

# ОСНОВНЫЕ ПОНЯТИЯ МОДУЛЯ



## **ВИДЕО**

Это упорядоченная последовательность кадров одинакового разрешения, захватываемая и отображаемая с заданной частотой.

## **ВИДЕОПОТОК**

Это поток изображений в реальном времени. В видеопотоке мы рассматриваем текущее изображение и предыдущие.

## **ВИДЕОРЯД**

Это видео фиксированной длины. В видеоряде у нас есть доступ ко всему видео, от первого изображения до последнего.

## **ОТСЛЕЖИВАНИЕ ОБЪЕКТОВ (OBJECT TRACKING)**

Это процесс обнаружения и перемещения объекта или нескольких объектов в видео с течением времени. Отслеживание выполняется путём вычисления сходства между кадрами 0 и 1.

## **VOT (VISUAL OBJECT TRACKING)**

Это краткосрочное отслеживание объекта.

## **MOT (MULTIPLE OBJECT TRACKING)**

Это долгосрочное одновременное отслеживание нескольких объектов.

## АВТОНОМНОЕ ОТСЛЕЖИВАНИЕ ОБЪЕКТОВ

Это отслеживание объектов на записанном видео, где все кадры, включая будущую активность, известны заранее.

## ОНЛАЙН-ОТСЛЕЖИВАНИЕ ОБЪЕКТОВ

Это отслеживание объектов в прямом видеопотоке, например, через камеры наблюдения.

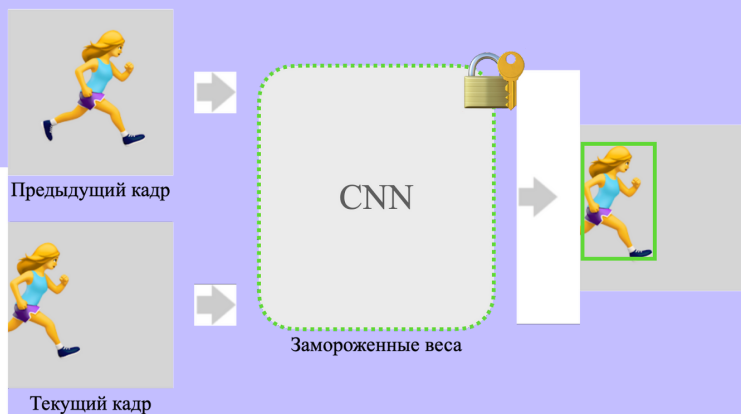
## ФИЛЬТР КАЛМАНА

Это алгоритм, который может предсказывать будущие позиции на основе текущей позиции. Он также может оценить текущее положение объекта лучше, чем датчик или алгоритм. Используется в SORT и DeepSORT.

## ВЕНГЕРСКИЙ АЛГОРИТМ

Может определить, совпадает ли объект в текущем кадре с объектом в предыдущем кадре.

# GENERIC OBJECT TRACKING USING REGRESSION NETWORK (GOTURN)



GOTURN обучается путём сравнения пар обрезанных кадров из тысяч видеопоследовательностей. В первом кадре («предыдущий кадр») местоположение объекта известно, и кадр обрезается до удвоенного размера ограничительной рамки вокруг объекта с центрированием объекта.

Затем алгоритм пытается предсказать местоположение того же объекта во втором кадре («текущем кадре»). Та же ограничительная рамка двойного размера используется для обрезки второго кадра. CNN обучена предсказывать расположение ограничивающей рамки во втором кадре.

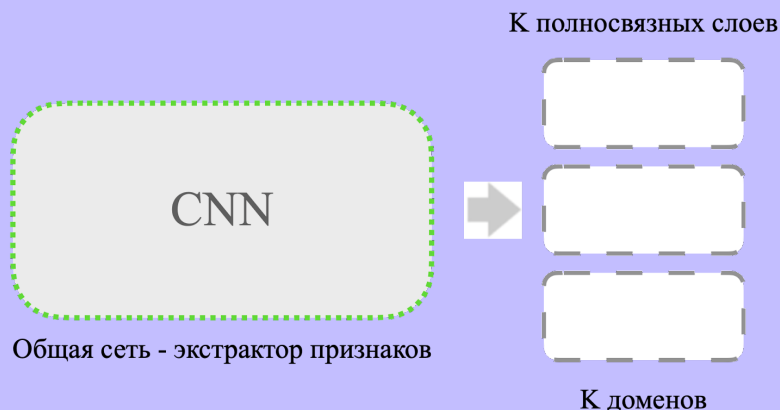
# MULTI-DOMAIN NETWORK (MDNET)



## МНОГОДОМЕННАЯ СЕТЬ (MDNET)

Это архитектура CNN, победившая в конкурсе VOT2015.

