

УДК 004.852

doi: 10.15622/rcai.2025.079

## РАСПОЗНАВАНИЕ ИЗОБРАЖЕНИЙ ЛЮДЕЙ ПО ДАННЫМ С ИНФРАКРАСНОЙ И ЦВЕТНОЙ КАМЕРЫ С ИСПОЛЬЗОВАНИЕМ СЕГМЕНТНОЙ СПАЙКОВОЙ НЕЙРОННОЙ СЕТИ<sup>1</sup>

И.С. Фомин (*i.fomin@rtc.ru*)

ГНЦ РФ Центральный научно-исследовательский  
и опытно-конструкторский институт робототехники и технической  
кибернетики, Санкт-Петербург

В работе рассматривается вопрос обработки изображений, получаемых при различном режиме работы камеры видеонаблюдения. В качестве примера рассматривается распознавание людей на подмножестве общедоступного набора данных. Показан алгоритм обучения сиамской нейронной сети для преобразования изображений ночной и дневной съемки в векторное пространство представлений. Приведен способ преобразования векторного представления в спайковое представление. Описан алгоритм классификации представлений инфракрасных и RGB изображений по отдельности, а также их преобразования в общий набор и совместной обработки. Проведенные экспериментальные исследования показывают высокое качество предложенных подходов. Качество автономной обработки составило 0,891 для ИК изображений, 0,965 для RGB изображений, совместной обработки – 0,957.

**Ключевые слова:** реидентификация людей, сверточные сети, сиамские сети, спайковые сети, нейронные сети, инфракрасные изображения, цветные изображения.

### Введение

Обнаружение и распознавание объектов в системах видеонаблюдения давно известная и хорошо решаемая задача технического зрения, многие готовые решения используются в различных коммерческих системах

---

<sup>1</sup> Результаты получены в рамках выполнения государственного задания Минобрнауки России «Исследование методов анализа слабоструктурированных данных, обработки знаний и создания когнитивных агентов на базе комбинированных глубоких нейронных сетей» (FNRG-2025-0008 1024050200009-5-1.2.1;2.2.2).

обеспечения безопасности. Обнаружение людей (наличия людей) на изображении в том числе включено и в некоторые готовые решения, где это используется как один из параметров, регулирующих начало и остановку записи. Однако даже так остается ряд вопросов, который освещен недостаточно и нередко возникает в научных и практических работах. К таким задачам стоит отнести вопросы реидентификации людей при сопровождении между несколькими камерами.

Ранее был предложен универсальный подход к реидентификации людей на основе комбинации сиамской сверточной и сегментной спайковой нейронных сетей. Сиамская сеть после обучения специальным образом отвечает за преобразование четырехмерного тензора данных, полученного с цветной камеры или инфракрасной камеры в пространство представлений. Спайковая сеть для классификации, собранная из нейронов сегментной спайковой модели [Бахшиев и др., 2021], отвечает за определение того, какому из предъявленных при обучении спайковых представлений, полученных из тензоров данных людей соответствует представление, полученное из изучаемого тензора. Ранее уже было показано [Корсаков и др., 2022], что сегментная спайковая сеть хорошо справляется с задачами классификации. Также в ходе исследования [Фомин, 2025] комбинация из сиамской сверточной и сегментной спайковой сетей показала уверенные результаты в задаче классификации и реидентификации изображений людей по данным с цветной камеры.

В данной работе предлагается обобщить ранее предложенный алгоритм реидентификации с использованием сиамской сверточной и сегментной спайковой сетей на обработку изображений с камеры видеонаблюдения, работающей в ночном (инфракрасном) режиме, что позволит применять алгоритм также и в темное время суток. Также предложен подход на основе комбинации сиамской сети для цветных векторов, сиамской сети для инфракрасных векторов и объединяющей их сегментной спайковой сети с многомерным входом для одновременной реидентификации по двум разным источникам данных.

