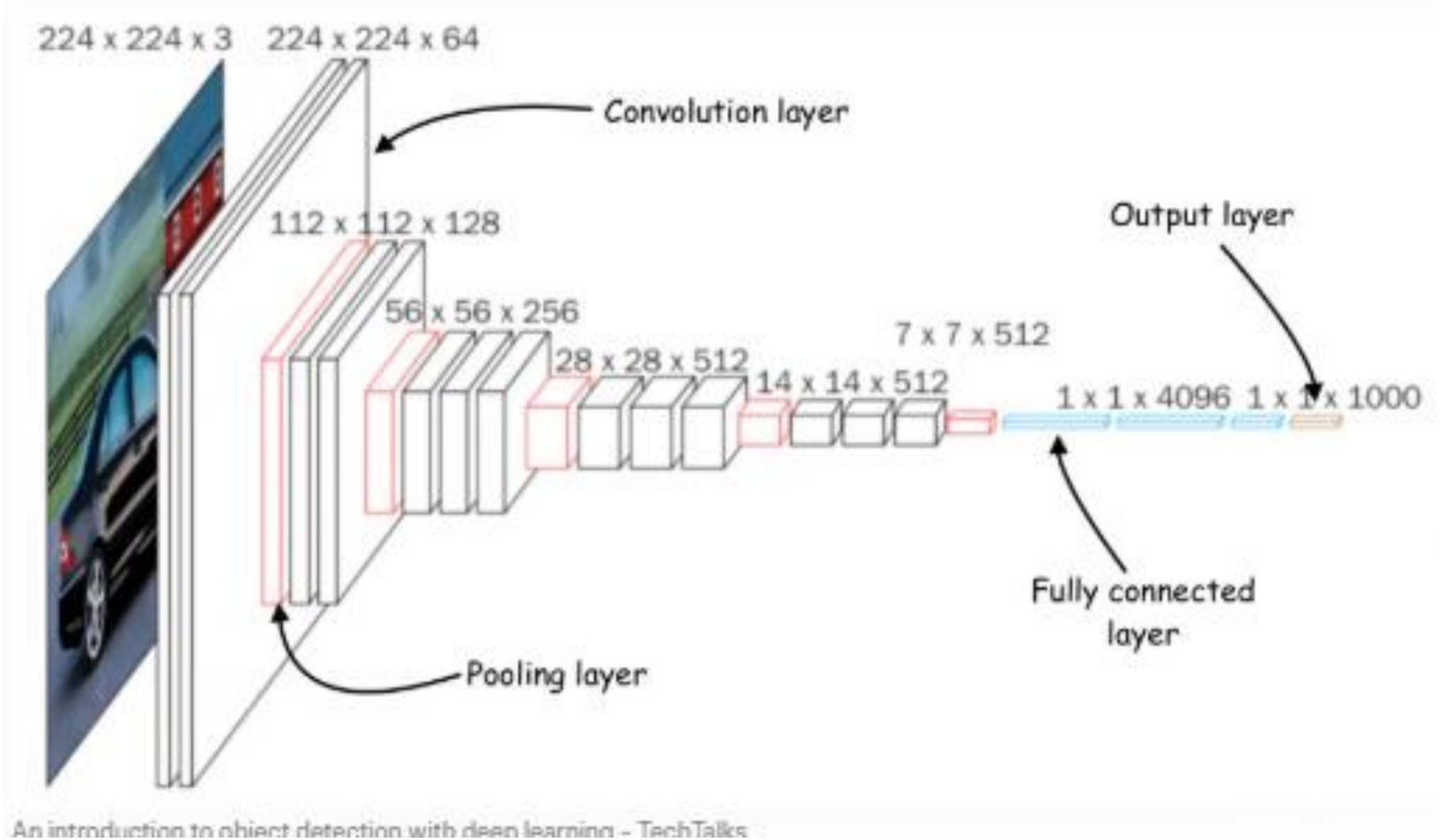
Модели глубокого обучения в задачах восприятия ADAS (CNN, YOLO, SSD, Transformer-based модели)
(Обнаружение, классификация, трекинг объектов в реальном времени)

1. Основы Convolutional Neural Networks (CNN) в задачах восприятия CNN - это фундамент всех современных моделей компьютерного зрения. **Ключевые операции**: Свёртка (convolution) - извлечение локальных признаков (края, текстуры, формы). Pooling (max/average) - снижение размерности и инвариантность к сдвигу. Activation (ReLU, Swish, GELU).Batch Normalization + Residual connections (ResNet, 2015) - решение проблемы vanishing gradient. Типичная архитектура backbone (ResNet-50/101, EfficientNet, CSPDarknet) используется как feature extractor.

