A Holistic View of Perception in Intel. Vehicles

Part IV: Key Takeaways and Future Directions

Objectives

Objectives in Part IV

- Takeaway Messages and Key Insights
- Unaddressed Challenges in Perception
  - Context Awareness
  - Embedded Perception
  - V2X Perception
- Future Research Directions
  - Temporal Processing
  - Sensor Processing Architectures
  - Sensors research
  - Infrastructure + AV Datasets

Objectives

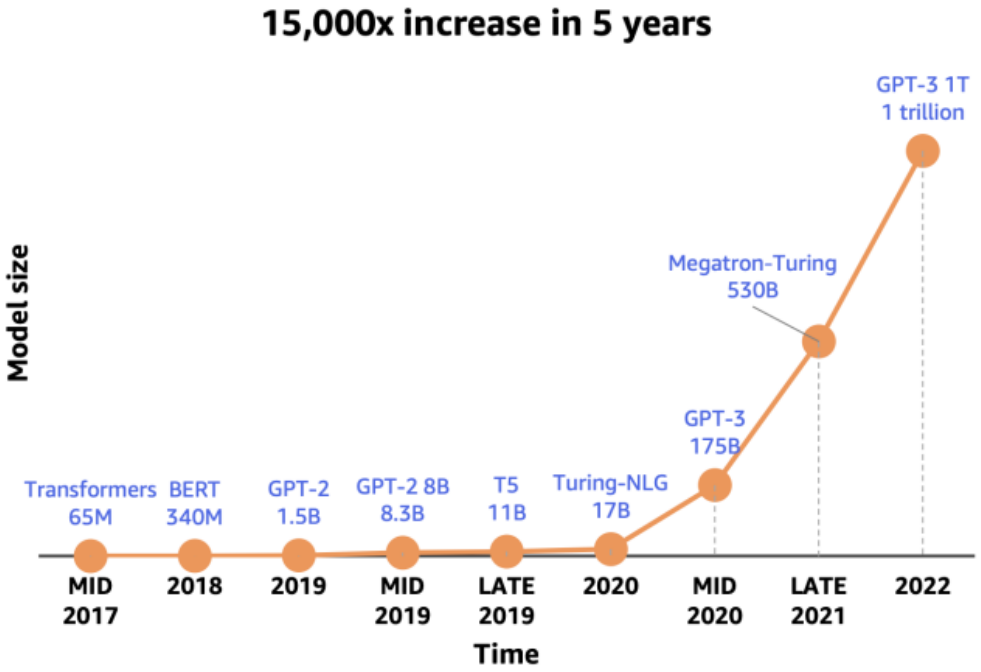
Takeaway Messages and Key Insights

- **Robustness** under challenging conditions, environments, context and surroundings-awareness are **challenges** in AV perception
  - **Deep Learning** provides a **holistic solution** to a number of the above challenges
- **Transfer Learning** and **training at scal**e help to create foundation models
  - **Self-supervised Learning** provides a framework for large scale learning on unannotated data
- It is not always clear if aberrant events and challenges must be incorporated in training
  - Instead, **model predictions** must be equipped with **diagnostic tools** at inference
  - These diagnostic tools are **anomaly and uncertainty scores** for decision making and **contextual explainability** for post-hoc stakeholders
  - **Gradients** provide the change induced by an aberrant event in the network and can be used to obtain the required **prediction diagnosis**

Perception in AVs

Unaddressed Technical Challenges for Level 3 Automation

- Challenging weather
- Challenging sensing
- Challenging environments
- Context awareness
- **Embedded perception**
- V2X perception