## **1. Обзор существующих подходов**

### **1.1. Обзор областей применения сиамских сетей**

Сравнение и сопоставление векторных представлений изображений сегодня регулярно применяются в задаче многообъектного обнаружения и сопровождения [Luo et al, 2021], [Bertinetto et al, 2016], можно обратить внимание на, то, что для решения этой задачи есть общий подход. Каждое изображение последовательности обрабатывается нейронной сетью для получения сжатого вектора представлений. Вектора представлений используются для сопоставления исходных матриц между собой на основе Евклидова расстояния, косинусного расстояния, или иных метрик дистан-

ции в  $n$ -мерном пространстве вещественных чисел. Различия между методами состоят в построении способа формирования векторов представлений.

Современные подходы к сопоставлению в процессе сопровождения в основном основаны на сиамских сетях. Их основная идея [Kim et al, 2017] – обучить нейронную сеть, использующую информацию с двух или более матриц (изображений) чтобы различить их наилучшим образом. Обычно часть сети для извлечения признаков для всех изображений общая, далее идут различные способы сравнения, сопоставления, разделения. В работе [Wang et al, 2015] предлагается сиамская сеть, обученная на контрастной функции ошибки. Она принимает на вход 2 изображения, их IoU, и выдает оценку схожести. В [Zhang et al, 2016] сеть принимает на вход не полные изображения, а их фрагменты и выдает оценку схожести уже между ними. В [Son et al, 2017] рассматривается триплетная функция ошибки (triplet loss). В статье [Zhu et al, 2018] авторы делают следующий шаг, и, добавляя еще один пример, формируют четверную функцию ошибки, где на входе 3 примера одного объекта за последние 3 кадра и 1 пример другого объекта. В режиме работы используется обученная сеть-дескриптор для построения представлений и рассчитывается сопоставление уже между ними. В [Hermans et al, 2017] выполняется сравнение на 128-мерном векторе представлений, извлекаемом тройной сверточной сетью, описанной в [Zhou et al, 2019]. В [Chen et al, 2018] приводится сверточная сеть для предсказания нового положения, которая способна извлекать представления из фрагментов изображения для уточнения сопоставления при получении нового кадра. В [Dai et al, 2016] показан алгоритм сопоставления с использованием GoogLeNet, обученный с использованием triplet loss. Сеть R-FCN [Bakhshiev et al, 2022] используется для генерации предположений о положении объекта на основе траектории и признаков с предыдущих кадров, потом предположения объединяются с реальными обнаружениями и используется обученная GoogLeNet для уточнения.

Кроме простых способов определения близости векторов и отнесения объекта к одной из групп, существуют и более интересные, биоинспирированные подходы. В работе [Корсаков и др., 2022] рассмотрена авторская биоподобная сегментная спайковая модель нейрона, которая хорошо показала себя в задаче классификации (см. [Бахшиев и др., 2021]). Модель способна успешно сопоставлять даже сложные, линейно не разделимые данные.

Исходя из рассмотренных работ можно сделать однозначный вывод о том, что использование сиамских сверточных сетей для извлечения векторов признаков желаемого размера является общепринятым подходом, который актуален для применения в данной работе. С точки зрения алгоритма обучения сиамской сверточной сети выделяются два основных подхода, тройная функция ошибки (triplet loss), ориентированная на раз-

делении точек данных при обучении на основе информации о положительных и отрицательных парах объектов и контрастивное обучение (contrastive loss), который обучается на парах изображений (положительных и отрицательных) независимо.

## 1.2. Обзор сегментной спайковой модели нейрона

В рамках выполнения настоящей работы использовалась сегментная спайковая модель нейрона (CSNM) [Бахшиев и др., 2021]. Настоящая статья не ставит своей задачей подробное математическое описание использовавшейся модели нейрона. Такое описание можно найти в вышеуказанной работе.

Предполагается, что на вход поступают спайки, которые преобразуются в синапсах в выходную величину, отображающую влияние синаптического тока на сегмент мембраны нейрона (рис. 1).

