



Аналитический обзор архитектур, моделей, методов и алгоритмов для локализации и трекинга неригидных объектов

Григорий Глебович **Гриценко**¹, Виталий Петрович **Фраленко**²

^{1,2}Институт программных систем им. А. К. Айламазяна РАН, Веськово, Россия

Аннотация. Компьютерное зрение требует анализа видеопотока, включающего извлечение информации из кадров, обнаружение определенных объектов и сбор данных о них. После обнаружения часто требуется выполнять *трекинг* или слежение за объектами в видеопотоке. *Неригидность* или изменчивость формы препятствует анализу объектов, усложняет их обнаружение и трекинг и ухудшает локализацию.

В обзоре рассмотрены архитектуры, модели, методы и алгоритмы, применяемые на практике при обнаружении и отслеживании неригидных объектов, и выделены перспективные решения.

Ключевые слова и фразы: неригидный объект, искусственная нейронная сеть, глубокое обучение, локализация объектов, трекинг объектов, обнаружение пожаров и задымлений, анализ медицинских изображений

Благодарности: Исследование выполнено за счет *гранта Российского научного фонда № 22-11-20001*  и гранта в форме субсидии из областного бюджета организациям Ярославской области.

Для цитирования: Гриценко Г. Г., Фраленко В. П. *Аналитический обзор архитектур, моделей, методов и алгоритмов для локализации и трекинга неригидных объектов* // Программные системы: теория и приложения. 2024. Т. 15. № 4(63). С. 111–151. https://psta.psiras.ru/read/psta2024_4_111-151.pdf

Введение

Обнаружение объектов – фундаментальная задача компьютерного зрения, направленная на выделение конкретных элементов из изображений или видеопотоков, таких как люди, транспортные средства и здания. Эта технология находит широкое применение в различных сферах, включая автономное вождение, медицинскую диагностику, системы безопасности и развлекательные приложения.

Ключевым элементом обнаружения объектов является классификация, то есть определение типа объекта на основе его визуальных характеристик. Алгоритмы машинного обучения анализируют пиксели изображения, распознавая формы, цвета, текстуры и другие особенности, чтобы, например, выделить пешехода на фоне или отличить автомобиль от велосипеда. Однако обнаружение объектов является лишь первым шагом. В динамике, когда объекты движутся, изменяют форму или исчезают из поля зрения, требуется более сложная технология – трекинг.

Трекинг – задача отслеживания объектов в видеопотоке с течением времени. Требуется не только определить местоположение объекта в каждом кадре, но и проследить его траекторию, скорость и направление движения. Точность трекинга напрямую зависит от уровня ригидности объекта, то есть его способности сохранять форму и геометрию.

Ригидные объекты, такие как автомобили или здания, легко отслеживаются, поскольку их внешние изменения незначительны. Неригидные объекты представляют собой более сложную задачу для алгоритмов трекинга. Эти объекты могут изменять форму, распадаться на части, сливаться с фоном или полностью исчезать из поля зрения. Например, облака, туман и дым, которые кажутся статичными, но на самом деле постоянно меняют свою форму и структуру. Огонь, пожары и взрывы – примеры неригидных объектов, которые сложно отследить из-за динамичного характера пламени, дыма и взрывной волны.

Алгоритмы трекинга должны учитывать деформацию объекта, его взаимодействие с окружающей средой, изменение освещения и множество других факторов, влияющих на его внешний вид. Решение этих задач имеет ключевое значение для совершенствования систем автономного вождения, прогнозирования стихийных бедствий, мониторинга окружающей среды и многих других областей. Дополнительные трудности включают изменение размера объекта, перекрытия объектов, появление и исчезновение объектов из поля зрения. Новые технологии, такие как глубокое обучение и 3D-трекинг, позволяют повысить точность и информативность трекинга неригидных объектов.

Цель настоящей работы заключается в сравнительном анализе архитектур, моделей, методов и алгоритмов обнаружения и трекинга неригидных объектов с выделением наиболее перспективных из них.

1. Способы повышения эффективности локализации неригидных объектов в кадре

1.1. Применение метода выравнивания гистограмм, модулей СВАМ и алгоритма CAPAN

Высокая влажность в морской среде ухудшает видимость. Поэтому для улучшения обнаружения пожара на судне с помощью камер в работе [1] предложен метод выравнивания гистограмм. Он позволяет повысить точность идентификации и локализации опасных неригидных объектов, таких как огонь и дым. Метод заключается в регулировании яркости изображений путем равномерного распределения значений по каналам RGB. При увеличении масштаба изображения его идентичность сохраняется. Анализ яркости изображений в исследовании [1] проводился с использованием функции кумулятивной плотности (*cumulative density*). Модель, обученная на обширном наборе данных, включающем более 25 тысяч изображений судов, достигла впечатляющей точности, превышающей 99%; для сравнения – 94% получено в ранней работе тех же авторов [2].

Авторы статьи [3] предлагают вариант усовершенствования архитектуры YOLOv5 путем добавления трех модулей сверточного блочного анализа (СВАМ, Convolutional Block Attention Modules) [4]. Для сбора детальной информации о градиентном потоке и повышения точности обнаружения небольших объектов использован модуль C2f ^{URL} вместо стандартного модуля C2. Модуль C2f реализуется двумя сегментационными головками (от англ «segmentation heads»), которые обучаются предсказывать семантические сегментационные маски для входного изображения. По своей природе модуль C2f является более быстрой реализацией модуля C2.

Эксперименты показали, что алгоритм достиг 82.36% mAP50 для обнаружения пожаров. Использование данных с маркировкой объектов, похожих на огонь, снизило количество ложных срабатываний. Модуль СВАМ улучшил mAP50 на 2.09%, а модуль C2f – на 1.35%. Объединение обоих модулей привело к увеличению mAP50 на 2.72%. В свою очередь, эксперименты по маркировке данных, подобных огню, показали, что возможно снизить количество ложных срабатываний на 3.7% для выбросов из дымовых труб и на 2.45% для облаков. Интеграция модулей СВАМ и C2f улучшает способность сети различать объекты, связанные с пожарами.

В статье [5] предложен алгоритм CAPAN (Channel Attention Path Aggregation Network), который позволяет искусственной нейронной

сети (ИНС) фокусироваться на особенностях канала с информацией о переднем плане. SARAN предлагает более эффективное объединение функций, используя механизм внимания к каналу (ECA, Efficient Channel Attention). Алгоритм, предложенный в статье, показал лучшую среднюю точность обнаружения и распознавания по сравнению с другими методами. Результаты тестирования показали улучшение точности распознавания на 8.3% для датасета Flame и на 2.1% для датасета Smoke по сравнению с базовой моделью.

1.2. Искусственная нейронная сеть со слоем Fire-RPG

В работе [6] представлен комбинированный подход к решению задачи раннего обнаружения возгораний в условиях городской среды. В рамках подхода рассматриваются ИНС Fire-RPG, метрика Wise-IoU, блоки для построения сетей CBAM [4], RepVGG [7] и GhostV2 [8]. В качестве основы Fire-RPG взята ИНС YOLOv8. Она состоит из трех основных частей: «позвоночника» (*backbone*), «шеи» (*neck*) и «головной части» (*head*), каждая из которых включает модули для локализации ключевых объектов.

Модули «позвоночника» и «шеи» содержат такие части как C2f, CBS и SPPF, которые обеспечивают эффективное представление объектов. CBS является базовым сверточным модулем, а C2f использует поэтапный метод повышения обучаемости и упрощения модели. SPPF (Spatial Pyramid Pooling Fast), в свою очередь, представляет собой специальный модуль для объединения знаний об объектах разного размера, оптимизированный для снижения вычислительных затрат.

«Шея» построена на структурах FPN (Feature Pyramid Network) и PAN (Path Aggregation Network), которые генерируют карты объектов разного масштаба с семантической и функциональной информацией. Эти структуры расширяют возможности обнаружения объектов разного размера. «Головная» часть состоит из трех слоев обнаружения объектов, которые выполняют операции вывода на основе информации, полученной от «шеи». Эти слои определяют местоположение объектов и их категории.

Для раннего обнаружения пожаров добавлен дополнительный слой обнаружения сверхмалых объектов – Fire-RPG, переход к обнаружению таких объектов снижает эффект от того, что целевые объекты неригидны. Этот слой увеличивает вычислительные затраты, что может повлиять на способность обнаружения в реальном времени. Пять высокочувствительных по используемым ресурсам модулей C2f заменены модулями GhostV2C2f (предложены авторами этой же работы).

Модули CBS заменены высокопроизводительными блоками RepVGG для более эффективного извлечения характеристик. В магистраль добавлены три блока СВМ, использующие пространственную информацию для выделения средних и мелких объектов. RepVGG применяет метод структурной репараметризации для трансформации сложных нейронных сетей в более простые и эффективные эквивалентные структуры. В работе приводится оптимизированная структура GhostNet и внедрен механизм внимания (*attention mechanism*) DFC (Decoupled Fully Connected). Полностью подключенные слои генерируют карты внимания с глобальным полем восприятия. Механизм внимания DFC и модуль Ghost работают в параллельных ветвях ИНС. Механизм DFC использует понижающую дискретизацию для сжатия карты объектов.

Авторами работы предложена функция потерь Wise-IoU (WIoU) для решения задачи обнаружения пожаров. Эта функция потерь направлена на снижение влияния низкокачественных данных и имеет три версии, каждая из которых ориентирована на различные типы якорных прямоугольников (*anchor boxes*). WIoU использует механизм внимания на расстоянии (*distance attention mechanism*). Второй важный фактор WIoU – коэффициент усиления по немонотонному градиенту (*non-monotonic gradient gain coefficient*), который определяет как функция потерь реагирует на низкокачественные и высококачественные якорные прямоугольники. Сравнение различных функций потерь приведено в таблице 1.

Таблица 1. Сравнение функций потери (в процентах)

Модель	Точность	Полнота	mAP50	mAP50-95
YOLOv8s-CIoU	84.8	70.9	79.1	47.7
YOLOv8s-GIoU	85.1	70.4	79.6	47.6
YOLOv8s-DIoU	82.3	71.6	79.8	47.5
YOLOv8s-SIoU	82.6	71.4	79.1	47.2
YOLOv8s-EIoU	82.5	71.9	79.2	47.4
YOLOv8s-WIoUv1	85.1	70.3	79.0	47.0
YOLOv8s-WIoUv2	84.7	70.7	79.6	47.6
YOLOv8s-WIoUv3	84.2	71.4	80.0	47.3

Механизмы внимания в нейронных сетях позволяют выделять ключевые функции на картах объектов и создавать карты внимания. Эти карты внимания затем умножаются на исходные карты объектов, что позволяет усилить информацию об этих объектах. Существует несколько типов механизмов внимания, каждый из которых имеет свои особенности. Например, SE Attention (Squeeze-and-Excitation Attention) фокусируется на канале, но не учитывает пространственную информацию. Это делает

его простым в вычислениях, но менее эффективным для задач, требующих учета пространственной информации. CA Attention (Channel and Spatial Attention), напротив, учитывает как канал, так и пространственную информацию, но может быть сложным в вычислениях и замедлять процесс обнаружения объектов. Канальное внимание фокусируется на том, *что* важно в данных, а пространственное внимание – на том, *где* эта информация находится.

СВАМ, который ранее уже упоминался в настоящем исследовании, сочетает преимущества обоих подходов, обеспечивая баланс между канальным и пространственным вниманием. Он имеет простую структуру и низкие затраты на вычисления, что делает его эффективным для широкого спектра задач.

Результаты работы показывают, что СВАМ значительно повышает mAP50 до 80.0% и mAP50-95 до 48.0%, по сравнению с базовой моделью YOLOv8s, имеющей показатели mAP50 и mAP50-95, равные соответственно 79.1% и 47.7%. Значения показателей эффективности работы механизмов внимания представлены в таблице 2.

Таблица 2. Сравнение механизмов внимания (в процентах)

Модель	Точность	Полнота	mAP50	mAP50-95
YOLOv8s	84.8	70.9	79.1	47.7
YOLOv8s-SE	82.7	71.7	79.6	47.9
YOLOv8s-SA [9]	84.6	70.4	79.4	48.0
YOLOv8s-ECA [10]	82.1	71.1	79.2	47.4
YOLOv8s-CA	83.5	71.5	79.8	48.0
YOLOv8s-SimAM [11]	82.0	71.1	79.3	47.8
YOLOv8s-CBAM	83.2	72.9	80.0	48.0

Fire-RPG демонстрирует высокую точность обнаружения малых и трудноразличимых объектов, таких как слабое пламя и тонкий дым. YOLOv8s пропускает эти объекты и имеет более низкие показатели достоверности. Для оценки способностей моделей YOLOv8s и Fire-RPG к обобщению, они проверены на трех наборах данных.

Набор данных D-Fire включает изображения дикой природы с двумя видами объектов: дымом и пламенем. Набор данных ForestFire сосредоточен на лесных сценах и усовершенствован с помощью методов аугментации данных. Набор данных DFS содержит изображения пожаров из различных сцен с тремя метками: пламя, дым и «другое».

