Challenges in Motion Prediction for Autonomous Driving

Yi Yang, KTH, Scania AB

EECS, Division of Robotics, Perception and Learning

1. Hard Cases Detection in Motion Prediction by Vision-Language Foundation Models

Yi Yang, Qingwen Zhang, Kei Ikemura, Nazre Batool, John Folkesson

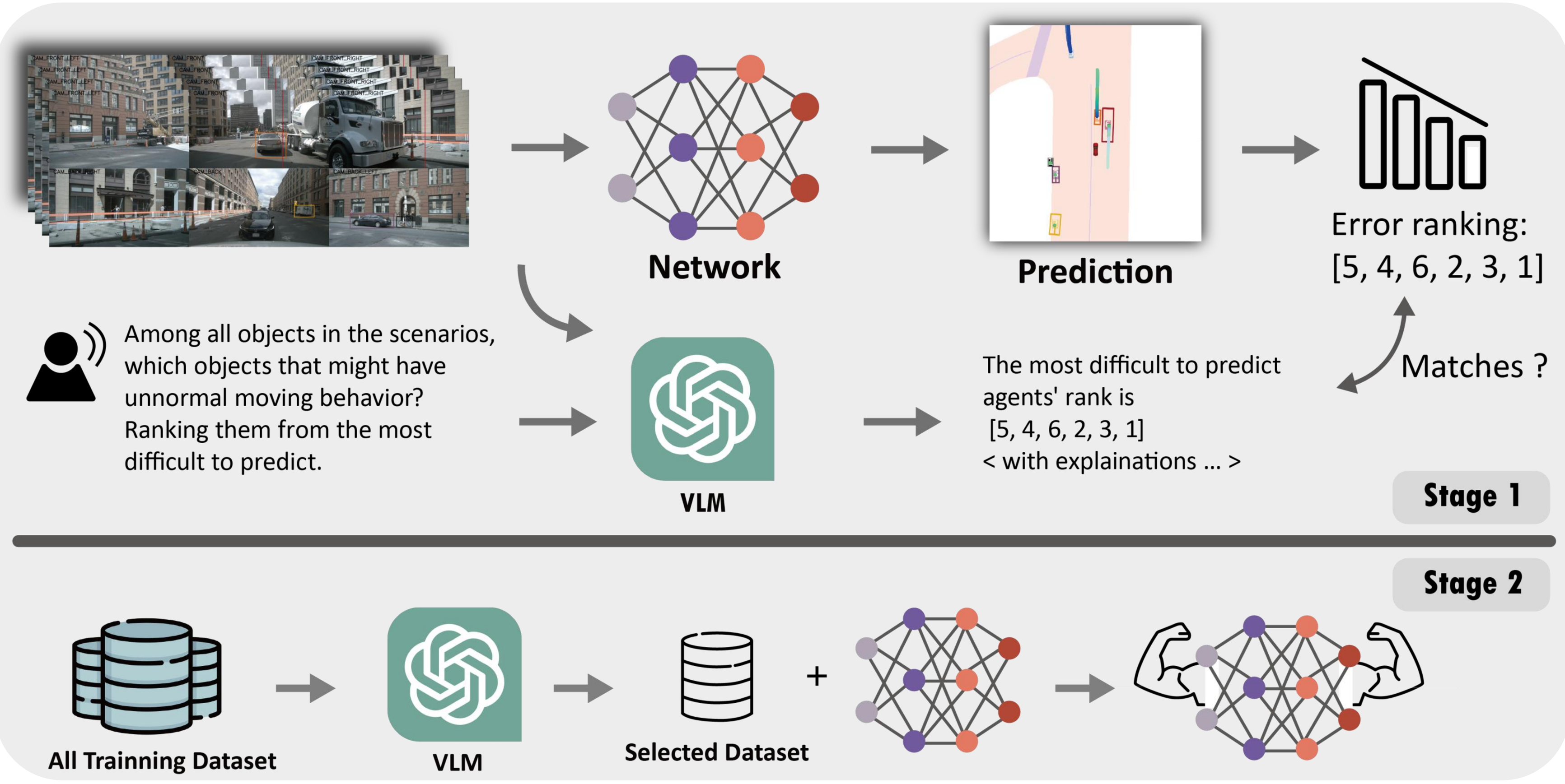
- Addressing hard cases is challenging!
  - o Sparsity & high variability