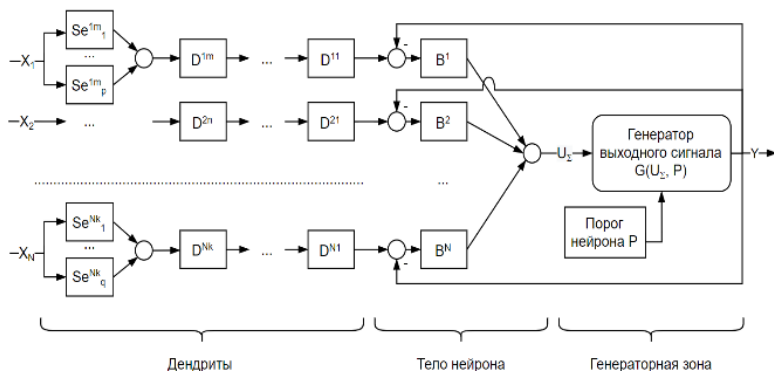


Рис. 1. Структура сегментной спайковой модели нейрона.  $Se^{ij}_l$  – входной синапс  $l$ , связанный с сегментом дендрита  $D^{ij}$ ;  $D^{ij}$  –  $j$ -й сегмент  $i$ -го дендрита;  $B^k$  –  $k$ -й сегмент тела нейрона

Представленная структурная организация модели нейрона позволяет проводить расчет систем дифференциальных уравнений для каждого нейрона независимо, поскольку нейроны обмениваются исключительно информацией о моменте времени возникновения спайка, а входной вектор системы уравнений модели нейрона формируется внутри модели как функция от моментов времени прихода спайков.

Полагаем, что нейроны обмениваются информацией через события (возникновение спайков), которые можно представить по формуле:

$$(1.1)$$

где  $x_i$  – выход нейрона в  $i$ -й интервал времени, а  $t$  – время.

Размер нейрона (число сегментов сомы), множество длин дендритов и число возбуждающих и тормозных синапсов для каждого сегмента мембраны определяют структуру нейрона, которая формирует требуемую реакцию на входные воздействия.

Данный нейрон характеризуется особенной методикой структурного обучения. Согласно ей, для заданного количества сегментов сомы, сначала выполняется процедура синхронизации, когда последовательно наращиваются длины дендритов, чтобы обеспечить одновременность прихода сигнала на сому, затем процедура нормализации, когда наращивается число синапсов, чтобы компенсировать падение уровня сигнала при прохождении по элементам дендрита, комбинация этих операций обеспечивает настройку на паттерн импульсов.

## **2. Предлагаемые решения**

### **2.1. Наборы данных для экспериментальных исследований**

Некоторые решения, принятые в рамках исследования предлагаемого алгоритма, учитывают отдельные характеристики набора данных, поэтому требуется начать с его общего описания.

В качестве источника данных для проведения экспериментов был выбран набор данных SYSU-MM01 [Wu et al, 2017]. Набор содержит изображения нескольких сотен разных людей (персон), полученные с 4 различных цветных и 2 различных инфракрасных камер. Для последующей обработки изображения людей были извлечены из исходных изображений и преобразованы в набор обучающих тензоров фиксированного размера, с дополнительным осреднением и случайными изменениями извлекаемой области, цветности, ориентации и т.д.. Все вектора были поделены по персонам, данные с различных камер были объединены. Оставлены только персоны, присутствующие и на цветных, и на ИК изображениях. В итоге, получено два набора серий векторов, сгруппированных по персонам, более двух сотен персон для обоих вариантов датасета.

Так как для каждой персоны представлено большое количество обучающих тензоров (особенно, извлеченных из данных, полученных с цветной камеры), для основного набора данных (табл. 1, «..., осн.») было выбрано по 50 первых персон для наборов, полученных из цветных и ИК изображений. Также были подготовлены полные наборы тензоров, извлеченных из обеих типов изображений (табл. 1, «..., полн.»), для исследования того, насколько предложенная концепция обучения и тестирования может быть успешно адаптирована кратно большему количеству классов. Параметры наборов представлены в табл. 1.