Результаты испытаний показали, что модель Fire-RPG продемонстрировала лучшие результаты по сравнению с YOLOv8s на всех трех наборах данных. В частности, Fire-RPG улучшила показатель mAP50 более чем

на 1% на всех наборах данных. На наборе данных DFS улучшение показателя mAP50-95 составило 0.8%, что свидетельствует о высокой способности модели к обобщению. Кроме того, модель Fire-RPG показала лучшую эффективность обнаружения по сравнению с обычными сверточными нейронными сетями (СНС), достигнув показателя mAP50 79.2%.

Модель Fire-RPG подходит для использования в городских системах пожарной сигнализации, где она может быть применена для раннего обнаружения пожаров, что способствует повышению скорости реагирования и безопасности. Результаты работы СНС представлены в таблице 3.

Таблица 3. Результаты работы СНС (в процентах)

Набор данных	Модель	Точность	Полнота	mAP50	mAP50-95
D-Fire	Fire-RPG	78.7	71.7	79.2	46.1
	YOLOv8s	76.6	72.9	78.0	45.6
ForestFire	Fire-RPG	63.2	60.0	63.1	27.8
	YOLOv8s	63.0	58.5	61.8	27.5
DFS	Fire-RPG	61.2	49.2	54.9	28.1
	YOLOv8s	60.7	48.8	53.8	27.3

1.3. Дистилляция знаний для улучшения обнаружения пожаров

В работе [12] представлен подход ILKDG (Intermediate Layer Knowledge Distillation with Gram Matrix-based Feature Flow) к извлечению знаний, основанный на распространении знаний (*knowledge propagation*) между слоями. Для этого используется матрица потока процедур решения (FSP, Flow of Solution Procedure), основанная на матрице Грама и коэффициенте корреляции Пирсона, что позволяет дистиллировать знания. Авторы предлагают интегрировать YOLOv7 с методами дистилляции знаний и использовать новый подход для улучшения обнаружения пожаров. При дистилляции знаний модель меньшего размера, обученная повторять поведение тяжелой и точной модели-учителя, достигает схожих с ней показателей, существенно выигрывая в размере и скорости за счет упрощенной архитектуры.

В отличие от предыдущих работ, в предлагаемом подходе для построения матрицы Грама используются признаки из разных слоев. ILKDG сопоставлен с классическими и современными подходами на специально созданном наборе данных. Обнаружено, что ILKDG существенно повышает точность и скорость обнаружения пожаров: значение mAP50 улучшено на 2.9%, а mAP50-95 – на 2.7% без необходимости изменения параметров сети. Дополнительные данные представлены в таблице 4.

Таблица 4. Результаты оценки на тестовом наборе данных (в процентах)

Модель/Подход	Точность	Полнота	mAP50	mAP50-95
Teacher (YOLOv7x)	88.1	84.2	89.5	63.7
Student (YOLOv7)	83.3	81.9	84.6	60.4
KD	84.4	82.4	85.9	61.8
FitNet	84.7	82.2	86.0	61.1
FGD	85.0	82.1	85.4	62.1
KD++	84.6	82.3	85.8	62.2
ILKDG	86.2	83.1	87.0	63.1

1.4. Архитектуры DMCNN, SqueezeNet, Light-FireNet, DCN_Fire, AFSNet, SE-EFFNet, EFDNet и GLCT

В работе [13] авторы рассмотрели основанные на глубоком обучении подходы к обнаружению пожаров. Далее перечислены наиболее перспективные варианты.

В исследовании [14] предложена архитектура DMCNN (Deep Multi-scale CNN). Она включает многомасштабную сверточную структуру Inception для достижения масштабной инвариантности и использует многомасштабные аддитивные слои слияния для снижения вычислительных затрат при сохранении более динамичных и статичных характеристик задымления.

В методе, описанном в [15], в качестве магистральной сети использована облегченная нейронная сеть *SqueezeNet*^{URL} [16], сделана тонкая настройка архитектуры, в том числе применены сверточные ядра меньшего размера и исключены полностью подключенные плотные слои (*dense*).

Рассматривались архитектура Light-FireNet [17], сочетающая в себе легкий механизм свертки и новый архитектурный дизайн для уменьшения размера модели, и модель DCN_Fire [18] для оценки риска лесных пожаров с использованием PCA (Principal Component Analysis) для улучшения межклассовой различимости.

Адаптивная сеть выбора кадров AFSNet [19] автоматически выбирает наиболее полезные видеокadres для уменьшения избыточности функций и обеспечивает улучшенную расширенную свертку для уменьшения потери критически важной информации. Сеть изучает дискриминационные представления путем рассмотрения многомасштабной, контекстной и пространственно-временной информации.

В архитектуре SE-EFFNet [20] используется нейросеть EfficientNet-B3 в качестве магистральной сети извлечения полезных функций. В ней применяются комплексные автоэнкодеры для достижения эффективного выбора функций. EfficientNet обеспечивает баланс между разрешением входа, глубиной и шириной сети, одновременно внедряя сеть плотного

подключения (DenseNet) для обеспечения эффективного распознавания места пожара. Результаты эксперимента продемонстрировали, что по сравнению с базовой моделью SE-EFFNet достигла более низкого уровня ложноположительных и ложноотрицательных срабатываний, обладает более высокой точностью. Однако эта архитектура может страдать от переобучения, а ее производительность при обработке в реальном времени на устройствах с ограниченными ресурсами является средней.

Также, в исследовании [13] рассмотрены методы определения цвета и угла для предварительной обработки областей пламени и оптического потока для обнаружения дыма. Ху и др. [21] попытались объединить сети с глубокой сверточной долговременной памятью (LSTM, Long Short-Term Memory) и оптическими потоковыми методами для обнаружения пожара в режиме реального времени. Предложен метод, который сочетает использование канала со смешанным вниманием и слиянием функций. Метод протестирован на данных с туманной средой. Результаты эксперимента показали, что метод достиг точности 96.73% и оценки F1 87.22%.

Анализировалась применимость EFDNet [22], задействующей много-масштабный механизм выделения признаков (*multiscale feature extraction*) для улучшения пространственных деталей на этапе выделения признаков более низкого уровня с целью улучшения способности различать объекты, похожие на огонь. Неявный механизм глубокого контроля с помощью плотных пропускных соединений улучшает взаимодействие между информационными потоками, преобразуя мелкие пространственные объекты в семантическую информацию высокого уровня. Подход позволил достичь точности в 95.3% при компактном размере модели – всего 4.80 МБ.

В работе [23] предложена нейронная сеть для обнаружения пожаров под названием GLCT. Модель основывается на блоке-трансформере *MobileViT^{URL}*, что позволяет извлекать как глобальную, так и локальную информацию. Благодаря сочетанию SPP (Spatial Pyramid Pooling) с BiFPN (Bi-directional Feature Pyramid Network) для объединения признаков и включению в архитектуру YOLO-головки, построена целостная архитектура GLCT. Экспериментальные результаты показали, что GLCT достигла mAP 80.71%, обеспечив баланс между скоростью и точностью.

1.5. Искусственная нейронная сеть CAGSA-YOLO

Новая архитектура CAGSA-YOLO [24] отличается использованием модуля CARAFE (Content-Aware ReAssembly of FEatures) для повышения дискретизации, добавлением нового уровня обнаружения масштаба для объектов порядка 4x4 и использованием облегченной конструкции Sampling Ghost, в которой C3-модуль заменяется модулем C3Ghost. Модель CAGSA-YOLO разработана на основе оригинальной архитектуры YOLOv5s.

В отличие от традиционных трехуровневых архитектур, CAGSA-YOLO использует четырехуровневую, что достигается заменой традиционной повышающей дискретизации на CARAFE. Механизм SA (Self-Attention Mechanism) повышает интерес к целевым объектам и подавляет ненужные функции. Изменения позволили улучшить производительность модели и повысить ее эффективность при обнаружении объектов. CAGSA-YOLO обнаруживает больше деталей и имеет более высокую точность по сравнению с традиционным YOLOv5. Значения показателей проведенных экспериментов представлены в таблице 5.

Таблица 5. Сравнительная таблица экспериментов (в процентах)

Модель	Точность	Полнота	mAP50
YOLOv5s	93.7	78.5	83.4
YOLOv5s+CARAFE	92.3	79.5	84.1
YOLOv5s+CARAFE+scale	89.7	79.7	84.5
YOLOv5s+CARAFE+scale+Ghost	87.7	81.4	84.5
CAGSA-YOLO (YOLOv5s+CARAFE+scale+Ghost+SA)	89.7	80.1	85.1

1.6. Двухуровневые СНС с пространственно-временным вниманием

В статье [25] представлена двухуровневая СНС STCNNsmoke для сегментации областей с дымом, основанная на механизме пространственно-временного внимания. В пространственной части сети извлекаются характеристики объектов переднего плана с использованием модели ранжирования с частичным контролем. Во временной (или темпоральной) части характеристики оптического потока используются для представления динамических характеристик дыма, таких как рассеивание и разветвление.

Эксперименты показали, что полученное среднее значение IoU составляет 83.52%, а среднее значение F1-меры – 85.75%. Это на 3.34% и 2.72% выше, чем у трех сравниваемых моделей соответственно. Для видеороликов, содержащих большое количество легкого и разреженного дыма, значения метрик IoU и F1 предлагаемой СНС немного улучшены по сравнению с другими СНС. Результаты экспериментов на тестовом наборе данных представлены в таблице 6.

В работе [26] рассмотрена пространственно-временная нейросеть *STCNet^{UR}* (Spatio-Temporal Cross Network). Предложена архитектура для обнаружения дыма, включающая обработку как пространственной, так и временной информации. Пространственная часть обрабатывает кадры как отдельные объекты, извлекая многомасштабную пространственную

Таблица 6. Средние значения IoU и F1-меры для тестовых примеров (в процентах)

Модель	IoU	F1-мера
FCN	75.25	79.97
Deeplab V3+	78.83	81.92
RANet	80.18	83.03
STCNetSmoke	83.52	85.75

пирамиду признаков. Остаточные кадры выделяются путем вычитания соседних кадров.

RGB-кадры используются для временного контура, ограничивая максимальное значение остаточного пикселя. Остаточные кадры фокусируются на движущихся объектах, подавляя помехи. Временная часть сети извлекает особенности движения из различий в кадрах, она имеет такую же архитектуру, как и пространственная, но весовые коэффициенты иные. Модель создает карты объектов с разной глубиной и размером.

Функциональная пирамидальная сеть (FPN, Feature Pyramid Network) улучшает представление признаков. Для многоуровневой структуры используются карты объектов от разных остаточных блоков. Пространственно-временная двойная пирамида улучшает способность распознавания признаков дыма. Карты пространственной и временной частей ИНС суммируются для участия в выводе модели. Суммирование карт объектов выполняется поэлементно. Для оптимизации весов модели использовался стохастический градиентный спуск.

Проведено сравнение с другими базовыми моделями, в том числе с MobileNetV2. Результаты сравнения с первичными методами распознавания видео показали, что SEResNeXt-50 обеспечивает лучшую производительность. Рассмотрены три варианта STCNet: STCNet-A, STCNet-B и STCNet-C. STCNet-A представляет собой типичную двухпоточную сеть, STCNet-B использует однонаправленное слияние элементов, а STCNet-C применяет пространственно-временное слияние объектов.

Результаты проведенных экспериментов показали эффективность многомасштабного слияния функций в STCNet. Для визуализации активных областей в кадрах применялся метод Grad-CAM. Продемонстрировано, что STCNet способна фокусироваться на области задымления, игнорируя помехи от пара. STCNet показывает значительные улучшения значений F-меры в обнаружении промышленного дыма по сравнению с конкурентами (до 6.20%).

1.7. Алгоритмическая модель «MS Transformer»

В статье [27] предложена алгоритмическая модель «MS Transformer», которая использует подход самоконтролируемого обучения (*self-supervised learning*) для применения случайной маски к входному изображению с целью восстановления входных признаков, получения более полного вектора признаков и фильтрации избыточного шума. Скользящее окно (*sliding window*) с механизмом локального самоанализа используется для повышения оценки внимания к обнаруживаемым небольшим объектам. Модель «MS Transformer» учитывает особенности медицинских изображений, включая низкое разрешение и общую зашумленность.

Процесс обработки изображений заключается в разделении входного изображения на части и выполнении операции маскирования, кодировании немаскированных патчей для получения признаков изображения, после чего происходит обучение модели с использованием декодера, восстанавливающего пиксели. Каждое медицинское изображение сегментируется на регулярные участки, затем они случайным образом выбираются и маскируются. Случайная маска способствует устранению избыточности и изучению признаков на глубоком уровне.

Архитектура «MS Transformer» включает слой реконструкции изображения, «Swin Transformer»^{URL} и фрагменты архитектуры YOLOv5. Иерархический трансформер состоит из двух последовательных трансформерных блоков, предсказывающих класс поражения и ограничивающую рамку.