Применение в ADAS: Классификация дорожных знаков (German Traffic Sign Dataset - >99 % точность).Семантическая сегментация (LaneNet, DeepLabv3+). Feature maps для последующих задач детекции и трекинга. CNN отлично работают с локальными паттернами, но плохо захватывают глобальный контекст сцены (дальний автомобиль + пешеход на переходе).



An introduction to object detection with deep learning - TechTalks

2. Обнаружение объектов (Object Detection)Задача: для каждого объекта на изображении выдать bounding box + класс + confidence. Два подхода:Two-stage (R-CNN → Fast R-CNN → Faster R-CNN): сначала region proposals (Selective Search или RPN), потом классификация. Высокая точность, но 5–15 FPS - не для реального времени в ADAS.One-stage (YOLO, SSD): предсказание сразу на всей карте признаков. Скорость 30–100 + FPS.3. Семейство YOLO - «король» реального времениПринцип «You Only Look Once» (Redmon, 2015): изображение делится на grid (S×S), в каждой ячейке предсказываются B bounding boxes + вероятности классов.Эволюция (по состоянию на март 2026):YOLOv1–v3 - классика.YOLOv4–v7 - CSPDarknet, PANet, anchor-free.YOLOv8–v11 (Ultralytics) - decoupled head, anchor-free, Distill.YOLOv12 (2025) - введение attention-механизмов.YOLO26 (январь 2026, Ultralytics) - полностью переработан для edge: NMS-free (end-to-end), удалён DFL, новый оптимизатор MuSGD, +43 % скорости на CPU, отличная точность по мелким объектам.

В бенчмарках ADAS (2025) YOLOv12 показывает mAP@50–95 = 82,2 %, 70–73 FPS; YOLO26 ещё быстрее и легче на embedded.

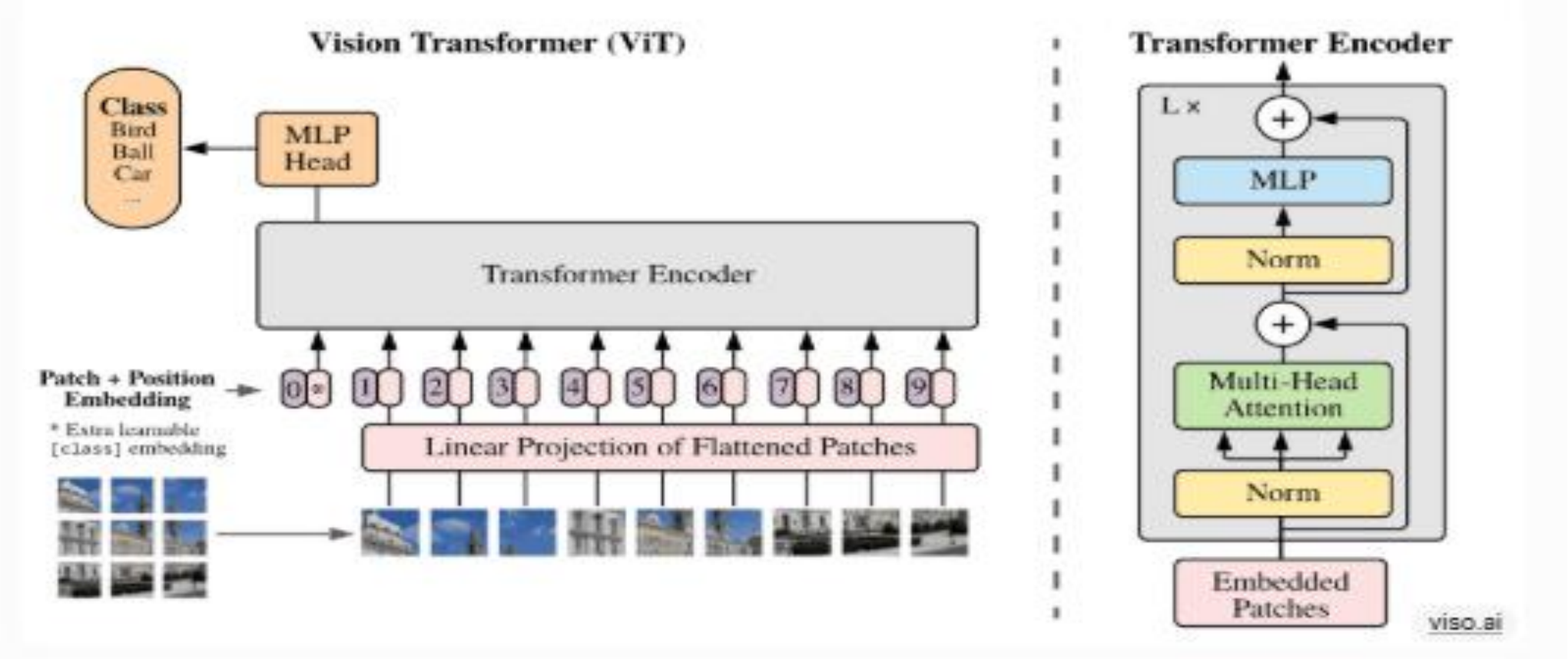
Преимущества в ADAS: Реальное время даже на Jetson Nano/Orin Nano.Легко дообучать на своих датасетах (KITTI, nuScenes, BDD100K, российские дороги). Много задач в одном модели (detection + segmentation + pose + classification).