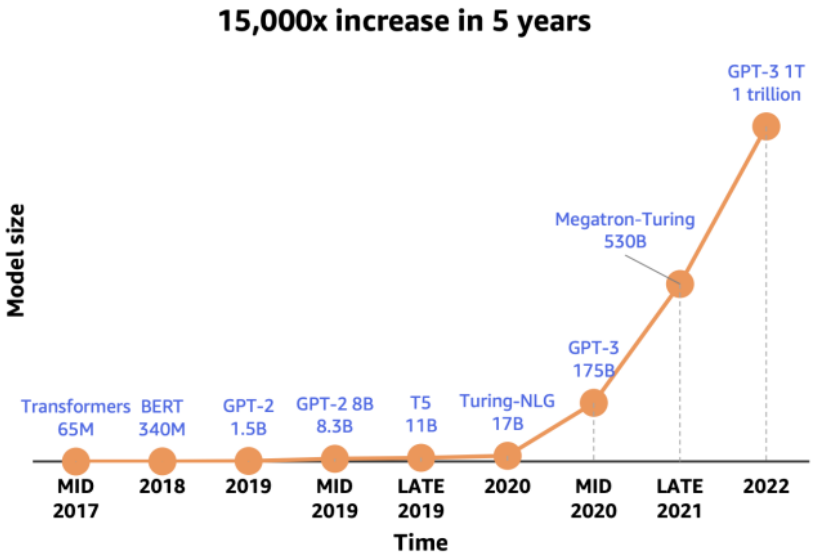
- Foundation models are great but the real-time feasibility is an issue
- The inaccuracies from model outputs is dangerous in urban settings

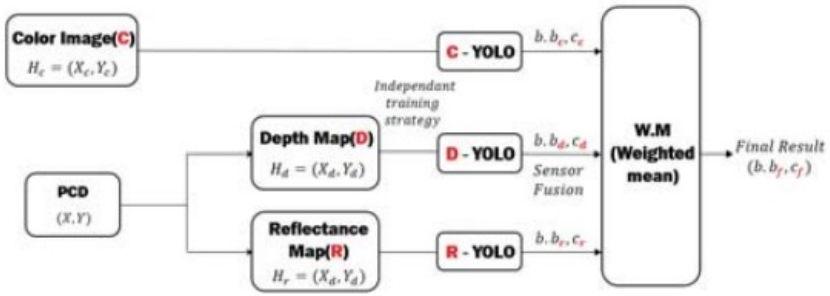
Perception in AVs

Unaddressed Technical Challenges for Levels 4 and 5

Foundation models with multiple sensor modalities

- Challenging weather
- Challenging sensing
- Challenging environments
- **Context awareness**
- **Embedded perception**
- **V2X perception**





- Levels 4 and 5 automation relies on roadside infrastructure to obtain high-resolution predictions
- 10x is the rough estimate of the increase in processing power between levels of automation

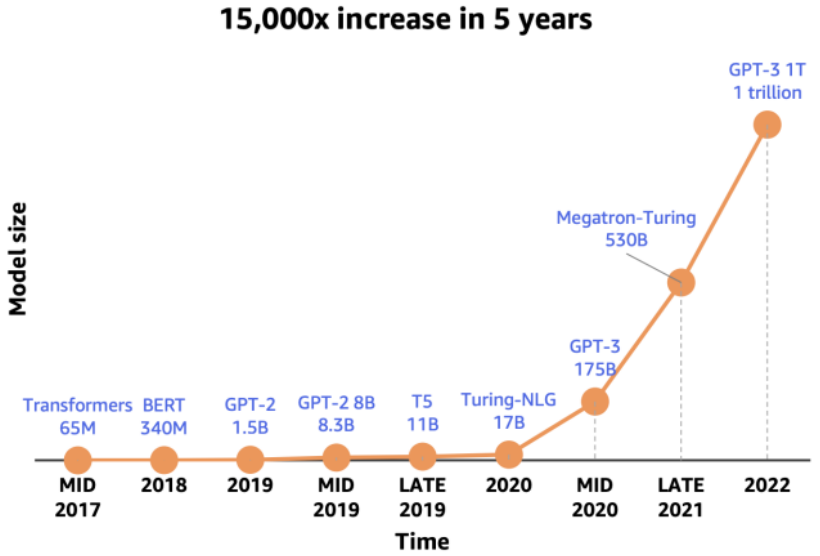
Kim, J., Kim, J., & Cho, J. (2019, December). An advanced object classification strategy using YOLO through camera and LiDAR sensor fusion. In *2019 13th International Conference on Signal Processing and Communication Systems (ICSPCS)* (pp. 1-5). IEEE.