Без механизма маскирования точность модели снижается на 8.6%, а mAP – на 9.0%. Эффект иерархического трансформера на модель выше, чем у механизма маски. Использование только маскирования для задачи обнаружения объектов снижает точность модели на 15.6% и уменьшает значение mAP на 16.1%.

Проведены эксперименты с коэффициентом маски в диапазоне от 10% до 80%. Когда коэффициент находится в диапазоне 10–30%, точность распознавания модели на эталонных наборах данных BCDD и DeepLesion составляет 86% и 82% соответственно. При коэффициенте маски 40% точность распознавания модели значительно улучшается, достигая 94.3% и 87.1% на тех же наборах данных.

1.8. Нейросетевой алгоритм GSSD⁺⁺

В работе [28] представлен алгоритм GSSD⁺⁺^{URL} для выявления пораженных участков печени на изображениях. Он использует динамическое

межфазное (*dynamic interphase*) извлечение пространственной салиентности (*spatial saliency*) из многопоточковых признаков (*multistream features*) для обнаружения очагов поражения печени.

Разработка базируется на автоматической сегментации паренхимы печени и сосудов на КТ с использованием глубокого обучения [29]. Для лучшего обнаружения повреждений предложен механизм многофазного пространственного управления вниманием, основанный на деформируемом ядре свертки. Этот механизм учитывает основные фазы многофазных входных данных, что повышает точность обнаружения повреждений и снижает вероятность ложных срабатываний.

Чтобы точнее обнаруживать повреждения, ядра свертки с обучаемыми и динамичными смещениями прогнозируют геометрические смещения для многофазных функций, а дополнительную визуальную подсказку для многофазного выравнивания дает карта врат самовнимания (*self-attention gate map*). Совместное использование карты врат самовнимания и модулей деформируемой свертки значительно улучшает эффективность обнаружения ошибочно зарегистрированных многофазных наборов данных и снижает вероятность ложных срабатываний.

Алгоритм GSSD⁺⁺ продемонстрировал точное и надежное обнаружение поражений печени, особенно при высоком пороге перекрытия. Показатели представлены в таблице 7.

Таблица 7. Анализ эффективности алгоритма GSSD⁺⁺ (в процентах)

Алгоритм	mAP для IoU50	mAP для IoBB50
GSSD ⁺⁺	67.87	77.22
GSSD ⁺⁺ без Phase-wise Offsets	61.50	74.08
GSSD ⁺⁺ без DC Module	61.60	71.89
GSSD ⁺⁺ без SA Modules@Base	62.59	74.40
GSSD ⁺⁺ без SA Modules@Fusion	64.03	76.49
GSSD ⁺⁺ без Interphase Attention	63.82	73.56

1.9. Искусственная нейронная сеть DEGPR

В статье [30] представлена ИНС *DEGPR*^{UR} (Deep Guided Posterior Regularization), предназначенная для подсчета и обнаружения клеток разных классов на медицинских изображениях. Модель DEGPR помогает в обнаружении объектов, акцентируя внимание на уникальных особенностях клеток, которые могут быть предоставлены патологоанатомами или извлечены из визуальных данных.

Дискриминация классов клеток базируется на пострегистрационной регуляризации (PR, Posterior Regularization) для двух типов признаков:

явных и неявных. Явные признаки вводятся под непосредственным руководством экспертов-патологоанатомов. Неявные признаки – вкрапления признаков для каждого класса, исследуемые с помощью супервизорного метода контрастных потерь (*contrastive loss*).

Векторы различий признаков используются для формирования Gaussian Mixture Model (GMM), применяемой при сравнении истинных и прогнозных ограничивающих рамок. Она запоминает базовое распределение объектов с помощью оценки плотности. Для каждого класса готовятся две отдельные GMM, моделирующие истинные и предсказанные ограничивающие рамки.

Чтобы выровнять распределения признаков, минимизируется KL-расхождение (*Kullback-Leibler divergence*) между истинными и прогнозируемыми GMM. Метод Монте-Карло, аппроксимирующий интеграл с помощью выборок, оценивает KL-расхождения. Потери для каждой пары классов вычисляются и нормализуются по количеству пар. Итоговая величина потерь рассчитывается как среднее значение потерь по всем парам классов.

Преобразование изображений в неявные векторы признаков реализует нейросеть ResNet18. Уменьшение размерности признаков осуществляет PCA, сохраняя при этом 90% дисперсии.

Модель протестирована на наборах данных CoNSeP, MoNuSAC и MuCeD и обеспечивает до 9% абсолютного прироста точности в обнаружении объектов, улучшает эффективность обнаружения и подсчета клеток на наборах данных. В случае набора MuCeD она повышает mAP на 3–9% и снижает среднюю абсолютную ошибку на 10–35%. Эффективность в улучшении производительности систем обнаружения объектов характеризует увеличение F1-меры модели для прогнозирования целиакии с 77% до 90%.

1.10. Искусственная нейронная сеть RCS-YOLO

В работе [31] предложена перепараметризованная СНС *RCS-YOLO^{UR}*, в которой объединены архитектура ShuffleNet и блоки RepVGG/RepConv. Новый модуль RCS-OSA извлекает семантическую информацию.

RCS опирается на структурную перепараметризацию для улучшения свертки, разделяет входной тензор на два канальных тензора и применяет для каждого свои операции обработки данных. Модуль одношаговой агрегации (OSA, One-Shot Aggregation) для объединения объектов в DenseNet переключает каналы в случайном порядке для повышения информативности представления. В свою очередь, перетасовка каналов улучшает объединение информации между тензорами, снижает

вычислительную сложность. Она обеспечивает полную взаимосвязь между функциями ввода и вывода, которые иначе ограничены внутри групп сверток. Проведены эксперименты, показавшие, что RCS-YOLO превосходит YOLOv6, YOLOv7 и YOLOv8: точность – на 1%; скорость вывода – на 60%.

1.11. Искусственная нейронная сеть BGF-YOLO

В статье [32] рассмотрена архитектура BGF-YOLO на основе YOLOv8. BGF-YOLO объединяет двухуровневую систему маршрутизации внимания (BRA, Bi-level Routing Attention) и сети с обобщенными функциональными возможностями (GFPN, Generalized Feature Pyramid Networks), в отличие от YOLOv8 добавлена четвертая детекторная головка.

Произведена оценка модели *BGF-YOLO*^{URL}, результаты отображены в таблице 8. В частности, BGF-YOLO показала абсолютное повышение на 1.2%, 4.5%, 4.7% и 0.7% по сравнению с YOLOv8x в показателях точности, полноты, mAP50 и mAP50-95 соответственно. BGF-YOLO превзошла DAMO-YOLO-L [33], RCS-YOLO [34] и другие высокоточные детекторы.

Таблица 8. Результаты сравнения показателей моделей (в процентах)

Модель	Точность	Полнота	mAP50	mAP50-95
YOLOv8	90.7	88.1	92.7	64.6
DAMO-YOLO-L	–	–	90.0	61.0
RCS-YOLO	90.8	88.5	87.8	68.0
BGF-YOLO	91.9	92.6	97.4	65.3

1.12. Искусственные нейронные сети FireNet и FireNet-v2

Авторами статьи [35] предложена архитектура нейронной сети *FireNet*^{URL}. Данная архитектура оптимизирована для мобильных и встроенных приложений, обеспечивает высокую производительность и более 24 кадров в секунду на Raspberry Pi 3B. Сеть состоит из 14 слоев, включая слои «объединения» (*pooling*), «выбывания» (*dropout*) и выходной слой Softmax. FireNet имеет три уровня свертки с ReLU (Rectified Linear Unit) в качестве функции активации, за исключением последнего слоя с Softmax. Сеть имеет следующие параметры входных данных и размеров слоев. Первый слой принимает на вход тензоры с полноцветными изображениями размером (64×64×3). Входной сигнал может быть увеличен до (128×128×3) без значительного снижения частоты кадров. В каждом из двух последующих слоев свертки удваиваются входные признаки, сохраняя размер ядра неизменным. Последующие слои включают плотный слой (*Dense layer*)

с 256 нейронами и два плотных слоя с 128 нейронами. Последний – плотный слой с двумя нейронами для предсказания. Используется «выбывание» со слоями свертки наряду с плотными слоями. Для слоев свертки выбрано стандартное значение «выбывания», равное 0.5. Для последующего плотного слоя выбрано значение, равное 0.2.

В работе [36] представлена ИНС FireNet-v2, превосходящая предыдущую версию в оптимизации и качестве обнаружения. Она имеет значительно меньшее количество параметров по сравнению с FireNet, сохраняя при этом эффективность обнаружения пожаров и производительность. Конкретные модификации в работе по сравнению с FireNet заключаются в том, что количество фильтров в первом, втором и третьем конволюционных слоях составляет 15, 20 и 30 соответственно, в то время как в FireNet – 16, 32 и 64, что привело к значительному уменьшению числа обучаемых параметров. Функция активации, используемая в последнем конволюционном слое и обоих внутренних плотных слоях, – сигмоидальная. В FireNet использовалась функция ReLU. Вместо разделения на тренировочные и тестовые наборы в соотношении 70 к 30, выбрано разделение 90 к 10, что позволило предлагаемой модели FireNet-v2 обучаться на большем количестве данных о пожарах. В модели в качестве входа используется тензоры с изображениями размером $64 \times 64 \times 3$. Промежуточные слои включают каскад сверток, слои отсева и усреднения, а функции активации – ReLU и сигмоид. Все три конволюционных слоя соединены с объединением средних значений. После сверточных слоев появляется выравнивающий слой (*Flatten layer*) и два плотных слоя с 256 и 128 нейронами каждый. После начального плотного слоя используется «выбывание» со значением 0.2. Полностью связанный плотный слой с выходом Softmax имеет два нейрона и является слоем предсказания, выводящим сигналы «пожар» и «не пожар». FireNet-v2 имеет точность 98.43%, используя всего 0.32 млн. параметров, по сравнению с 96.53% точности FireNet, использующей 0.65 млн. параметров. Результаты тестирования моделей представлены в таблицах 9–10.

ТАБЛИЦА 9. Сравнение точности при тестировании FireNet-v2 (на наборе данных Foggia)

Модель	Точность, %	Количество параметров, млн.
FireNet	96.53	0.65
FireNet-v2	98.43	0.32
Muhammad 1 [15]	94.50	0.42
Abdullah [37]	97.50	0.51
Yakhyokhuja [38]	99.53	9.08
Muhammad 2 [39]	94.43	7.00

Таблица 10. Сравнение точности при тестировании модели FireNet-v2 (на наборе данных, используемом в работе [35])

Модель	Точность, %	Количество параметров, тыс.
FireNet	93.91	646.820
FireNet-v2	94.95	318.460
Saponara [40]	96.58	171.296
Ayala [41]	96.33	956.226
Elhanashi [42]	93.60	23.482

1.13. Искусственные нейронные сети NASNet-A-OnFire и ShuffleNetV2-OnFire

В статье [43] предложены две компактные ЧНС с пониженной сложностью архитектуры для повышения эффективности – *NASNet-A-OnFire* и *ShuffleNetV2-OnFire*^{URL}. NASNet-A-Mobile [44] и ShuffleNetV2 [45] экспериментально оптимизированы с помощью специальных фильтров, предложенных в работе [46]. В результате NASNet-A-OnFire и ShuffleNetV2-OnFire обеспечили точность локализации 95% и 97% соответственно. Для локализации пожаров на изображениях использовалась суперпиксельная сегментация.

Сети NASNet-A-Mobile и ShuffleNetV2 выбраны за их компактность и высокую производительность в ImageNet-классификации. Обе архитектуры имеют модульную структуру, позволяющую легко изменять или удалять отдельные ячейки. NASNet-A-Mobile содержит последовательность из нормальных и восстановленных ячеек, повторяющуюся три раза. Нормальная ячейка состоит из трех 3×3 сверток и двух 5×5 , в то время как восстановленная ячейка имеет одну 3×3 , две 5×5 и две 7×7 свертки. Остальные слои в NASNet-A-Mobile включают усреднение или максимальное объединение. ShuffleNetV2 состоит из начального слоя свертки 3×3 , слоя объединения и трех типов ячеек: нормальной, уменьшающей и восстанавливающей. Нормальная ячейка разделяется на две половинки, каждая из которых взаимодействует с тремя ядрами с различными типами сверток. Уменьшающая ячейка объединяет все входные данные, в то время как восстановленная ячейка похожа на нормальную, но с шагом 2 в глубине свертки.

Авторами исследования [43] проведены эксперименты с упрощением описанных выше архитектур ЧНС. Упрощенные архитектуры включают удаление последнего полностью подключенного слоя и создание нового линейного слоя для бинарной классификации. Базовая модель NASNet-A-Mobile была предварительно обучена для классификации ImageNet, далее весовые коэффициенты замораживались за исключением последнего слоя. Выполнено сокращение количества фильтров в NASNet-A-Mobile и

ShuffleNetV2 для улучшения обобщения (с 1056 до 480). Архитектура ShuffleNetV2 содержит 340 тысяч параметров. Как и в экспериментах с моделью NASNet-A-Mobile, проведено замораживание параметров на первую половину обучения, удаление фильтров с низкими значениями L2-нормы и последующее переобучение модели.