4. SSD (Single Shot MultiBox Detector, 2016) Ключевая идея - multi-scale feature maps (из разных слоёв backbone). На каждой карте - default boxes (аналог anchors) разных масштабов и aspect ratio.

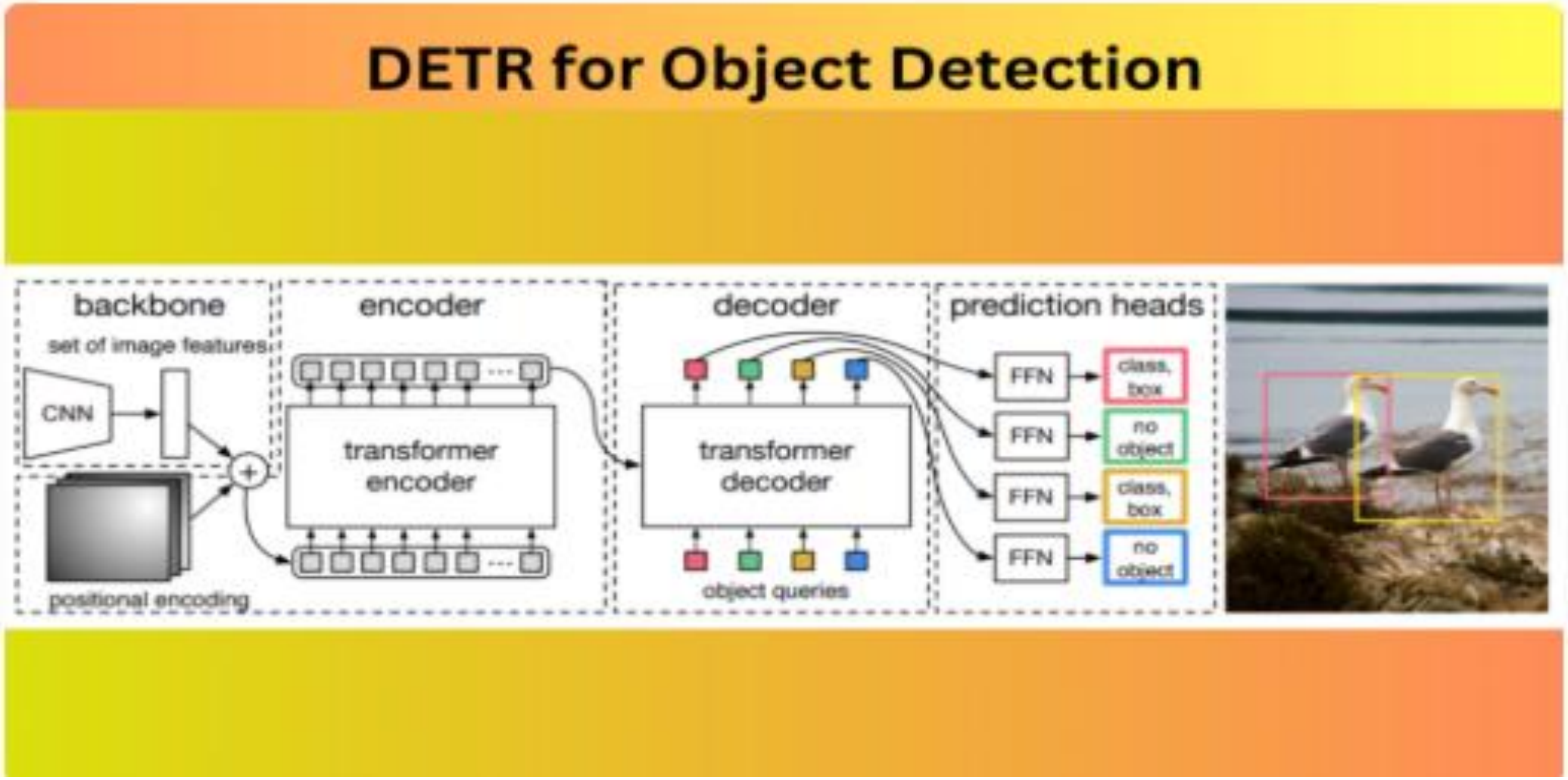


Сильные стороны по сравнению с ранними YOLO: Лучше детектирует мелкие объекты (пешеходы на расстоянии 50–100 м).Нет жёсткого grid-ограничения. **Слабые стороны**:Требует большего количества default boxes → больше вычислений. В 2026 году уступает YOLO26 и RT-DETR по скорости/точности, но всё ещё используется в мобильных и low-power системах (SSD-MobileNetV3).

5. Transformer-based модели - новый стандарт 2024–2026Vision Transformer (ViT, 2020) разбил изображение на патчи 16×16, применил self-attention. **Плюсы**: глобальный контекст уже на первом слое, отличное понимание сложных сцен. **Минусы**: квадратичная сложность $O(N^2)$, где N - число патчей.