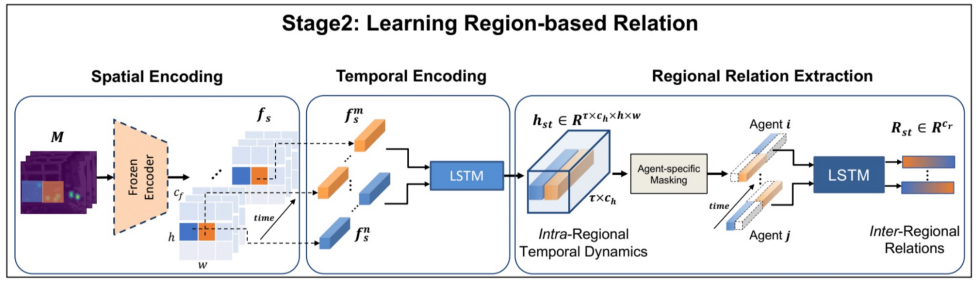
Perception in AVs

Unaddressed Technical Challenges for Levels 4 and 5

Foundation models with multiple sensor modalities and on temporal data

- Challenging weather
- Challenging sensing
- Challenging environments
- **Context awareness**
- **Embedded perception**
- **V2X perception**





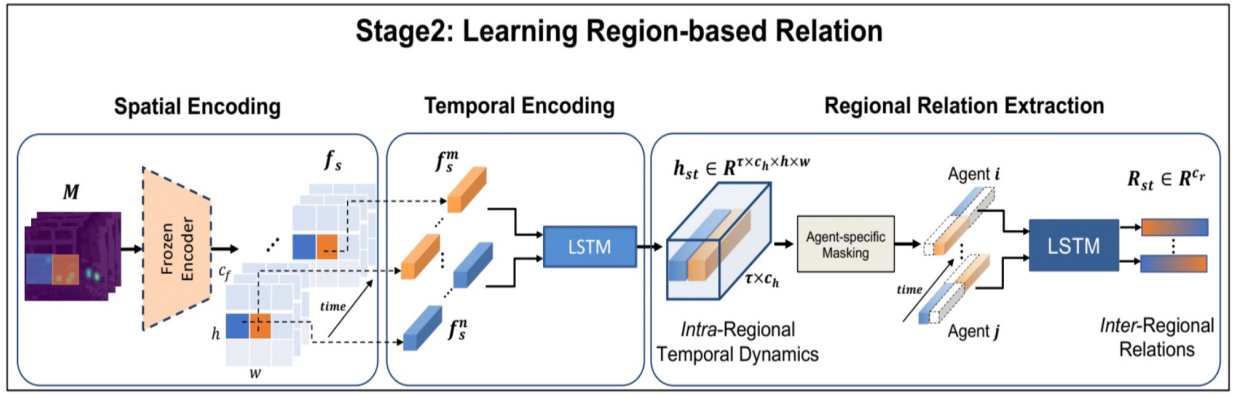
- Levels 4 and 5 automation relies on roadside infrastructure to obtain high-resolution predictions
- 10x is the rough estimate of the increase in processing power between levels of automation
- Current temporal processing = linear spatial processing in time

C. Zhou, G. AlRegib, A. Parchami, and K. Singh, "TrajPRed: Trajectory Prediction With Region-Based Relation Learning," *IEEE Transactions on Intelligent Transportation Systems*, submitted on Dec. 28 2022