Таблица 1

Набор	Тип	Классов	Прим. мин.	Прим. макс.	Прим. среднее	Прим. медиана
ИК, осн.	Обуч.	50	32	36	32,28	32
ИК, осн.	Тест	50	8	10	8,3	8
ИК, полн.	Обуч.	233	32	36	32	32
ИК, полн.	Тест	233	8	10	8	8
Цв., осн.	Обуч.	50	56	102	75,84	76
Цв., осн.	Тест	50	14	26	19,46	19
Цв., полн.	Обуч.	233	48	115	74,4	73
Цв., полн.	Тест	233	13	29	19,07	19

## 2.2. Организация сиамской сети и подход к ее обучению

Так как набор данных достаточно большой, то было принято решение выстроить сиамскую сверточную сеть на основе полноценной сверточной сети, аналогично тому, как было сделано ранее для другого набора [Фомин, 2025]. В качестве основы была взята сеть ResNet-18 [He et al, 2016], предобученная на наборе данных ImageNet [Deng et al, 2009]. Конечный слой, предназначенный для обучения на данных для классификации, был отброшен и заменен на слой линейной рекомбинации с двумя выходами. Также в качестве сети для обучения рассматривалась и собственная сеть, представленная в работе [Фомин, 2025], но ряд экспериментов с разными параметрами обучения показал низкую степень сходимости, для существенно большего по размеру набора данных числа обучаемых параметров оказалось недостаточно.

В качестве функции ошибки исходно рассматривалось два варианта, triplet loss с оригинальным способом формирования обучающих отрицательных примеров [Фомин, 2025], и классический contrastive loss, представленный в том числе и в стандартной библиотеке функций ошибки pytorch, но специально подготовленный для адаптации к имевшейся программной части. Эксперименты с triplet loss продемонстрировали низкую сходимость, при том что на одну эпоху обучения (обработку полного набора данных) требовалось очень много времени в силу необходимости статистической обработки результатов для формирования отрицательных обучающих примеров на следующей эпохе. Замена triplet loss на contrastive loss позволила на треть сократить память, занимаемую одним батчем обучения, и существенно уменьшить время обсчета одной эпохи. Увеличение числа эпох позволило добиться сходимости к качественному распределению кластеров классов в пространстве представлений.

Таким образом, все обучение выполнялось на сиамской сети на базе ResNet-18, с функцией ошибки contrastive loss, до 1000 эпох обучения в обоих случаях, с фиксированными значениями скорости обучения оптимизатора и параметрами функции ошибки.

### **2.3. Подготовка данных и обучение сегментной спайковой сети**

Для распознавания был использован подход к классификации (реидентификации), подробно описанный в работе [Фомин, 2025]. Все особенности классификации (расознавания) последовательностей импульсов подробно изложены в основополагающих работах по этой теме [Бахшиев и др., 2021], [Корсаков и др., 2022]. Кодирование результатов работы из пространства представлений сиапской сверточной нейронной сети в спайковое представление для использования в сегментной спайковой нейронной сети выполнено прямым пропорциональным преобразованием, аналогично тому, как описано в [Фомин, 2025]. Иными словами, каждая точка данных кодируется 3 импульсами задержки по 3 каналам, 1 импульс соответствует 1-й координате вектора представлений, 2 импульс 2-й координате, 3-й используется для синхронизации. Задержка импульса прямо пропорциональна соответствующей координате в векторе представлений. Кодирование выполнено единообразно для представлений тензоров, извлеченных из цветных и инфракрасных изображений.

Исходная архитектура сегментной сети – 3 (5) сегмента тела нейрона, к каждому подключен дендрит, на вход которого поступает импульс от генератора. По завершении обучения на каждом из каналов в рамках процедуры синхронизации увеличивается число сегментов дендрита обратно пропорционально значению задержки сигнала, а также число синапсов на входном дендрите в рамках процедуры нормализации. Подробнее процедура обучения описана в [Бахшиев и др., 2021].