Цель MDNet — ускорить обучение, чтобы обеспечить результаты в режиме реального времени.



Стратегия состоит в том, чтобы разделить сеть на две части.

- Первая часть действует как экстрактор общих признаков, который учится отличать объекты от их фона.
- Вторая часть обучается на конкретном обучающем наборе и учится идентифицировать объекты в кадрах видео.

Таким образом, MDNet позволяет изменять веса только нескольких последних слоёв CNN во время обучения, что значительно сокращает время вычислений объектов.

# ОБЩИЙ ПРИНЦИП DETECTION-BASED АЛГОРИТМОВ MOT

Этап	Описание
Обнаружение	Сначала обнаруживаются все объекты в кадре. Может быть одиночное или множественное обнаружение.
Ассоциация	Как только появляются обнаружения для кадра, выполняется сопоставление для аналогичных обнаружений по отношению к предыдущему кадру. Совпавшие кадры прослеживаются по всей последовательности, чтобы получить отслеживание объекта.

## DEEP SORT

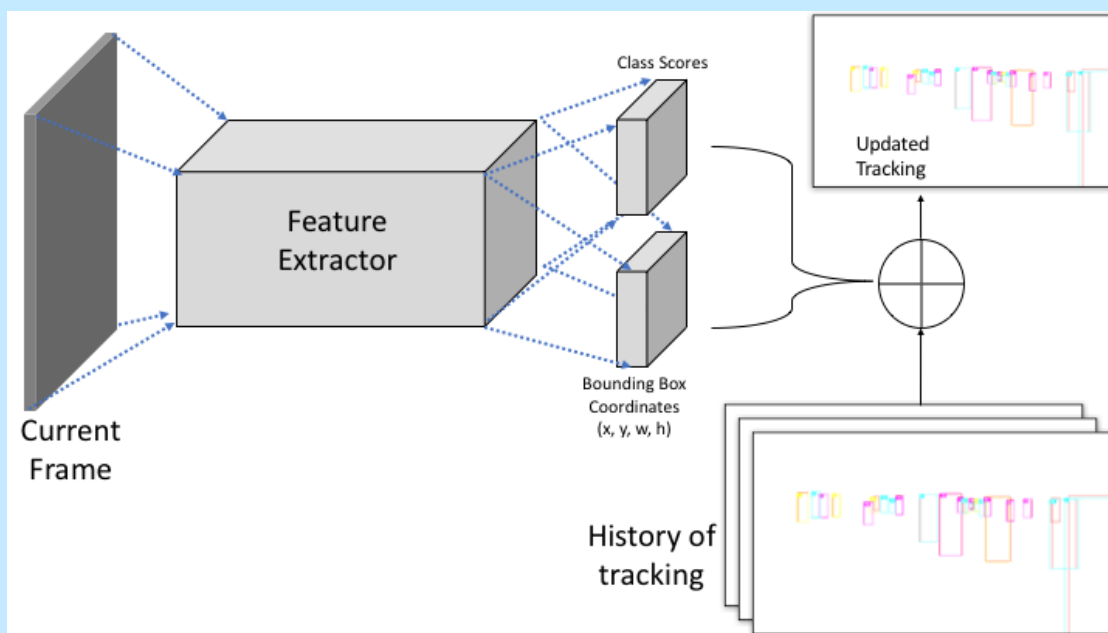
Этап	Описание
Детектирование	Для вычисления обнаружений на каждом кадре используется популярный метод детектирования объектов на основе CNN (Faster-RCNN в оригинальной статье).

## Оценка

Промежуточный шаг перед ассоциацией данных состоит из модели оценки. При этом состояние каждого трека используется в виде вектора восьми величин, то есть центра ограничивающей рамки ( $x, y$ ), масштаба рамки ( $s$ ), соотношения сторон ( $a$ ) и их производных по времени в качестве скоростей. Фильтр Калмана используется для моделирования этих состояний как динамической системы.

## Ассоциация

На последнем этапе, учитывая предсказанные состояния от фильтрации Калмана с использованием предыдущей информации и вновь обнаруженной рамки в текущем кадре, создаётся ассоциация для нового обнаружения со старыми треками объекта в предыдущем кадре. Это вычисление происходит с использованием Венгерского алгоритма.



# ВИДЕО-АНАЛИЗ