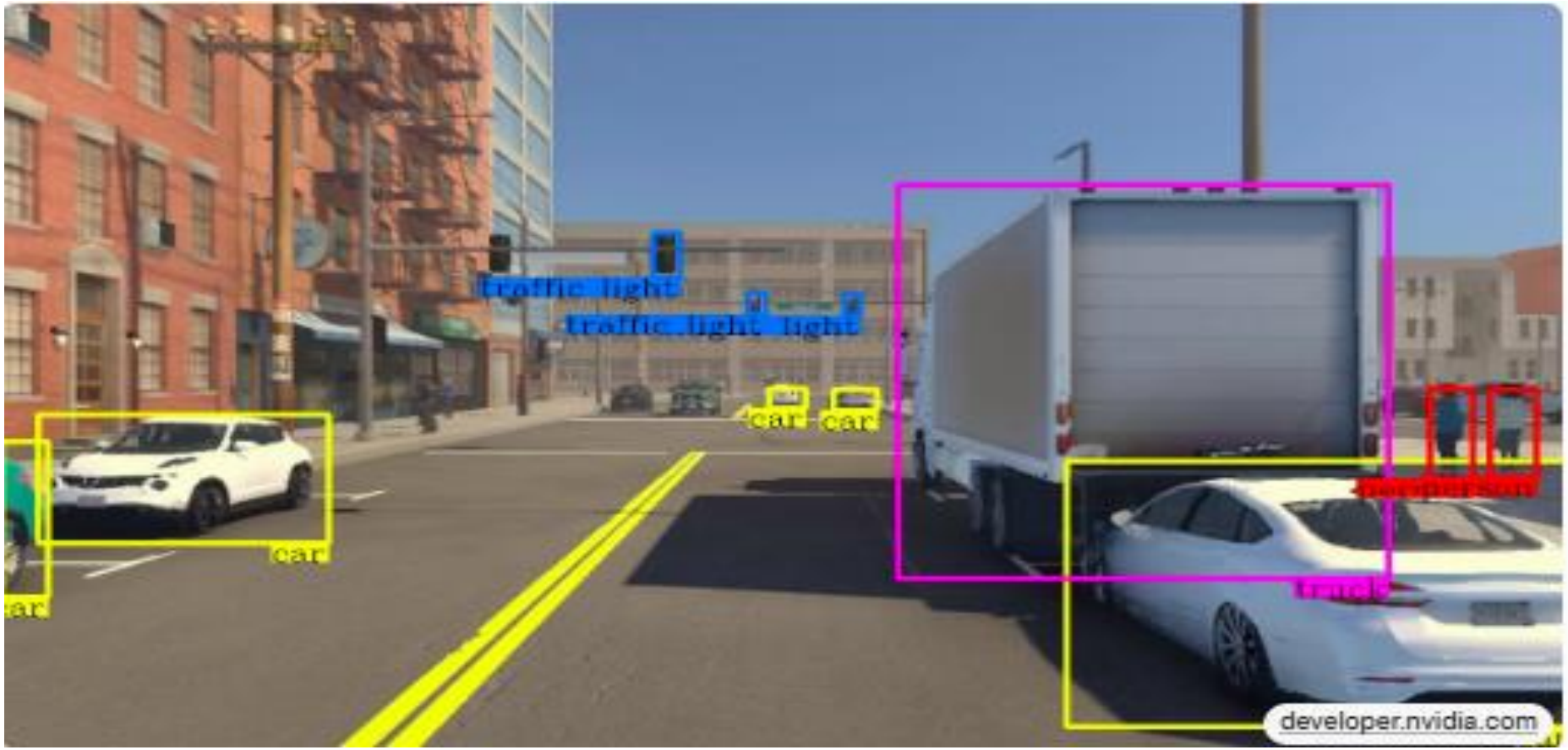


Для детекции - DETR (2020):
- Backbone CNN → Transformer encoder-decoder.
- Объекты - как набор предсказаний (set prediction).
- Hungarian matching + bipartite loss - нет NMS и anchors!
RT-DETR (Baidu, CVPR 2024 + обновления v2/v3/v4 до 2025):
- Hybrid encoder (CNN + efficient transformer).
- IoU-aware query selection.
- Scalable decoder (можно менять число слоёв на инференсе без переобучения).
- Превосходит многие YOLO по точности при той же скорости.
- Идеален для ADAS: лучше работает с окклюзиями, дальними объектами, сложным фоном.