Для экспериментов по совместной реидентификации по данным, извлеченным из цветных и инфракрасных изображений, подготовленные импульсные представления были объединены следующим образом. Для каждой пары импульсов из пространства представлений для цветных данных, подбирались случайная пара импульсов из пространства представлений для инфракрасных данных, которых было меньше, потому пары иногда повторялись. Нейрон при обучении имел вместо двух входов данных 4 входа, по 2 для каждого из типов, плюс 1 синхронизирующий вход. В качестве обучающих данных классов брались те же центроиды, что и при автономном обучении.

## **3. Результаты экспериментального исследования**

Экспериментальные исследования в данной работе преследовали три основные цели. Во-первых, требовалось удостовериться, что подход, подразумевающий применение сиапской сверточной нейронной сети для перевода тензоров данных, извлеченных из изображений в малоразмерное пространство признаков, может быть обобщен до существенно больших наборов данных по сравнению с тем, что применялся при первичном обучении. Во-вторых, необходимо было экспериментально подтвердить, что

данный подход может быть без существенных изменений обобщен на обработку тензоров, извлеченных из инфракрасных изображений, следовательно, метод реидентификации может быть обобщен на работу в темное время суток и в сумерках. В-третьих, необходимо было проверить гипотезу о том, что универсальный компонент классификации на основе сегментной спайковой нейронной сети принципиально может быть обобщен на совместную идентификацию результатов преобразования цветных и ИК данных, на примере объединения не связанных между собой напрямую данных. В-четвертых, был поставлен отдельный эксперимент по обучению siamoй сверточной нейронной сети на полном наборе данных цветных изображений для исследования принципиальной возможности обобщения предложенного подхода на большое количество классов и большой объем обучающих данных.

В силу вычислительной сложности модели даже с учетом переноса вычислений на GPU и использования параллельных вычислений там, где это допустимо обучение и тестирование модели занимает продолжительное время. Для сокращенного набора данных обучение и тестирование занимает порядка 8-10 часов на Nvidia 1080Ti, для полного набора – 22 тыс. примеров – 211593,59 с на том же вычислителе. Существенное влияние оказывает количество нейронов (233 нейрона на полном наборе). Обучение сверточной сети на малом наборе потребовало нескольких дней, на полном наборе – около полутора недель вычислительного времени.

Оценка качества реидентификации во всех трех случаях выполнялась одинаковым образом. Велся подсчет правильно классифицированных точек пространства представления (TP), ошибок 2 рода, когда срабатывал неправильный нейрон (FP) и ошибок 1 рода, когда не срабатывал правильный нейрон (FN). На основе этих данных по формулам (3.1) были рассчитаны показатели точности (Precision, Prec), полноты (Recall, Rec), F<sub>1</sub>-меры.

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}, \quad F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (3.1)$$

Приводятся показатели и для тестового набора данных, и для обучающего. Для нейронной сети в качестве обучающих примеров выступают центры кластеров классов в обучающем наборе, а все точки данных как из обучающего набора, так и из тестового выступают в качестве тестовых примеров, поэтому такая оценка корректна.

Результаты тестирования представлены в табл. 2.



Таблица 2

Набор	Обучающий набор данных			Тестовый набор данных		
	Prec	Rec	$F_1$	Prec	Rec	$F_1$
ИК	0,951	0,996	0,973	0,807	0,993	0,891
Цв., 500	0,966	0,995	0,981	0,911	0,995	0,952
Цв., 1000	0,996	1,000	0,998	0,932	1,000	0,965
ИК+Цв., 500	0,996	0,992	0,994	0,982	0,932	0,957
ИК+Цв., 1000	0,998	0,996	0,997	0,986	0,920	0,952
Цв., полный	0,556	0,999	0,714	0,518	0,999	0,682