1.14. Искусственная нейронная сеть MVMNet

В статье [47] предложен метод мультиориентированного обнаружения целевых объектов, основанный на модуле VAM (Value conversion-Attention Mechanism) и смешанной NMS (Non-Maximum Suppression). Используется softpool-объединение пространственных пирамид для сохранения информации о характеристиках; рассматриваемые мини-блоки изображения делятся на части в разных масштабах, далее над этими частями производятся операции свертки. Изображение конвертируется методом главных компонент, далее результат конвертации считывается как по строкам, так и по столбцам. Применяется комбинация гибридных не максимальных методов DIoU-NMS и Skew-NMS с двумя порогами для расстояний и углов для предотвращения ложного и пропущенного обнаружения. MVMNet превосходит другие методы, такие как D-RFCN+SNIP и RDD, по точности и скорости.

Архитектура MVMNet успешно справляется с задачами обнаружения дыма в сложных условиях, таких как наличие облаков и тумана. На тестовом наборе данных *Forest_Fire_Smoke_DATA*^{[URL](#)} mAP50 достигла 88.05%.

1.15. 3D-сверточная нейронная сеть с расширенным региональным контекстом

В статье [48] представлена ИНС *3DCE*^{[URL](#)} (3D-СНС с расширенным региональным контекстом), которая объединяет информативные признаки из нескольких двумерных изображений для эффективного использования контекстной информации. ИНС выполняет агрегацию карт объектов для окончательного прогнозирования. 3DCE объединяет M трехканальных изображений в M двумерных канальных карт объектов. Центральный срез содержит ограничивающую рамку «исходная истина», а другие фрагменты обеспечивают трехмерный контекст.

Результаты экспериментов показали, что доработанная R-FCN с ключевым фрагментом обеспечивает низкую точность из-за важности трехмерного контекста; даже при работе с 11 срезами RCNN не удается превзойти 3DCE, которая показала лучшие результаты для небольших (<10 мм) поражений. Показатели чувствительности СНС представлены в таблицах 11–12.

Таблица 11. Чувствительность (%) при различных FP (False Positive) на изображении на тестовом наборе DeepLesion; в качестве критерия расчета перекрытия используется IoU

FP на изображение	0.5	1	2	4	8	16
Без 3D-контекста	48.60	60.57	71.19	79.15	84.77	88.42
Faster RCNN, 3 среза	56.90	67.26	75.57	81.62	85.83	88.74
Оригинальная R-FCN, 3 среза	55.70	67.26	75.37	82.21	86.26	89.19
Доработанная R-FCN, 3 среза	56.49	67.65	76.89	82.76	87.03	89.82
Data-level fusion, 11 срезов	58.49	70.03	77.89	83.02	86.71	89.19
3DCE, 9 срезов	59.32	70.68	79.09	84.34	87.81	89.62
3DCE, 27 срезов	62.48	73.37	80.70	85.65	89.09	91.06

Таблица 12. Чувствительность (%) при четырех ложных срабатываниях на изображении на тестовом наборе DeepLesion

	Тип поражения								Диаметр поражения, мм			Интервал между срезами, мм	
	LU	ME	LV	ST	PV	AB	KD	BN	<10	10~30	>30	<2.5	>2.5
RCNN	86	83	88	70	80	79	79	65	75	84	81	81	82
3DCE	89	88	90	74	84	84	82	75	80	87	84	86	86

2. Способы трекинга неригидных объектов

2.1. Трекинг с помощью многомасштабных пространственно-временных дискриминантных карт салиентности

В статье [49] представлен метод отслеживания объектов, основанный на принципе пространственно-временной согласованности и идентификации характерных признаков. Разработана специальная глубокая полностью конволюционная нейронная сеть (TFCN, Tailored Fully Convolutional Neural network) для моделирования локального приоритета салиентности (*local saliency prior*) для заданного региона изображения, которая не только обеспечивает попиксельный вывод, но и объединяет семантическую информацию. Предложен многомасштабный мультирегиональный механизм для генерации карт локальной салиентности региона, который эффективно учитывает визуальное восприятие с различными пространственными схемами и вариациями масштаба. Карты салиентности объединяются с помощью метода взвешенной энтропии, что позволяет получить окончательную дискриминационную карту салиентности. Отмечается алгоритм слежения, основанный на предложенной пространственно-временной согласованной карте салиентности (STCSM, Spatial-Temporal Consistent Saliency Map). Алгоритм используется для классификации по фону и онлайн-обновлений. TFCN обеспечивает попиксельные выходные данные

и интегрирует семантическую информацию о целях. TFCN обучается с помощью стохастического градиентного спуска и адаптируется к онлайн-визуальному отслеживанию.

Эксперименты и результаты показали, что предлагаемый метод превосходит другие по точности и F-мере, в том числе проведено сравнение с методами LEGS, MDF и DRFI. Предлагаемый трекер показывает точность 84.2% и коэффициент успешности 56.1%. В бенчмарке OTB-50 трекер работает лучше, чем второй лучший трекер неригидных объектов – OGBT [51]. Производительность трекера OGBT: точность 74.8%, коэффициент успешности 52.4%. Отмечается, что разработанный трекер может генерировать попиксельные карты салиентности, что полезнее, чем результаты отслеживания по ограничительным рамкам. Детали реализации имеются в репозитории [*tracker_benchmark_v1.0*](https://github.com/zhengyu1992/tracker_benchmark_v1.0)^{URL}.

2.2. Трекинг с помощью деформируемых патчей

В работе [52] была предложена структура для отслеживания объектов с использованием деформируемых патчей и ядерного корреляционного фильтра с сохранением формы (SP-KCF, Shape-Preserved Kernelized Correlation Filter) для учета формы цели. Адаптация деформируемых патчей к сложной топологии и изменениям целей обеспечивается за счет фотометрической дискриминации (*photometric discrimination*) и вариативности формы патчей. Это позволяет отслеживать отдельные области и вычислять контуры объектов на основе информации о форме патчей. KCF-трекер учитывает неригидную форму патчей, что обеспечивает надежное отслеживание. Обучение классификатора происходит в области Фурье с использованием участков изображения вокруг мишени. Результаты сравнения методов представлены в таблице 13.

Таблица 13. Результаты оценки сравниваемых методов (F-мера в %)

Pix [53]	DF [54]	HT [55]	SLSM [56]	RPT [57]	SP-KCF
40.33	42.06	45.73	47.84	54.02	56.76

2.3. Трекинг с использованием параметрического активного контура и модели распределения точек

В работе [58] исследуется метод отслеживания неригидных объектов на изображениях с загроможденным фоном в многолюдных сценах. Авторы предлагают комбинированное использование параметрических активных контуров (PAC, Parametric Active Contours) и модели распределения точек (PDM, Point Distribution Model). PAC и PDM дополняют друг друга,

повышая устойчивость системы к окклюзии и изменениям внешнего вида объекта. Трекер состоит из этих двух частей и уровня управления, обеспечивающего обмен информацией между частями трекера, что позволяет продолжить работу в случае отказа одного из трекеров.

При рассмотрении частей трекера по отдельности стоит выделить, что PDM-трекер вычисляет вектор признаков и генерирует вектор формы, а PAC-трекер инициализируется на основе PDM и точно определяет область с целевым объектом. Уровень управления объединяет результаты PAC- и PDM-трекеров для обновления PDM. Неотъемлемой частью трекера является алгоритм обмена данными между его компонентами. Алгоритм использует ROI (Region-Of-Interest) для извлечения точек интереса и сопоставления с предыдущими кадрами. PAC инициализируется на основе точечной модели или контура предыдущего кадра. С целью повышения точности отслеживания применяется конвергенция PAC и обновление PDM. Стоит отметить, что априорная модель PDM используется для отслеживания работы системы и инициализации контура при окклюзиях. Для повышения надежности работы трекера обновления PDM отключены в моментах перекрытий.

2.4. Трекер на основе многообразий распределения точек

В работе [59] рассматривается модель трекинга неригидных объектов, разработанная для обработки текстурированных объектов в многолюдных сценах. Модель опирается на отслеживание характерных точек на изображениях и их сопоставлении. Модель представляет собой набор геометрических фигур, которые могут быть адаптированы к изменениям в объекте. Порог стробирования θ используется для фильтрации расстояний между объектами в разных кадрах, превышающих θ . Изображение центрируется и масштабируется для соответствия модели. Преобразование подобия T используется для отображения формы из системы отсчета изображения. Параметры сходства модели с изображением отсеиваются с помощью фильтра Калмана. Фильтр настраивается для различных типов движений.

Особенностью предложенной модели является ее универсальность: помимо людей она может успешно применяться для отслеживания других объектов, например, автомобилей. Важным отличием является сохранение меток объектов даже при частичных перекрытиях, что особенно актуально в условиях динамичной сцены.

2.5. Алгоритм трехмерной реконструкции неригидных объектов с использованием камеры глубины

В статье [60] рассматривается алгоритм выделения особых точек, предназначенный для создания трехмерных моделей объектов реального мира с использованием RGB-D-камеры Microsoft Kinect. Входные данные для алгоритма включают RGB-изображение, карту глубины и облако точек, полученные с помощью камеры Kinect. Отметим, что на этапе выделения признаков алгоритм находит общие точки между двумя объектами, используя для этого вычисляемые дескрипторы.

В процессе поиска подходящих ключевых точек исследованы четыре детектора: Harris [61], ISS [62], SUSAN [63] и SIFT [64]. Для описания признаков протестированы пять дескрипторов: PFH [65], SHOT [66], SC [67], RIFT [68] и RSD [69]. Для сравнения эффективности комбинаций детектор-дескриптор использован набор данных «freiburg1_teddy» из бенчмарка [70]. На этапе сопоставления признаков алгоритм находит соответствия между облаками точек с использованием модифицированного RANSAC (Random Sample Consensus). Для выравнивания совпадений используется алгоритм Kabsch, на этапе неригидной регистрации этот алгоритм находит преобразования для выравнивания облаков точек между собой. Итеративный алгоритм ближайших точек ICP используется для согласования выровненных поверхностей. На заключительном этапе строится трехмерная модель объекта на основе выровненных облаков точек. Сравнение работы различных модификаций ICP сделано с использованием среднего квадратического отклонения (СКВО), результаты тестирования алгоритмов представлены в таблице 14.

Таблица 14. Сравнение эффективности работы различных модификаций ICP

Алгоритм регистрации	СКВО, мм
ICP-R	4.7
ICP-nR	3.4
ICP-nRC	3.2

3. Оценка и выбор наиболее перспективных решений

С целью формирования более предметного обоснования выбора наиболее перспективных решений для работы с неригидными объектами составлены сводные таблицы 15–16 для каждого из основных разделов настоящего аналитического обзора. В них представлены показатели эффективности и общая характеристика обнаруженных в результате проведения аналитического обзора решений.

Таблица 15. Сводная таблица решений, рассмотренных в разделе «Способы повышения эффективности локализации неригидных объектов в кадре»

№	Решение	Эффективность	Комментарий
1	Применение метода выравнивания гистограмм [1], модулей СВМ [3] и алгоритма CAPAN [5]	Метод выравнивания гистограмм обеспечил рост точности с 94 до 99%; СВМ улучшил mAP50 на 2.09%; CAPAN улучшил mAP50 на 2.1–8.3%	Метод выравнивания гистограмм регулирует яркость изображений, равномерно распределяя значения по каналам RGB. Предложен ряд усовершенствований архитектуры YOLOv5, в том числе – использование модуля сверточного блочного анализа СВМ и алгоритма CAPAN, позволяющего ИНС фокусироваться на особенностях переднего плана.
2	Fire-RPG [6]	Рост mAP50 на 1.1–1.3%	Слой Fire-RPG для обнаружения сверхмалых объектов позволяет снизить негативный эффект неригидности целевых объектов. Слой увеличивает вычислительные затраты и может повлиять на способность обнаружения в реальном времени. Используется функция потерь Wise-IoU.
3	Дистилляция знаний на основе потока процедур решения [12]	Рост mAP50 на 2.9%, рост mAP50-95 на 2.7%	Предлагается интегрировать YOLOv7 с методами дистилляции знаний и использовать метод ILKDG.
4	DMCNN [14]	–	Используется многомасштабная сверточная структура Inception для достижения масштабной инвариантности, применяются аддитивные слои слияния для снижения вычислительных затрат при сохранении характеристик задымления.
5	<i>SqueezeNet</i> ^(m) [16]	–	Тонкая настройка архитектуры, в том числе осуществляется использование сверточных ядер меньшего размера и исключение плотных полностью подключенных слоев.
6	Light-FireNet [17]	–	Облегченные свертки и методы минимизации размера модели.