Anomalous road users *Extreme weather* *Complex traffic*

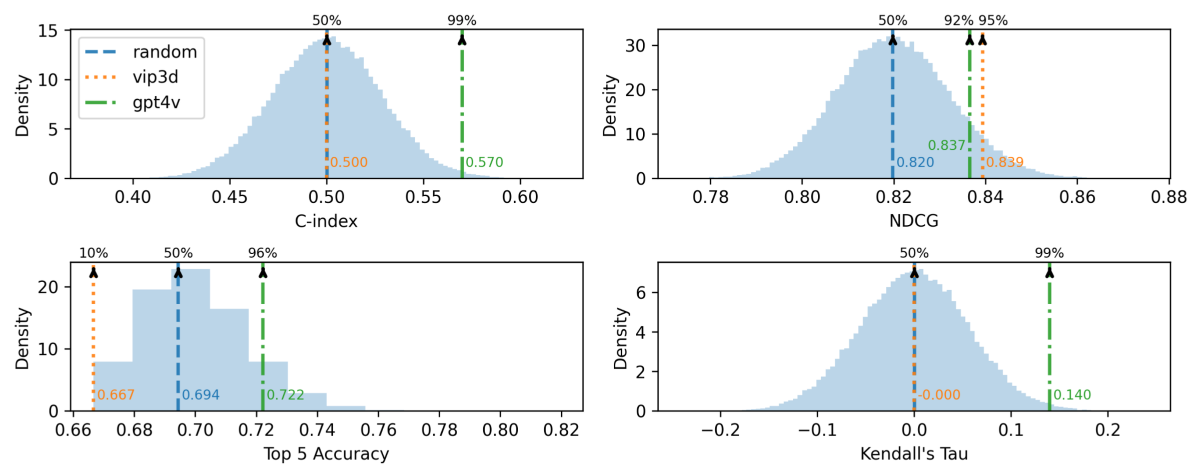
- Existing method:
  - o collect more real-world data? -> *expensive!*
  - o synthetic data?
    - generate with deep generative models conditioned on specific needs
    - manipulate the 3D reconstructed environment, like moving/adding road users

    -> *require much human intervention!*
  - o Incremental learning? -> *dependence on the network training!*

Q:
Is there a more explainable and independent method available?



Stage 1: Agent-level
- o **Verify** the ability of VLM to detect hard cases
- o using existing motion prediction NN as ground truth



Stage 2: Scene-level
- o **Improve** training efficiency by training the network with a smaller subset of data selected by VLM.

Class		Vehicle		Pedestrain	
# Samples [Ratio%]		minADE	minFDE	minADE	minFDE
Whole	28130 [100]	0.71	1.02	0.82	1.11
Random	2000 [7.1]	0.80 ↑13%	1.22 ↑19%	0.93 ↑13%	1.31 ↑18%
	1000 [3.6]	0.82 ↑16%	1.27 ↑24%	0.92 ↑12%	1.29 ↑16%
	500 [1.8]	0.93 ↑31%	1.45 ↑42%	0.99 ↑21%	1.42 ↑28%
	200 [0.7]	0.97 ↑36%	1.55 ↑51%	1.03 ↑25%	1.45 ↑30%
	100 [0.4]	1.08 ↑52%	1.73 ↑69%	1.13 ↑38%	1.63 ↑47%
GPT-4v	200 [0.7]	0.93 ↑31%	1.48 ↑45%	1.03 ↑25%	1.43 ↑29%
	100 [0.4]	0.97 ↑37%	1.56 ↑52%	1.07 ↑31%	1.56 ↑40%

- *4 real examples of GPT4 outputs*

9 "**nighttime** driving and **wet road** surfaces, which can affect visibility and vehicle behavior. The **reflections** and **glare** from the lights..."