Полученные результаты полностью подтверждают высокую результативность предложенного подхода во всех рассмотренных экспериментальных сценариях. Снижение показателей на полном наборе связано в первую очередь с тем, что несмотря на заметную сходимость использованная сверточная нейронная сеть с линейным классификатором не показала достаточной способности к обобщению. Также контрастная функция ошибки как способ обучения недостаточно хороша для малых наборов (что было показано в предыдущих работах). С увеличением количества точек данных выше некоторого предела кластеры перестают быть линейно разделимыми, поэтому часть точек оказывается ближе к другим классам, что увеличивает ошибку. Доработка функции ошибки и выбор сети для больших наборов данных есть предмет дальнейшего исследования.

### Заключение

В работе исследовано обобщение ранее предложенного подхода к реидентификации людей на основе комбинации сиамской сверточной и сегментной спайковой нейронных сетей на в несколько раз больший набор данных. Предложено обобщение на обработку данных, полученных в ночном (инфракрасном) режиме камеры видеонаблюдения. Автономное качество ( $F_1$ -мера здесь и далее) для цветных данных составило 0,932 и 0,965 для двух разных эпох, для инфракрасных 0,891. Качество на объединенном датасете составило 0,957 и 0,952 для двух пар результатов. Наконец, экспериментальное исследование на полном наборе (в десятки раз больше рассмотренного в прошлых работах) показало качество 0,714 для обучающих и 0,682 для тестовых данных, извлеченных из цветных изображений, что подтверждает работоспособность предложенного подхода в этих условиях.

Так как исследование на данных, извлеченных из изображений демонстрирует устойчивую успешную классификацию как для наборов данных малого размера, так и для наборов среднего размера, в качестве направления дальнейших исследований предполагается обобщить принцип на тензоры данных, извлекаемые из сенсоров иной физической природы (облака точек и линии с LiDAR, тепловизионные изображения, карты глубины, многоспектральные данные и др.).

## Список литературы

- [Бахшиев и др., 2021] Бахшиев А.В., Демчева А.А. Сегментная спайковая модель нейрона // Нелинейная динамика в когнитивных исследованиях (Нижний Новгород, 2021): Труды VII Всероссийской конференции. – 2021. – С. 26-29.
- [Корсаков и др., 2022] Корсаков А.М., Астапова Л.А., Бахшиев А.В. Применение сегментной спайковой модели нейрона со структурной адаптацией для решения задач классификации // Информатика и автоматизация. – 2022. – Т. 21, № 3. – С. 493-520.
- [Фомин, 2025] Фомин И.С. Сравнение сжатых векторных представлений матриц с помощью сегментной спайковой модели нейрона (CSNM) // Робототехника и техническая кибернетика. – 2025. – Т. 13, № 1. – С. 33-40.
- [Bakhshiev et al, 2022] Bakhshiev A., Demcheva A., Stankevich L. CSNM: The Compartmental Spiking Neuron Model for Developing Neuromorphic Information Processing Systems // In: Advances in Neural Computation, Machine Learning, and Cognitive Research V: Studies in Computational Intelligence, eds. B. Kryzhanovsky et al. – Cham: Springer International Publishing, 2022. – P. 327-333. – doi: 10.1007/978-3-030-91581-0\_43.
- [Bertinetto et al, 2016] Bertinetto L., Valmadre J., Henriques J.F., Vedaldi A., Torr P.H.S. Fully-Convolutional Siamese Networks for Object Tracking. In: Computer Vision – ECCV 2016 Workshops : Lecture Notes in Computer Science / eds. G. Hua, H. Jégou. – Cham: Springer International Publishing, 2016. – P. 850-865. – doi: 10.1007/978-3-319-48881-3\_56.
- [Chen et al, 2018] Chen L., Ai H., Zhuang Z., Shang C. Real-Time Multiple People Tracking with Deeply Learned Candidate Selection and Person Re-Identification // In: 2018 IEEE International Conference on Multimedia and Expo (ICME). – 2018. – P. 1-6. – doi: 10.1109/ICME.2018.8486597.
- [Dai et al, 2016] Dai J., Li Y., He K., Sun J. R-FCN: Object Detection via Region-based Fully Convolutional Networks // In: Advances in Neural Information Processing Systems, Curran Associates, Inc. – 2016. – P. 29. – doi:10.48550/arXiv.1605.06409.
- [Deng et al, 2009] Deng J., Dong W., Socher R., Li L.J., Li K., Fei-Fei L. Imagenet: A large-scale hierarchical image database // In: 2009 IEEE conference on computer vision and pattern recognition, 2009 Jun 20. – P. 248-255. – doi: 10.1109/CVPR.2009.5206848.
- [He et al, 2016] He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition // In: Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – P. 770-778. – doi: 10.1109/CVPR.2016.90.
- [Hermans et al, 2017] Hermans A., Beyer L., Leibe B. In Defense of the Triplet Loss for Person Re-Identification [Электронный ресурс] // arXiv:1703.07737 [cs] // arXiv. – 2017. – URL: <https://arxiv.org/abs/1703.07737> (дата обращения: 24.05.2025). doi:10.48550/arXiv.1703.07737.
- [Kim et al, 2017] Kim M., Alletto S., Rigazio L. Similarity Mapping with Enhanced Siamese Network for Multi-Object Tracking [Электронный ресурс] // arXiv:1609.09156 [cs] // arXiv. – 2017. – URL: <https://arxiv.org/abs/1609.09156> (дата обращения: 24.05.2025). – doi: 10.48550/arXiv.1609.09156.
- [Luo et al, 2021] Luo W., Xing J., Milan A., Zhang X., Liu W., Kim T.-K. Multiple Object Tracking: A Literature Review // Artificial Intelligence. – 2021. – Vol. 293. – P. 103448. – doi: 10.1016/j.artint.2020.