7	DCN_Fire [18]	—	Использование PCA для улучшения межклассовой различимости.
8	AFSNet [19]	—	Реализуется автоматический выбор полезных видеок кадров для уменьшения избыточности функций и улучшение свертки для минимизации потери важной информации. Сеть изучает дискриминационные представления, используя многомасштабную, контекстную и пространственно-временную информацию.
9	SE-EFFNet [20]	—	EfficientNet-B3 используется как магистральная сеть для извлечения полезных функций, применяя комплексные автоэнкодеры для эффективного выбора функций. Сеть балансирует разрешение входа, глубину и ширину сети.
10	Методы определения цвета и угла для предварительной обработки областей пламени и оптического потока, LSTM-сеть [21]	Точность 96.73%, F1-мера 87.22%	Объединение LSTM-сети с оптическими потоковыми методами.
11	EFDNet [22]	Точность 95.3%	Используется многомасштабный механизм выделения признаков для улучшения способности различать объекты, похожие на огонь. Применяется неявный механизм глубокого контроля с плотными пропускными соединениями для улучшения взаимодействия между информационными потоками, преобразуя мелкие пространственные объекты в семантическую информацию высокого уровня.
12	GLCT [23]	mAP 80.71%	Нейронная сеть на основе блока-трансформера MobileViT. Модель извлекает глобальную и локальную информацию, сочетая SPP и BiFPN для объединения признаков.
13	CAGSA-YOLO [24]	Рост mAP50 на 1.7%	Используется модуль CARAFE для повышения дискретизации, новый уровень обнаружения масштаба для объектов 4x4, облегченная конструкция Sampling Ghost с заменой C3-модуля на C3Ghost.

14	Двухуровневые СНС с пространственно-временным вниманием [25]	Рост F-меры на 2.72–6.20%	В пространственной части извлекаются характеристики объектов переднего плана с использованием модели ранжирования с частичным контролем. Во временной части используются характеристики оптического потока для представления динамических характеристик дыма, таких как рассеивание и разветвление.
15	MSTransformer [27]	Точность 87.1–94.3%	Используется самоконтролируемое обучение для применения случайной маски к входному изображению для восстановления признаков, получения полного вектора признаков и фильтрации шума. Скользящее окно с локальным самоанализом повышает внимание к небольшим объектам.
16	Нейросетевой алгоритм GSSD++ [28]	mAP 67.87–77.22%	Используется динамическое межфазное извлечение пространственной салиентности из многопоточковых признаков для обнаружения очагов поражения. Механизм многофазного пространственного управления вниманием на основе деформируемого ядра свертки повышает точность обнаружения повреждений и снижает вероятность ложных срабатываний.
17	DEGPR [30]	Рост точности на 9%, рост полноты на 4.5%, рост mAP на 3–9%, рост F1–меры на 13%	Модель акцентирует внимание на уникальных особенностях клеток, опираясь на пострегистрационную регуляризацию для явных и неявных признаков.
18	RCS-YOLO [31]	Рост точности на 1%	Используется структурная перепараметризация модели, входной тензор разделяется на два канальных тензора, для каждого нового тензора применяются свои операции обработки данных. Перетасовка каналов улучшает объединение информации и снижает вычислительную сложность.
19	BGF-YOLO [32]	Рост точности на 1.2%, рост mAP50 на 4.7%, рост mAP50-95 на 0.7%	Объединены двухуровневая система маршрутизации внимания и нейросеть с обобщенными функциональными возможностями. Добавлена четвертая детекторная головка для фокусировки на важных объектах.
20	FireNet [35]	Точность 93.91–96.53%	Сеть состоит из 14 слоев, включая слой свертки с ReLU и выходной слой Softmax. Первый слой принимает на вход тензоры размером $(64 \times 64 \times 3)$, которые могут быть увеличены до $(128 \times 128 \times 3)$ без снижения частоты кадров. Последующие слои включают плотный слой с 256 нейронами и два плотных слоя с 128 нейронами.

21	FireNet-v2 [36]	Точность 94.95–98.43%	Модификация FireNet, в том числе уменьшение количества фильтров в первых трех конволюционных слоях до 15, 20 и 30 соответственно (в FireNet – 16, 32 и 64), что значительно уменьшает число обучаемых параметров. Функция активации в последнем конволюционном и обоих внутренних плотных слоях – сигмоидальная.
22	NASNet-A-OnFire [43] и ShuffleNetV2-OnFire [45]	Точность 95–97%	Используются специальные фильтры, предложенные в работе [46].
23	MVMNet [47]	mAP50 88.05%	Метод мультиориентированного обнаружения объектов на основе модуля VAM и смешанной NMS. Используется softpool-объединение пространственных пирамид для сохранения характеристик. Применяется комбинация гибридных не максимальных методов DIoU-NMS и Skew-NMS с двумя порогами для расстояний и углов для предотвращения ложного и пропущенного обнаружения.
24	3D-CHC с расширенным региональным контекстом [48]	См. таблицы 11–12	Объединяет информативные признаки из нескольких двумерных изображений для эффективного использования контекстной информации.

Таблица 16. Сводная таблица решений, рассмотренных в разделе «Способы трекинга неригидных объектов»

№	Решение	Эффективность	Комментарий
1	Трекинг с помощью глубоких многомасштабных пространственно-временных дискриминантных карт салиентности [49]	Точность 84.2%	Используется мультирегиональный механизм для генерации карт локальной салиентности, учитывающий визуальное восприятие с различными пространственными схемами и вариациями масштаба.
2	Трекинг с помощью деформируемых патчей [52]	F-мера 56.76%	Предлагается структура для отслеживания объектов с использованием деформируемых патчей и ядерного корреляционного фильтра с сохранением формы для учета формы цели. Адаптация патчей к сложной топологии и изменениям целей обеспечивается за счет фотометрической дискриминации и вариативности формы.
3	Трекинг с использованием параметрического активного контура и модели распределения точек [58]	—	Метод отслеживания неригидных объектов на изображениях с загроможденным фоном в многолюдных сценах. Предлагается использовать параметрические активные контуры (PAC) и модели распределения точек (PDM).
4	Трекер на основе многообразий распределения точек [59]	—	Модель опирается на набор геометрических фигур, которые могут быть адаптированы к изменениям в целевых объектах.
5	Алгоритм трехмерной реконструкции с использованием камеры глубины [60]	—	Входные данные: RGB-изображение, карта глубины и облако точек. Алгоритм находит общие точки между объектами, используя вычисляемые дескрипторы.

Выбор представленных решений при написании аналитического обзора обусловлен наличием качественных улучшений в области исследуемой задачи. Главный критерий, по которому то или иное решение можно назвать перспективным, – рост показателей эффективности от добавления новых слоев, модификации существующих или в целом от применения совершенно новой архитектуры. Другой критерий – высокое абсолютное значение ключевых показателей качества.

Исходя из данных, приведенных в обзоре, и тех же данных, структурированных в таблицах 15–16, следует отметить, что все архитектуры, модели, методы и алгоритмы, рассмотренные в данном аналитическом обзоре, в целом можно считать перспективными для использования. В представленных обобщающих таблицах имеются решения, для которых отсутствуют данные о четком числовом значении эффективности. Несмотря на это, включение их в настоящую работу целесообразно, так как в дальнейшем предлагается провести для этих решений дополнительные исследования на собственных наборах данных. К таким работам относятся решения из таблицы 15 – строки 4–9; из таблицы 16 – строки 3–5.

Заключение

По результатам проведенной работы над аналитическим обзором, для создания эффективных инструментов, позволяющих осуществлять обнаружение неригидных объектов в видеопотоке, потребуется объединение ряда решений из рассмотренных научных работ.

Перспективным для локализации неригидных объектов, на наш взгляд, выглядит следующий ряд архитектур, моделей, методов и алгоритмов: архитектура Fire-RPG [6]; нейросетевой алгоритм GSSD⁺⁺ [28]; модель DEGPR [30]; архитектура RCS-YOLO [31]; архитектура BGF-YOLO [32]; архитектура FireNet-v2 [36]; архитектуры NASNet-A-OnFire и ShuffleNetV2-OnFire [43]. Наиболее перспективные решения для трекинга: метод на основе многомасштабных пространственно-временных дискриминантных карт салиентности [49] и метод, базирующийся на применении деформируемых патчей [52].

Список использованных источников

- [1] Ergasheva A., Akhmedov F., Abdusalomov A., Kim W. *Advancing maritime safety: early detection of ship fires through computer vision, deep learning approaches, and histogram equalization techniques* // Fire. – 2024. – Vol. 7. – No. 3. – id. 84. – 15 pp.  [↑113](#), 133
- [2] Farkhod A., Abdusalomov A., Makhmudov F., Cho Y. I. *LDA-based topic modeling sentiment analysis using topic/document/sentence (TDS)* // Applied Sciences. – 2021. – Vol. 11. – No. 23. – id. 11091. – 15 pp.  [↑113](#)

- [3] Xu F., Zhang X., Deng T., Xu W. *An image-based fire monitoring algorithm resistant to fire-like objects* // Fire.– 2024.– Vol. 7.– No. 1.– id. 3.– 12 pp. [↑113, 133](#)
- [4] Woo S., Park J., Lee J.-Y. *CBAM: convolutional block attention module.*– 2018.– 17 c. arXiv: 1807.06521v2~[cs.CV] [↑113, 114](#)
- [5] Li G., Chen P., Xu C., Sun C., Ma Y. *Anchor-free smoke and flame recognition algorithm with multi-loss* // Fire.– 2023.– Vol. 6.– No. 6.– id. 225.– 16 pp. [↑113, 133](#)
- [6] Li X., Liang Y. *Fire-RPG: an urban fire detection network providing warnings in advance* // Fire.– 2024.– Vol. 7.– No. 7.– id. 214.– 22 pp. [↑114, 133, 138](#)
- [7] Ding X., Zhang X., Ma N., Han J., Ding G., Sun J. *RepVGG: Making VGG-style ConvNets great again.*– 2021.– 10 pp. arXiv: 2101.03697~[cs.CV] [↑114](#)
- [8] Tang Y., Han K., Guo J., Xu C., Xu C., Wang Y. *GhostNetV2: enhance cheap operation with long-range attention.*– 2022.– 12 pp. arXiv: 2211.12905~[cs.CV] [↑114](#)
- [9] Zhang Q. L., Yang Y. B. *SA-Net: shuffle attention for deep convolutional neural networks.*– 2021.– 9 pp. arXiv: 2102.00240~[cs.CV] [↑116](#)
- [10] Wang Q., Wu B., P. Zhu, P. Li, W. Zuo, Hu Q. *ECA-Net: efficient channel attention for deep convolutional neural Networks.*– 2020.– 12 pp. arXiv: 1910.03151v4~[cs.CV] [↑116](#)
- [11] Yang L., Zhang R. Y., Li L., Xie X. *Simple attention module based speaker verification with iterative noisy label detection.*– 2021.– 5 pp. arXiv: 2110.06534~[cs.CV] [↑116](#)
- [12] Xie J., Zhao H. *Forest fire object detection analysis based on knowledge distillation* // Fire.– 2023.– Vol. 6.– No. 12.– id. 446.– 15 pp. [↑117, 133](#)
- [13] Jin C., Wang T., Alhusaini N., Zhao S., Liu H., Xu K., Zhang J. *Video fire detection methods based on deep learning: datasets, methods, and future directions* // Fire.– 2023.– Vol. 6.– No. 8.– id. 315.– 27 pp. [↑118, 119](#)
- [14] Yuan F., Zhang L., Wan B., Xia X., Shi J. *Convolutional neural networks based on multi-scale additive merging layers for visual smoke recognition* // Machine Vision and Applications.– 2019.– Vol. 30.– Pp. 345–358. [↑118, 133](#)
- [15] Muhammad K., Ahmad J., Lv Z., Bellavista P., Yang P., Baik S. W. *Efficient deep CNN-based fire detection and localization in video surveillance applications* // IEEE Transactions on Systems, Man, and Cybernetics: Systems.– 2019.– Vol. 49.– No. 7.– Pp. 1419–1434. [↑118, 126](#)
- [16] Iandola F. N., Han S., Moskewicz M. W., Ashraf K., Dally W. J., Keutzer K. *SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size.*– 2016.– 13 pp. arXiv: 1602.07360~[cs.CV] [↑118, 133](#)
- [17] Khudayberdiev O., Zhang J., Abdullahi S. M., Zhang S. *Light-FireNet: an efficient lightweight network for fire detection in diverse environments* // Multimedia Tools and Applications.– 2022.– Vol. 81.– Pp. 24553–24572. [↑118, 133](#)
- [18] Zheng S., Gao P., Wang W., Zou X. *A highly accurate forest fire prediction model based on an improved dynamic convolutional neural network* // Applied Sciences.– 2022.– Vol. 12.– No. 13.– id. 6721.– 15 pp. [↑118, 134](#)
- [19] Tao H., Duan Q. *An adaptive frame selection network with enhanced dilated convolution for video smoke recognition* // Expert Systems with Applications.– 2023.– Vol. 215.– id. 119371.– 11 pp. [↑118, 134](#)