В 2025-2026 годах RT-DETRv4 и гибридные YOLO26 + attention часто выбирают для Level 3+ автономности.

6. Пример работы детекции в реальной дорожной сцене



Обратите внимание: модель одновременно детектирует автомобили, грузовик, пешеходов, светофоры - всё в одном проходе.

7. Трекинг объектов (Multi-Object Tracking, MOT)

Детекция на каждом кадре - недостаточно. Нужно связать объекты между кадрами и предсказывать траектории.

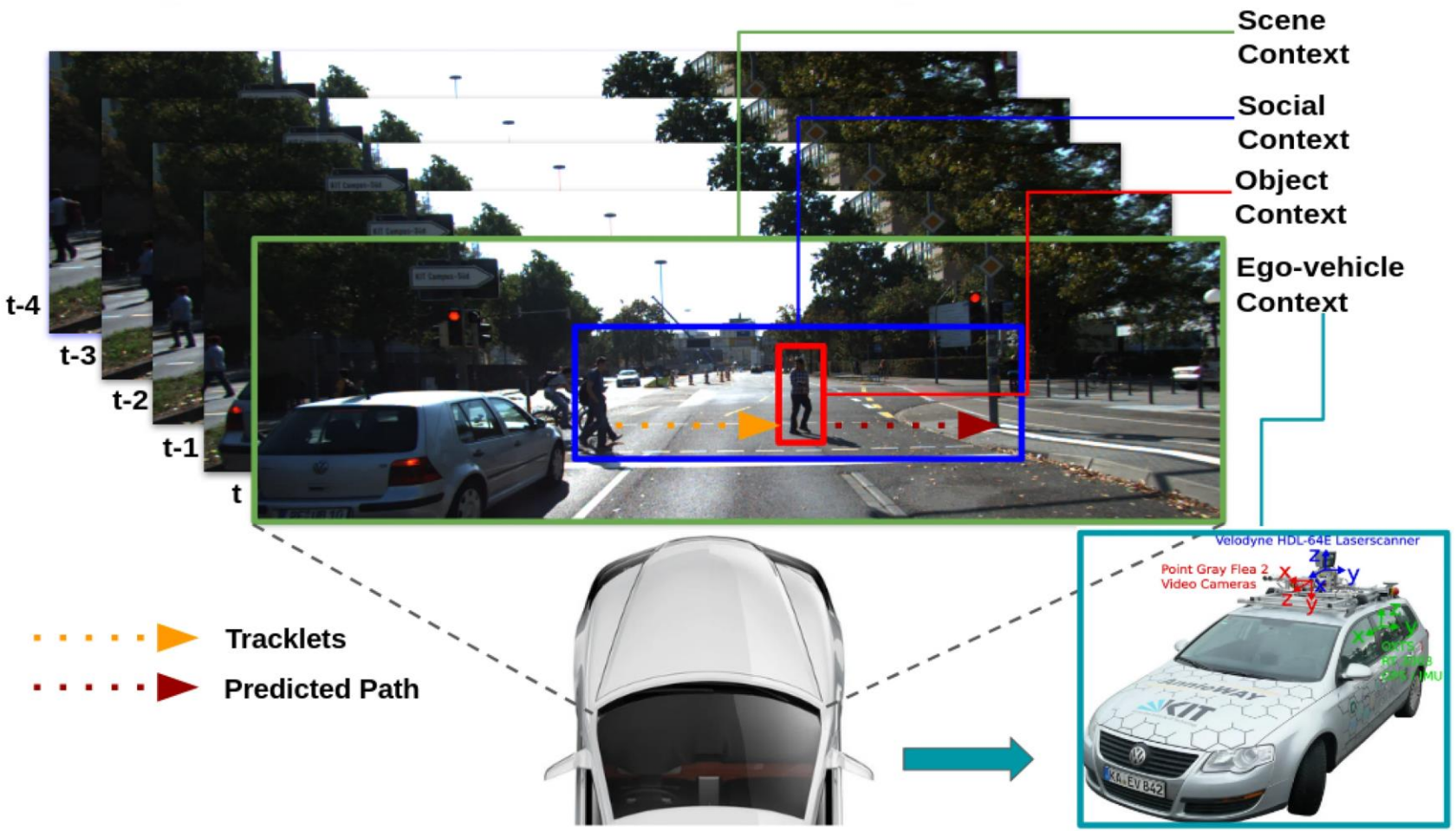
Классическая парадигма tracking-by-detection:

- Детектор (YOLO/RT-DETR) → bounding boxes.
- Motion model - Kalman filter (или LSTM/Transformer).
- Data association - Hungarian algorithm + IoU / ReID (DeepSORT, ByteTrack, BOT-SORT).
- Re-identification - отдельная CNN или embedding из backbone.

Современные end-to-end трекеры:

- TrackFormer, MOTR, TransTrack - полностью на трансформерах.
- В YOLO26 и RT-DETR есть встроенные трекинговые режимы (tracking mode).

Пример трекинга в ADAS:



Жёлтые пунктиры - tracklets, красные стрелки - предсказанная траектория. Это критично для прогнозирования столкновения за 2–3 секунды.

8. Сравнительная таблица (2025–2026 данные для ADAS)

Модель	Тип	mAP@50–95 (типично)	FPS (на Orin)	Маленькие объекты	Глобальный контекст	NMS-free	Рекомендация для ADAS
SSD	One-stage CNN	72–78 %	40–60	Хорошо	Слабый	Нет	Low-power, legacy
YOLOv12 / YOLO26	One-stage	80–85 %	70–100+	Отлично	Средний+	Да	Универсальный выбор №1
RT-DETR (v4)	Transformer	83–87 %	60–85	Отлично	Отличный	Да	Сложные сцены, Level 3+
Faster R-CNN	Two-stage	85–90 %	8–15	Хорошо	Хороший	Нет	Только оффлайн/ высокая точность

9. Основные вызовы и пути решения
Маленькие/дальние объекты → multi-scale + focal loss + augmentations (Mosaic, MixUp).
Погода/ночь → domain adaptation, synthetic data (CARLA + дождь/снег), multi-modal (камера + радар/лидар).
Реальное время на edge → quantization (INT8), pruning, TensorRT/OpenVINO, дистилляция.
Безопасность → ISO 26262, объяснимость (Grad-CAM, attention maps), adversarial robustness.
10. Перспективы (2026–2030)
Гибридные CNN-Transformer (YOLO26 + local attention).
Foundation models (Segment-Anything-2, GroundingDINO) для zero-shot обнаружения новых классов.
End-to-end восприятие + планирование (BEVFormer, UniAD).
Нейроморфные чипы и spiking neural networks для ультранизкого энергопотребления.

Сегодняшние модели глубокого обучения позволяют создавать системы ADAS, которые в реальном времени с высокой точностью обнаруживают, классифицируют и отслеживают объекты. YOLO26 и RT-DETR - это актуальный фронтир, который уже внедряется в серийные автомобили Level 2+/3. Знание архитектур, сильных и слабых сторон позволит вам выбирать правильную модель под конкретную задачу и аппаратную платформу.