- [**Son et al, 2017**] Son J., Baek M., Cho M., Han B. Multi-Object Tracking With Quadruplet Convolutional Neural Networks // In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2017. – P. 5620-5629. – doi: 10.1109/CVPR.2017.403.
- [**Wang et al, 2015**] Wang B., Wang L., Shuai B., Zuo Z., Liu T., Luk Chan K., Wang G. Joint Learning of Convolutional Neural Networks and Temporally Constrained Metrics for Tracklet Association // In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. – 2016. – P. 1-8. – doi: 10.1109/CVPRW.2016.55.
- [**Wu et al, 2017**] Wu A., Zheng W.S., Yu H.X., Gong S., Lai J. RGB-infrared cross-modality person re-identification // In: Proceedings of the IEEE international conference on computer vision. – 2017. – P. 5380-5389. – doi: 10.1007/s11263-019-01290-1.
- [**Zhang et al, 2016**] Zhang S., Gong Y., Huang J.-B., Lim J., Wang J., Ahuja N., Yang M.-H. Tracking Persons-of-Interest via Adaptive Discriminative Features // In: Computer Vision – ECCV 2016: Lecture Notes in Computer Science, eds. B. Leibe et al. – Cham: Springer International Publishing, 2016. – P. 415-433. – doi: 10.1007/978-3-319-46454-1\_26.
- [**Zhou et al, 2019**] Zhou H., Ouyang W., Cheng J., Wang X., Li H. Deep Continuous Conditional Random Fields With Asymmetric Inter-Object Constraints for Online Multi-Object Tracking // IEEE Transactions on Circuits and Systems for Video Technology. – 2019. – Vol. 29(4). – P. 1011-1022. doi:10.1109/TCSVT.2018.2825679.
- [**Zhu et al, 2018**] Zhu J., Yang H., Liu N., Kim M., Zhang W., Yang M.-H. Online Multi-Object Tracking with Dual Matching Attention Networks // In: Proceedings of the European Conference on Computer Vision (ECCV). – 2018. – P. 366-382. – doi: 10.1007/978-3-030-01228-1\_23.