- [20] Khan Z. A., Hussain T., Ullah F. U. M., Gupta S. K., Lee M. Y., Baik S. W. *Randomly initialized CNN with densely connected stacked autoencoder for efficient fire detection* // Engineering Applications of Artificial Intelligence.– 2022.– Vol. **116**.– id. 105403.– 11 pp. doi ↑_{118, 134}
- [21] Hu C., Tang P., Jin W., He Z., Li W. *Real-time fire detection based on deep convolutional long-recurrent networks and optical flow method* // Proceedings of the 2018 37th Chinese Control Conference (CCC), CCC 2018 (Wuhan, China, 25–27 July, 2018).– IEEE.– 2018.– ISBN 978-1-538-64968-8.– Pp. 9061–9066. doi ↑_{119, 134}
- [22] Li S., Yan Q., Liu P. *An efficient fire detection method based on multiscale feature extraction, implicit deep supervision and channel attention mechanism* // IEEE Transactions on Image Processing.– 2020.– Vol. **29**.– Pp. 8467–8475. doi ↑_{119, 134}
- [23] Yang C., Pan Y., Cao Y., Lu X. *CNN-transformer hybrid architecture for early fire detection* // Proceedings of the Artificial Neural Networks and Machine Learning.– V. IV, ICANN 2022: 31st International Conference on Artificial Neural Networks (Bristol, UK, 6–9 September, 2022), Lecture Notes in Computer Science.– vol. **13532**, Berlin: Springer.– 2022.– ISBN 978-3-031-15936-7.– Pp. 570–581. doi ↑_{119, 134}
- [24] Wang X., Cai L., Zhou S., Jin Y., Tang L., Zhao Y. *Fire safety detection based on CAGSA-YOLO network* // Fire.– 2023.– Vol. **6**.– No. 8.– id. 297.– 19 pp. doi ↑_{119, 134}
- [25] Ding Z., Zhao Y., Li A., Zheng Z. *Spatial-temporal attention two-stream convolution neural network for smoke region detection* // Fire.– 2021.– Vol. **4**.– No. 4.– id. 66.– 12 pp. doi ↑_{120, 135}
- [26] Cao Y., Tang Q., Lu X., Li F., Cao J. *STCNet: spatio-temporal cross network for industrial smoke detection*.– 2020.– 10 c. arXiv:2011.04863v1[cs.CV] doi ↑₁₂₀
- [27] Shou Y., Meng T., Ai W., Xie C., Liu H., Wang Y. *Object detection in medical images based on hierarchical transformer and mask mechanism* // Computational Intelligence and Neuroscience.– 2022.– Vol. **2022**.– id. 5863782.– 12 pp. doi ↑_{122, 135}
- [28] Lee S.-G., Kim E., Bae J. S., Kim J. H., Yoon S. *Robust end-to-end focal liver lesion detection using unregistered multiphase computed tomography images* // IEEE Transactions on Emerging Topics in Computational Intelligence.– 2023.– Vol. **7**.– No. 2.– Pp. 319–329. doi ↑_{122, 135, 138}
- [29] De Frutos J. P., Pedersen A., Pelanis E., Bouget D., Survarachakan S., Langø T., Elle O.-J., Lindseth F. *Learning deep abdominal CT registration through adaptive loss weighting and synthetic data generation* // PLOS ONE.– 2023.– Vol. **18**.– No. 2.– Pp. 1–14. doi ↑₁₂₃
- [30] Tyagi A. K., Mohapatra C., Das P., Makharia G., Mehra L., AP P., Mausam *DeGPR: deep guided posterior regularization for multi-class cell detection and counting*.– 2023.– 11 c. arXiv:2304.00741v1[cs.CV] doi ↑_{123, 135, 138}
- [31] Kang M., Ting C.-M., Ting F. F., Phan R. C.-W. *RCS-YOLO: a fast and high-accuracy object detector for brain tumor detection*.– 2023.– 11 c. arXiv:2307.16412v2[cs.CV] doi ↑_{124, 135, 138}
- [32] Kang M., Ting C.-M., Ting F. F., Phan R. C.-W. *BGF-YOLO: enhanced YOLOv8 with multiscale attentional feature fusion for brain tumor detection*.– 2023.– 5 c. arXiv:2309.12585v2[cs.CV] doi ↑_{125, 135, 138}

- [33] Xu X., Jiang Y., Chen W., Huang Y., Zhang Y., Sun X. *DAMO-YOLO : a report on real-time object detection design.*— 2023.— 10 с. arXiv:2211.15444v4[cs.CV] doi ↑¹²⁵
- [34] Kang M., Ting C.-M., Ting F. F., Phan R. C.-W. *RCS-YOLO: a fast and high-accuracy object detector for brain tumor detection.*— 2023.— 11 pp. arXiv:2307.16412v2[cs.CV] doi ↑¹²⁵
- [35] Jadon A., Omama M., Varshney A., Ansari M. S., Sharma R. *FireNet: a specialized lightweight fire & smoke detection model for real-time IoT applications.*— 2019.— 6 pp. arXiv:1905.11922v2[cs.CV] doi ↑^{125, 127, 135}
- [36] Shees A., Ansari M. S., Varshney A., Asghar M. N., Kanwal N. *FireNet-v2: improved lightweight fire detection model for real-time IoT applications* // *Procedia Computer Science.*— 2023.— Vol. **218.**— Pp. 2233–2242. doi ↑^{126, 136, 138}
- [37] Altowaijri A. H., Alfaifi M. S., Alshawi T. A., Ibrahim A. B., Alshebeili S. A. *A privacy-preserving IoT-Based fire detector* // *IEEE Access.*— 2021.— Vol. **9.**— Pp. 51393–51402. doi ↑¹²⁶
- [38] Valikhujaev Y., Abdusalomov A., Cho Y. I. *Automatic fire and smoke detection method for surveillance systems based on dilated CNNs* // *Atmosphere.*— 2020.— Vol. **11.**— No. 11.— id. 1241.— 15 pp. doi ↑¹²⁶
- [39] Muhammad K., Ahmad J., Mehmood I., Rho S., Baik S. W. *Convolutional neural networks based fire detection in surveillance videos* // *IEEE Access.*— 2018.— Vol. **6.**— Pp. 18174–18183. doi ↑¹²⁶
- [40] Saponara S., Elhanashi A., Gagliardi A. *Real-time video fire/smoke detection based on CNN in antifire surveillance systems* // *Journal of Real-Time Image Processing.*— 2021.— Vol. **18.**— Pp. 889–900. doi ↑¹²⁷
- [41] Ayala A., Lima E., Fernandes B., Bezerra B. L., Cruz F. *Lightweight and efficient octave convolutional neural network for fire recognition* // *Proceedings of the 2019 IEEE Latin American Conference on Computational Intelligence, LA-CCI'2019* (Guayaquil, Ecuador, 11–15 November, 2019).— IEEE.— 2019.— ISBN 978-1-7281-5666-8.— 6 pp. doi ↑¹²⁷
- [42] Saponara S., Elhanashi A., Gagliardi A. *Exploiting R-CNN for video smoke/fire sensing in antifire surveillance indoor and outdoor systems for smart cities* // *Proceedings of the 2020 IEEE International Conference on Smart Computing, SMARTCOMP'2020* (Bologna, Italy, 14–17 September, 2020).— IEEE.— 2020.— ISBN 978-1-7281-6997-2.— Pp. 392–397. doi ↑¹²⁷
- [43] Thomson W., Bhowmik N., Breckon T. P. *Efficient and compact convolutional neural network architectures for non-temporal real-time fire detection.*— 2020.— 6 pp. arXiv:2010.08833[cs.CV] doi ↑^{127, 136, 138}
- [44] Zoph B., Vasudevan V., Shlens J., Le Q. V. *Learning transferable architectures for scalable image recognition* // *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), CVPR'18* (Salt Lake City, Utah, 18–22 June, 2018).— IEEE.— 2018.— ISBN 978-1-728-13294-5.— Pp. 8697–8710. doi ↑¹²⁷

- [45] Ma N., Zhang X., Zheng H.-T., Sun J. *Shufflenet v2: practical guidelines for efficient CNN architecture design* // *Proceedings of the 2018 European Conference on Computer Vision (ECCV)*, ECCV'18 (Munich, Germany, 8–14 September, 2018), Lecture Notes in Computer Science.– vol. **11218**, Cham: Springer.– 2018.– ISBN 978-3-030-01263-2.– Pp. 122–138.  [↑127, 136](#)
- [46] Li H., Kadav A., Durdanovic I., Samet H., Graf H. P. *Pruning filters for efficient ConvNets*.– 2017.– 13 pp. arXiv: 1608.08710v3[cs.CV]  [↑127, 136](#)
- [47] Hu Y., Zhan J., Zhou G., Chen A., Cai W., Guo K., Hu Y., Li L. *Fast forest fire smoke detection using MVMNet* // *Knowledge-Based Systems*.– 2022.– Vol. **241**.– 20 pp.  [↑128, 136](#)
- [48] Yan K., Bagheri M., Summers R. M. *3D context enhanced region-based convolutional neural network for end-to-end lesion detection*.– 2018.– 11 pp. arXiv: 1806.09648v2[cs.CV]  [↑128, 136](#)
- [49] Zhang P., Liu W., Wang D., Lei Y., Wang H., Shen C., Lu H. *Non-rigid object tracking via deep multi-scale spatial-temporal discriminative saliency maps*.– 2019.– 12 pp. arXiv: 1802.07957v2[cs.CV]  [↑129, 137, 138](#)
- [50] Hong S., You T., Kwak S., Han B. *Online tracking by learning discriminative saliency map with convolutional neural network* // *Proceedings of the 32nd International Conference on Machine Learning, ICML'15* (Lille, France, 6–11 July, 2015), PMLR.– vol. **37**.– 2015.– ISBN 978-1-510-81058-7.– Pp. 597–606.  [↑](#)
- [51] Son J., Jung I., Park K., Han B. *Tracking-by-segmentation with online gradient boosting decision tree* // *Proceedings of the 2015 IEEE International Conference on Computer Vision, ICCV'15* (Santiago, Chile, 07–13 December, 2015).– IEEE.– 2015.– ISBN 978-1-4673-8391-2.– Pp. 3056–3064.  [↑130](#)
- [52] Sun X., Cheung N.-M., Yao H., Guo Y. *Non-rigid object tracking via deformable patches using shape-preserved KCF and level sets* // *Proceedings of the 2017 IEEE International Conference on Computer Vision, ICCV'17* (Venice, Italy, 22–29 October, 2017).– IEEE.– 2017.– ISBN 978-1-5386-1032-9.– Pp. 5496–5504.  [↑130, 137, 138](#)
- [53] Duffner S., Garcia C. *PixelTrack: a fast adaptive algorithm for tracking non-rigid objects* // *Proceedings of the 2013 IEEE International Conference on Computer Vision, ICCV'13* (Sydney, NSW, Australia, 1–8 December, 2013).– IEEE.– 2013.– ISBN 978-1-4799-2840-8.– Pp. 2480–2487.  [↑130](#)
- [54] Sevilla-Lara L., Learned-Miller E. *Distribution fields for tracking* // *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, CVPR'12* (Providence, RI, USA, 16–21 June, 2012).– 2012.– ISBN 978-1-4673-1226-4.– Pp. 1910–1917.  [↑130](#)
- [55] Godec M., Roth P. M., Bischof H. *Hough-based tracking of non-rigid objects* // *Proceedings of the 2011 IEEE International Conference on Computer Vision, ICCV'11* (Barcelona, Spain, 06–13 November, 2011).– 2011.– ISBN 978-1-4577-1101-5.– Pp. 81–88.  [↑130](#)
- [56] Sun X., Yao H., Zhang S., Li D. *Non-rigid object contour tracking via a novel supervised level set model* // *IEEE Transactions on Image Processing*.– 2015.– Vol. **24**.– No. 11.– Pp. 3386–3399.  [↑130](#)