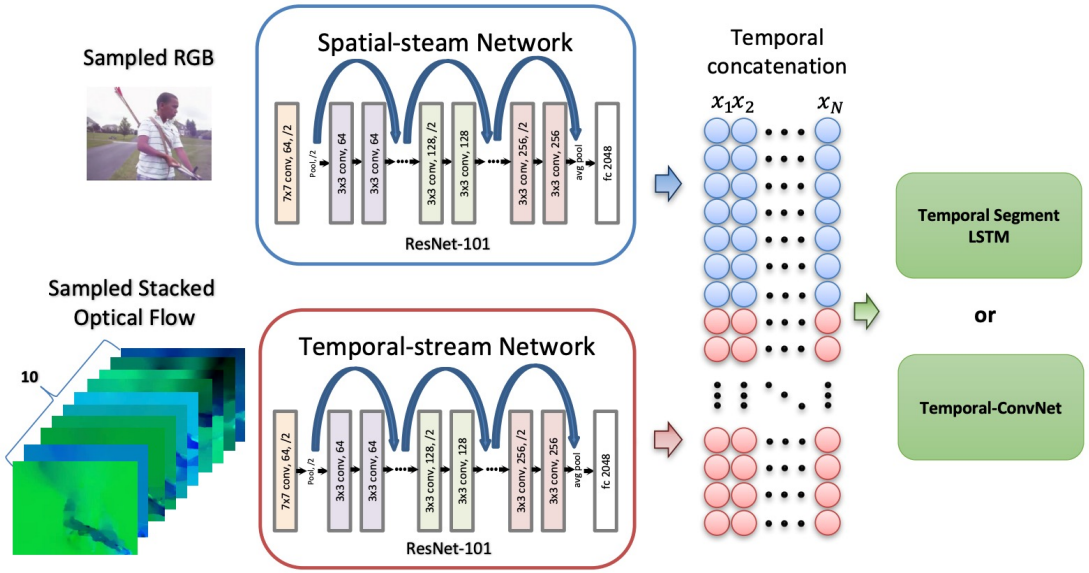
Future Direction 1

Temporal processing of data

Temporal processing ≠ Linear spatial processing



Late temporal fusion: Encode all spatial data in a time-wise fashion and determine temporal relationships



Early temporal fusion: Encode both spatial and temporal information together and fuse them within the network

C. Zhou, G. AlRegib, A. Parchami, and K. Singh, "TrajPRed: Trajectory Prediction With Region-Based Relation Learning," *IEEE Transactions on Intelligent Transportation Systems*, submitted on Dec. 28 2022

Future Direction 2

Sensor processing architectures

Vision data processing was revolutionized by CNNs

Language data processing was revolutionized by Transformers

LIDAR data processing is revolutionized by ?

RADAR data processing is revolutionized by ?

…

Future Direction 3

More data with less sensors!

4 Fisheye cameras provide a 360 degree surround view of the car

Results from Zero-shot (i.e. using the trained model out of the box) Segment Anything Model on Woodscape dataset

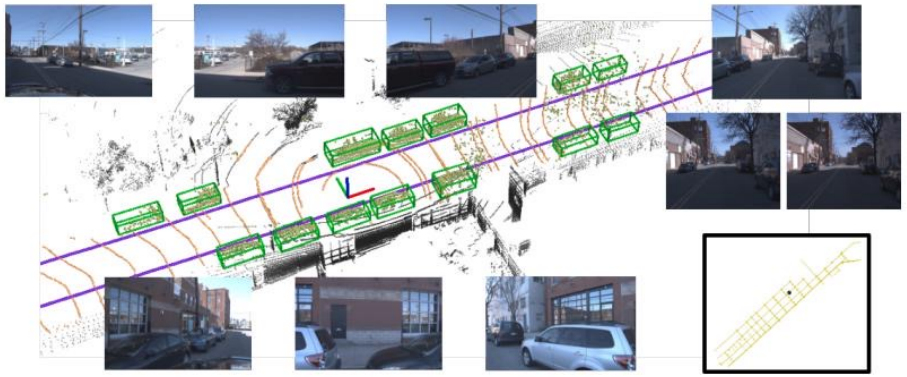
Important context and objects are not segmented

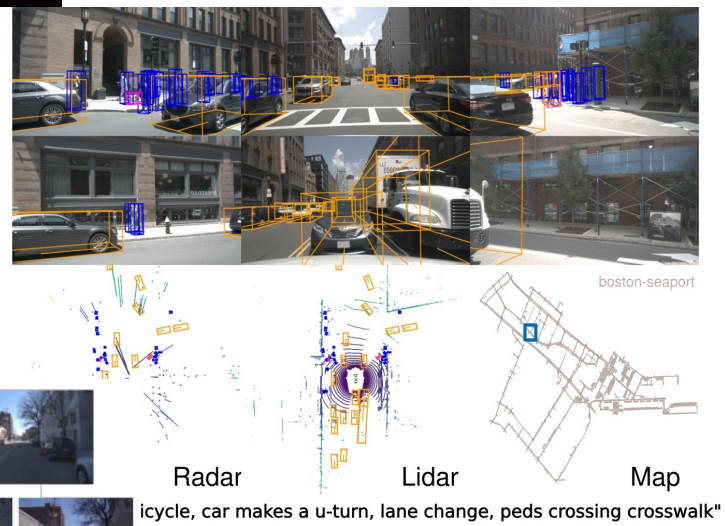
Future Direction 4

Infrastructure + AV Datasets

Abundance of egocentric AV datasets! Dearth of Infrastructure + AV datasets

Argoverse





NuScenes

- Infrastructure datasets: Stationary sensors at traffic junctures, streets, heavy pedestrian traffic areas etc.
- Infrastructure + AV datasets: Egocentric sensors on vehicles + stationary sensors for the same scenes

Some Memes to Wrap it Up