- [57] Li Y., Zhu J., Hoi S. *Reliable Patch Trackers: Robust visual tracking by exploiting reliable patches* // *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR'15 (Boston, MA, USA, 07–12 June, 2015).– IEEE.– 2015.– ISBN 978-1-4673-6964-0.– Pp. 353–361. doi ↑130
- [58] Olszewska J. I., Mathes T., Vleeschouwer C. D., Piater J., Macq B. *Non-rigid object tracker based on a robust combination of parametric active contour and point distribution model*, Visual Communications and Image Processing 2007 (San Jose, CA, USA, 28 January–1 February, 2007), Proc. SPIE.– vol. **6508**.– 2007.– ISBN 978-0-8194-6621-1.– id. 65082A.– 8 pp. doi URL ↑130, 137
- [59] Mathes T., Piater J. *Robust non-rigid object tracking using point distribution manifolds* // *Pattern Recognition*, Lecture Notes in Computer Science.– vol. **4174**, Berlin–Heidelberg: Springer.– 2006.– ISBN 978-3-540-44414-5.– Pp. 515–524. doi ↑131, 137
- [60] Руиз-Родригез М., Кобер В. И., Карнаухов В. Н., Мозеров М. Г. *Алгоритм трехмерной реконструкции нежестких объектов с использованием камеры глубины* // *Информационные процессы*.– 2019.– Т. **19**.– № 4.– С. 388–398. URL ↑132, 137
- [61] Sipiran I., Bustos B. H. *Harris 3D: a robust extension of the harris operator for interest point detection on 3D meshes* // *The Visual Computer*.– 2011.– Vol. **27**.– No. 11.– Pp. 963–976. doi ↑132
- [62] Zhong Y. *Intrinsic shape signatures: A shape descriptor for 3D object recognition* // *Proceedings of the 2009 IEEE Conference on Computer Vision Workshops*, ICCVW'09 (Kyoto, Japan, 27 September–4 October, 2009).– IEEE.– 2009.– ISBN 978-1-4244-4442-7.– Pp. 689–696. doi ↑132
- [63] Smith S. M., Brady J. M. *SUSAN — a new approach to low level image processing* // *International Journal of Computer Vision*.– 1997.– Vol. **23**.– No. 1.– Pp. 45–78. doi ↑132
- [64] Lowe D. G. *Distinctive image features from scale-invariant keypoints* // *International Journal of Computer Vision*.– 2004.– Vol. **60**.– No. 2.– Pp. 91–110. doi ↑132
- [65] Rusu R. B., Marton Z. C., Blodow N., Beetz M. *Persistent point feature histograms for 3D point clouds* // *Proceedings of the 10th International Conference on Intelligent Autonomous Systems*, IAS-10 (Baden-Baden, Germany, 23–25 July, 2008).– IOS Press.– 2008.– ISBN 978-1-58603-887-8.– Pp. 119–128. doi ↑132
- [66] Tombari F., Salti S., Stefano L. D. *Unique signatures of histograms for local surface description* // *Proceedings of the 2010 European Conference on Computer Vision*, ECCV'10 (Crete, Greece, 5–11 September, 2010), Lecture Notes in Computer Science.– vol. **6313**, Berlin–Heidelberg: Springer.– 2010.– ISBN 978-3-642-15557-4.– Pp. 356–369. doi ↑132
- [67] Frome A., Huber D., Kolluri R., Bulow T., Malik J. *Recognizing objects in range data using regional point descriptors* // *Proceedings of the 2004 European Conference on Computer Vision*, ECCV'04 (Prague, Czech Republic, 11–14 May, 2004), Berlin–Heidelberg: Springer.– 2004.– ISBN 978-3-540-21982-8.– Pp. 224–237. doi ↑132

- [68] Lazebnik S., Schmid C., Ponce J. *A sparse texture representation using local affine regions* // IEEE Transactions on Pattern Analysis and Machine Intelligence.– 2005.– Vol. **27**.– No. 8.– Pp. 1265–1278. doi ↑132
- [69] Marton Z. C., Pangercic D., Blodow N., Kleinhellefort J., Beetz M. *General 3D modelling of novel objects from a single view* // Proceedings of the 2010 IEEE/RSJ Conference on Intelligent Robots and Systems, IROS'10 (Taipei, Taiwan, 18–22 October, 2010).– IEEE.– 2010.– ISBN 978-1-4244-6674-0.– Pp. 3700–3705. doi ↑132
- [70] Sturm J., Engelhard N., Endres F., Burgard W., Cremers D. *A Benchmark for the evaluation of RGB-D SLAM systems* // Proceedings of the 2012 IEEE/RSJ Conference on Intelligent Robots and Systems (IROS), IROS'12 (Vilamoura-Algarve, Portugal, 7–12 October, 2012).– IEEE.– 2012.– ISBN 978-1-4673-1737-5.– Pp. 573–580. doi ↑132

Поступила в редакцию

08.10.2024;

одобрена после рецензирования

22.12.2024;

принята к публикации

22.12.2024;

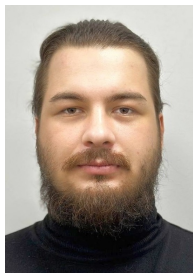
опубликована онлайн

26.12.2024.

Рекомендовал к публикации

д.т.н. В. М. Хачумов

Информация об авторах:



Григорий Глебович Гриценко

Аспирант ИПС им. А.К. Айламазяна РАН. Область научных интересов: локализация и трекинг неригидных объектов.



0009-0006-6838-9632

e-mail: GregorGre@mail.ru



Виталий Петрович Фраленко

Кандидат технических наук, ведущий научный сотрудник ИЦМС ИПС им. А.К. Айламазяна РАН. Область научных интересов: интеллектуальный анализ данных и распознавание образов, искусственный интеллект и принятие решений, параллельно-конвейерные вычисления, сетевая безопасность, диагностика сложных технических систем, графические интерфейсы.



0000-0003-0123-3773

e-mail: alarmod@pereslavl.ru

Все авторы сделали эквивалентный вклад в подготовку публикации.

Декларация об отсутствии личной заинтересованности: благополучие авторов не зависит от результатов исследования.



An analytical review of architectures, models, methods and algorithms for localization and tracking of non-rigid objects

Grigory Glebovich **Gricenko**¹, Vitaly Petrovich **Fralenko**²

^{1,2} Ailamazyan Program Systems Institute of RAS, Ves'kovo, Russia

Abstract. Computer vision requires video stream analysis, including extracting information from frames, detecting specific objects, and collecting data about them. After detection, *tracking* or following objects in the video stream is often required. *Non-rigidity* or shape variability hinders object analysis, complicates their detection and tracking, and worsens localization.

The review considers architectures, models, methods, and algorithms used in practice for detection and tracking of non-rigid objects, and highlights promising solutions. (*In Russian*).

Key words and phrases: non-rigid object, artificial neural network, deep learning, object localization, object tracking, fire and smoke detection, medical image analysis

2020 *Mathematics Subject Classification:* 68T45; 68T07

Acknowledgments: This work was financially supported by the *Russian Science Foundation*, project № 21-71-10056^{RF} and a grant in the form of a subsidy from the regional budget to organizations of the Yaroslavl region.

For citation: Grigory G. Gricenko, Vitaly P. Fralenko. *An analytical review of architectures, models, methods and algorithms for localization and tracking of non-rigid objects*. Program Systems: Theory and Applications, 2024, 15:4(63), pp. 111–151. (*In Russ.*). https://psta.psisras.ru/read/psta2024_4_111-151.pdf

References

- [1] A. Ergasheva, F. Akhmedov, A. Abdusalomov, W. Kim. “Advancing maritime safety: early detection of ship fires through computer vision, deep learning approaches, and histogram equalization techniques”, *Fire*, **7**:3 (2024), id. 84, 15 pp. 
- [2] A. Farkhod, A. Abdusalomov, F. Makhmudov, Y. I. Cho. “LDA-based topic modeling sentiment analysis using topic/document/sentence (TDS)”, *Applied Sciences*, **11**:23 (2021), id. 11091, 15 pp. 
- [3] F. Xu, X. Zhang, T. Deng, W. Xu. “An image-based fire monitoring algorithm resistant to fire-like objects”, *Fire*, **7**:1 (2024), id. 3, 12 pp. 
- [4] S. Woo, J. Park, J.-Y. Lee. *CBAM: convolutional block attention module*, 2018, 17 pp.  1807.06521v2 [cs.CV] 
- [5] G. Li, P. Chen, C. Xu, C. Sun, Y. Ma. “Anchor-free smoke and flame recognition algorithm with multi-loss”, *Fire*, **6**:6 (2023), id. 225, 16 pp. 
- [6] X. Li, Y. Liang. “Fire-RPG: an urban fire detection network providing warnings in advance”, *Fire*, **7**:7 (2024), id. 214, 22 pp. 
- [7] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, J. Sun. *RepVGG: Making VGG-style ConvNets great again*, 2021, 10 pp.  2101.03697 [cs.CV] 
- [8] Y. Tang, K. Han, J. Guo, C. Xu, C. Xu, Y. Wang. *GhostNetV2: enhance cheap operation with long-range attention*, 2022, 12 pp.  2211.12905 [cs.CV] 
- [9] Q. L. Zhang, Y. B. Yang. *SA-Net: shuffle attention for deep convolutional neural networks*, 2021, 9 pp.  2102.00240 [cs.CV] 
- [10] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, Q. Hu. *ECA-Net: efficient channel attention for deep convolutional neural Networks*, 2020, 12 pp.  1910.03151v4 [cs.CV] 
- [11] L. Yang, R. Y. Zhang, L. Li, X. Xie. *Simple attention module based speaker verification with iterative noisy label detection*, 2021, 5 pp.  2110.06534 [cs.CV] 
- [12] J. Xie, H. Zhao. “Forest fire object detection analysis based on knowledge distillation”, *Fire*, **6**:12 (2023), id. 446, 15 pp. 
- [13] C. Jin, T. Wang, N. Alhusaini, S. Zhao, H. Liu, K. Xu, J. Zhang. “Video fire detection methods based on deep learning: datasets, methods, and future directions”, *Fire*, **6**:8 (2023), id. 315, 27 pp. 
- [14] F. Yuan, L. Zhang, B. Wan, X. Xia, J. Shi. “Convolutional neural networks based on multi-scale additive merging layers for visual smoke recognition”, *Machine Vision and Applications*, **30** (2019), pp. 345–358. 
- [15] K. Muhammad, J. Ahmad, Z. Lv, P. Bellavista, P. Yang, S. W. Baik. “Efficient deep CNN-based fire detection and localization in video surveillance applications”, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, **49**:7 (2019), pp. 1419–1434. 
- [16] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, K. Keutzer. *SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size*, 2016, 13 pp.  1602.07360 [cs.CV] 
- [17] O. Khudayberdiev, J. Zhang, S. M. Abdullahi, S. Zhang. “Light-FireNet:

- an efficient lightweight network for fire detection in diverse environments”, *Multimedia Tools and Applications*, **81** (2022), pp. 24553–24572. [doi](#)
- [18] S. Zheng, P. Gao, W. Wang, X. Zou. “A highly accurate forest fire prediction model based on an improved dynamic convolutional neural network”, *Applied Sciences*, **12**:13 (2022), id. 6721, 15 pp. [doi](#)
 - [19] H. Tao, Q. Duan. “An adaptive frame selection network with enhanced dilated convolution for video smoke recognition”, *Expert Systems with Applications*, **215** (2023), id. 119371, 11 pp. [doi](#)
 - [20] Z. A. Khan, T. Hussain, F. U. M. Ullah, S. K. Gupta, M. Y. Lee, S. W. Baik. “Randomly initialized CNN with densely connected stacked autoencoder for efficient fire detection”, *Engineering Applications of Artificial Intelligence*, **116** (2022), id. 105403, 11 pp. [doi](#)
 - [21] C. Hu, P. Tang, W. Jin, Z. He, W. Li. “Real-time fire detection based on deep convolutional long-recurrent networks and optical flow method”, *Proceedings of the 2018 37th Chinese Control Conference (CCC)*, CCC 2018 (Wuhan, China, 25–27 July, 2018), IEEE, 2018, ISBN 978-1-538-64968-8, pp. 9061–9066. [doi](#)
 - [22] S. Li, Q. Yan, P. Liu. “An efficient fire detection method based on multiscale feature extraction, implicit deep supervision and channel attention mechanism”, *IEEE Transactions on Image Processing*, **29** (2020), pp. 8467–8475. [doi](#)
 - [23] C. Yang, Y. Pan, Y. Cao, X. Lu. “CNN-transformer hybrid architecture for early fire detection”, *Proceedings of the Artificial Neural Networks and Machine Learning. V. IV, ICANN 2022: 31st International Conference on Artificial Neural Networks* (Bristol, UK, 6–9 September, 2022), Lecture Notes in Computer Science, vol. **13532**, Springer, Berlin, 2022, ISBN 978-3-031-15936-7, pp. 570–581. [doi](#)
 - [24] X. Wang, L. Cai, S. Zhou, Y. Jin, L. Tang, Y. Zhao. “Fire safety detection based on CAGSA-YOLO network”, *Fire*, **6**:8 (2023), id. 297, 19 pp. [doi](#)
 - [25] Z. Ding, Y. Zhao, A. Li, Z. Zheng. “Spatial-temporal attention two-stream convolution neural network for smoke region detection”, *Fire*, **4**:4 (2021), id. 66, 12 pp. [doi](#)
 - [26] Y. Cao, Q. Tang, X. Lu, F. Li, J. Cao. *STCNet: spatio-temporal cross network for industrial smoke detection*, 2020, 10 pp. [arXiv:2011.04863 \[cs.CV\]](#) [doi](#)
 - [27] Y. Shou, T. Meng, W. Ai, C. Xie, H. Liu, Y. Wang. “Object detection in medical images based on hierarchical transformer and mask mechanism”, *Computational Intelligence and Neuroscience*, **2022** (2022), id. 5863782, 12 pp. [doi](#)
 - [28] S.-G. Lee, E. Kim, J. S. Bae, J. H. Kim, S. Yoon. “Robust end-to-end focal liver lesion detection using unregistered multiphase computed tomography images”, *IEEE Transactions on Emerging Topics in Computational Intelligence*, **7**:2 (2023), pp. 319–329. [doi](#)
 - [29] de Frutos J. P., A. Pedersen, E. Pelanis, D. Bouget, S. Survarachakan, S. Survarachakan, O.-J. Elle, F. Lindseth. “Learning deep abdominal CT registration through adaptive loss weighting and synthetic data generation”, *PLOS ONE*, **18**:2 (2023), pp. 1–14. [doi](#)

- [30] A. K. Tyagi, C. Mohapatra, P. Das, G. Makharia, L. Mehra, P. AP, P. AP. *DeGPR: deep guided posterior regularization for multi-class cell detection and counting*, 2023, 11 pp. [arXiv:2304.00741](#) [cs.CV] [doi](#)
- [31] M. Kang, C.-M. Ting, F. F. Ting, R. C.-W. Phan. *RCS-YOLO: a fast and high-accuracy object detector for brain tumor detection*, 2023, 11 pp. [arXiv:2307.16412v2](#) [cs.CV] [doi](#)
- [32] M. Kang, C.-M. Ting, F. F. Ting, R. C.-W. Phan. *BGF-YOLO: enhanced YOLOv8 with multiscale attentional feature fusion for brain tumor detection*, 2023, 5 pp. [arXiv:2309.12585v2](#) [cs.CV] [doi](#)
- [33] X. Xu, Y. Jiang, W. Chen, Y. Huang, Y. Zhang, X. Sun. *DAMO-YOLO: a report on real-time object detection design*, 2023, 10 pp. [arXiv:2211.15444v4](#) [cs.CV] [doi](#)
- [34] M. Kang, C.-M. Ting, F. F. Ting, R. C.-W. Phan. *RCS-YOLO: a fast and high-accuracy object detector for brain tumor detection*, 2023, 11 pp. [arXiv:2307.16412v2](#) [cs.CV] [doi](#)
- [35] A. Jadon, M. Omama, A. Varshney, M. S. Ansari, R. Sharma. *FireNet: a specialized lightweight fire & smoke detection model for real-time IoT applications*, 2019, 6 pp. [arXiv:1905.11922v2](#) [cs.CV] [doi](#)
- [36] A. Shees, M. S. Ansari, A. Varshney, M. N. Asghar, N. Kanwal. “FireNet-v2: improved lightweight fire detection model for real-time IoT applications”, *Procedia Computer Science*, **218** (2023), pp. 2233–2242. [doi](#)
- [37] A. H. Altowaijri, M. S. Alfaifi, T. A. Alshawi, A. B. Ibrahim, S. A. Alshebeili. “A privacy-preserving IoT-Based fire detector”, *IEEE Access*, **9** (2021), pp. 51393–51402. [doi](#)
- [38] Y. Valikhujaev, A. Abdusalomov, Y. I. Cho. “Automatic fire and smoke detection method for surveillance systems based on dilated CNNs”, *Atmosphere*, **11**:11 (2020), id. 1241, 15 pp. [doi](#)
- [39] K. Muhammad, J. Ahmad, I. Mehmood, S. Rho, S. W. Baik. “Convolutional neural networks based fire detection in surveillance videos”, *IEEE Access*, **6** (2018), pp. 18174–18183. [doi](#)
- [40] S. Saponara, A. Elhanashi, A. Gagliardi. “Real-time video fire/smoke detection based on CNN in antifire surveillance systems”, *Journal of Real-Time Image Processing*, **18** (2021), pp. 889–900. [doi](#)
- [41] A. Ayala, E. Lima, B. Fernandes, B. L. Bezerra, F. Cruz. “Lightweight and efficient octave convolutional neural network for fire recognition”, *Proceedings of the 2019 IEEE Latin American Conference on Computational Intelligence*, LA-CCI’2019 (Guayaquil, Ecuador, 11–15 November, 2019), IEEE, 2019, ISBN 978-1-7281-5666-8, 6 pp. [doi](#)
- [42] S. Saponara, A. Elhanashi, A. Gagliardi. “Exploiting R-CNN for video smoke/fire sensing in antifire surveillance indoor and outdoor systems for smart cities”, *Proceedings of the 2020 IEEE International Conference on Smart Computing*, SMARTCOMP’2020 (Bologna, Italy, 14–17 September, 2020), IEEE, 2020, ISBN 978-1-7281-6997-2, pp. 392–397. [doi](#)
- [43] W. Thomson, N. Bhowmik, T. P. Breckon. *Efficient and compact convolutional neural network architectures for non-temporal real-time fire detection*, 2020, 6 pp. [arXiv:2010.08833](#) [cs.CV] [doi](#)

- [44] B. Zoph, V. Vasudevan, J. Shlens, Q. V. Le. “Learning transferable architectures for scalable image recognition”, *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, CVPR’18 (Salt Lake City, Utah, 18–22 June, 2018), IEEE, 2018, ISBN 978-1-728-13294-5, pp. 8697–8710. 
- [45] N. Ma, X. Zhang, H.-T. Zheng, J. Sun. “Shufflenet v2: practical guidelines for efficient CNN architecture design”, *Proceedings of the 2018 European Conference on Computer Vision (ECCV)*, ECCV’18 (Munich, Germany, 8–14 September, 2018), Lecture Notes in Computer Science, vol. **11218**, Springer, Cham, 2018, ISBN 978-3-030-01263-2, pp. 122–138. 
- [46] H. Li, A. Kadav, I. Durdanovic, H. Samet, H. P. Graf. *Pruning filters for efficient ConvNets*, 2017, 13 pp. [arXiv:1608.08710v3](https://arxiv.org/abs/1608.08710v3) [cs.CV] 
- [47] Y. Hu, J. Zhan, G. Zhou, A. Chen, W. Cai, K. Guo, Y. Hu, L. Li. “Fast forest fire smoke detection using MVMNet”, *Knowledge-Based Systems*, **241** (2022), 20 pp. 
- [48] K. Yan, M. Bagheri, R. M. Summers. *3D context enhanced region-based convolutional neural network for end-to-end lesion detection*, 2018, 11 pp. [arXiv:1806.09648v2](https://arxiv.org/abs/1806.09648v2) [cs.CV] 
- [49] P. Zhang, W. Liu, D. Wang, Y. Lei, H. Wang, C. Shen, H. Lu. *Non-rigid object tracking via deep multi-scale spatial-temporal discriminative saliency maps*, 2019, 12 pp. [arXiv:1802.07957v2](https://arxiv.org/abs/1802.07957v2) [cs.CV] 
- [50] S. Hong, T. You, S. Kwak, B. Han. “Online tracking by learning discriminative saliency map with convolutional neural network”, *Proceedings of the 32nd International Conference on Machine Learning, ICML’15* (Lille, France, 6–11 July, 2015), PMLR, vol. **37**, 2015, ISBN 978-1-510-81058-7, pp. 597–606. 
- [51] J. Son, I. Jung, K. Park, B. Han. “Tracking-by-segmentation with online gradient boosting decision tree”, *Proceedings of the 2015 IEEE International Conference on Computer Vision, ICCV’15* (Santiago, Chile, 07–13 December, 2015), IEEE, 2015, ISBN 978-1-4673-8391-2, pp. 3056–3064. 
- [52] X. Sun, N.-M. Cheung, H. Yao, Y. Guo. “Non-rigid object tracking via deformable patches using shape-preserved KCF and level sets”, *Proceedings of the 2017 IEEE International Conference on Computer Vision, ICCV’17* (Venice, Italy, 22–29 October, 2017), IEEE, 2017, ISBN 978-1-5386-1032-9, pp. 5496–5504. 
- [53] S. Duffner, C. Garcia. “PixelTrack: a fast adaptive algorithm for tracking non-rigid objects”, *Proceedings of the 2013 IEEE International Conference on Computer Vision, ICCV’13* (Sydney, NSW, Australia, 1–8 December, 2013), IEEE, 2013, ISBN 978-1-4799-2840-8, pp. 2480–2487. 
- [54] L. Sevilla-Lara, E. Learned-Miller. “Distribution fields for tracking”, *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, CVPR’12* (Providence, RI, USA, 16–21 June, 2012), 2012, ISBN 978-1-4673-1226-4, pp. 1910–1917. 
- [55] M. Godec, P. M. Roth, H. Bischof. “Hough-based tracking of non-rigid objects”, *Proceedings of the 2011 IEEE International Conference on Computer*

- Vision*, ICCV'11 (Barcelona, Spain, 06–13 November, 2011), 2011, ISBN 978-1-4577-1101-5, pp. 81–88. [doi](#)
- [56] X. Sun, H. Yao, S. Zhang, D. Li. “Non-rigid object contour tracking via a novel supervised level set model”, *IEEE Transactions on Image Processing*, **24**:11 (2015), pp. 3386–3399. [doi](#)
- [57] Y. Li, J. Zhu, S. Hoi. “Reliable Patch Trackers: Robust visual tracking by exploiting reliable patches”, *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition*, CVPR'15 (Boston, MA, USA, 07–12 June, 2015), IEEE, 2015, ISBN 978-1-4673-6964-0, pp. 353–361. [doi](#)
- [58] J. I. Olszewska, T. Mathes, C. D. Vleeschouwer, J. Piater, B. Macq. “Non-rigid object tracker based on a robust combination of parametric active contour and point distribution model”, *Visual Communications and Image Processing 2007* (San Jose, CA, USA, 28 January–1 February, 2007), Proc. SPIE, vol. **6508**, 2007, ISBN 978-0-8194-6621-1, id. 65082A, 8 pp. [doi](#) [URL](#)
- [59] T. Mathes, J. Piater. “Robust non-rigid object tracking using point distribution manifolds”, *Pattern Recognition*, Lecture Notes in Computer Science, vol. **4174**, Springer, Berlin–Heidelberg, 2006, ISBN 978-3-540-44414-5, pp. 515–524. [doi](#)
- [60] M. Ruiz-Rodríguez, V. I. Kober, V. N. Karnauxov, M. G. Mozerov. “Algorithm for three-dimensional reconstruction of non-rigid objects using a depth camera”, *Informacionnye processy*, **19**:4 (2019), pp. 388–398 (in Russian). [URL](#)
- [61] I. Sipiran, B. H. Bustos. “Harris 3D: a robust extension of the harris operator for interest point detection on 3D meshes”, *The Visual Computer*, **27**:11 (2011), pp. 963–976. [doi](#)
- [62] Y. Zhong. “Intrinsic shape signatures: A shape descriptor for 3D object recognition”, *Proceedings of the 2009 IEEE Conference on Computer Vision Workshops*, ICCVW'09 (Kyoto, Japan, 27 September–4 October, 2009), IEEE, 2009, ISBN 978-1-4244-4442-7, pp. 689–696. [doi](#)
- [63] S. M. Smith, J. M. Brady. “SUSAN — a new approach to low level image processing”, *International Journal of Computer Vision*, **23**:1 (1997), pp. 45–78. [doi](#)
- [64] D. G. Lowe. “Distinctive image features from scale-invariant keypoints”, *International Journal of Computer Vision*, **60**:2 (2004), pp. 91–110. [doi](#)
- [65] R. B. Rusu, Z. C. Marton, N. Blodow, M. Beetz. “Persistent point feature histograms for 3D point clouds”, *Proceedings of the 10th International Conference on Intelligent Autonomous Systems*, IAS-10 (Baden-Baden, Germany, 23–25 July, 2008), IOS Press, 2008, ISBN 978-1-58603-887-8, pp. 119–128. [doi](#)
- [66] F. Tombari, S. Salti, L. D. Stefano. “Unique signatures of histograms for local surface description”, *Proceedings of the 2010 European Conference on Computer Vision*, ECCV'10 (Crete, Greece, 5–11 September, 2010), Lecture Notes in Computer Science, vol. **6313**, Springer, Berlin–Heidelberg, 2010, ISBN 978-3-642-15557-4, pp. 356–369. [doi](#)
- [67] A. Frome, D. Huber, R. Kolluri, T. Bulow, J. Malik. “Recognizing objects in range data using regional point descriptors”, *Proceedings of the 2004 European Conference on Computer Vision*, ECCV'04 (Prague, Czech Republic, 11–14

- May, 2004), Springer, Berlin–Heidelberg, 2004, ISBN 978-3-540-21982-8, pp. 224–237. 
- [68] S. Lazebnik, C. Schmid, J. Ponce. “A sparse texture representation using local affine regions”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**:8 (2005), pp. 1265–1278. 
- [69] Z. C. Marton, D. Pangercic, N. Blodow, J. Kleinhellefort, M. Beetz. “General 3D modelling of novel objects from a single view”, *Proceedings of the 2010 IEEE/RSJ Conference on Intelligent Robots and Systems, IROS’10* (Taipei, Taiwan, 18–22 October, 2010), IEEE, 2010, ISBN 978-1-4244-6674-0, pp. 3700–3705. 
- [70] J. Sturm, N. Engelhard, F. Endres, W. Burgard, D. Cremers. “A Benchmark for the evaluation of RGB-D SLAM systems”, *Proceedings of the 2012 IEEE/RSJ Conference on Intelligent Robots and Systems (IROS)*, IROS’12 (Vilamoura-Algarve, Portugal, 7–12 October, 2012), IEEE, 2012, ISBN 978-1-4673-1737-5, pp. 573–580. 