7 "at an intersection... There is a **large truck** on the left that may **obstruct the view** and movement... increase the difficulty of prediction due to **potential blind spots** ..."

5 "The **intersection** ahead adds complexity to the driving scenario, but overall traffic **density** is **not high**..."

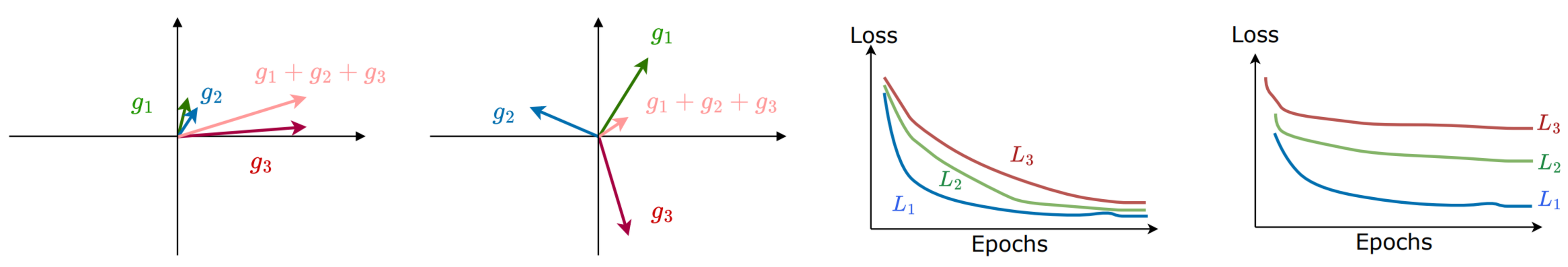
3 "The traffic situation appears to be **straightforward** with **light** traffic and **clear** road markings..."

2. *AutoScale*: Combining Multi-Task Optimization with Linear Scalarization

Yi Yang*, Kei Ikemura*, Qingwen Zhang, Ci Li, Nazre Batool, Sina Mansouri Sharif, John Folkesson

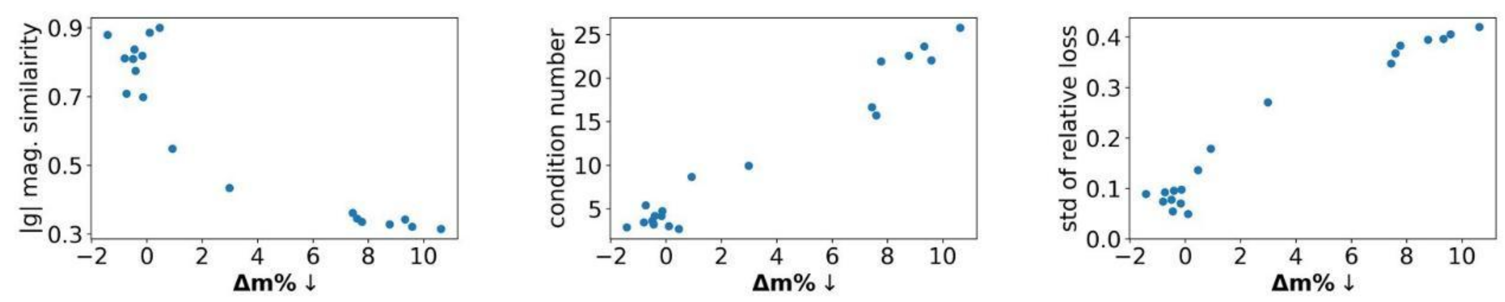
- There exists multiply MTL training issues!
  - Multi-objective learning: gradient & loss

$$\mathcal{L}(\boldsymbol{\theta};\boldsymbol{w}) \equiv \sum_{i=1}^{K} \boldsymbol{w}_i \mathcal{L}_i(\boldsymbol{\theta}), \quad \boldsymbol{w} > 0, \quad \sum_i \boldsymbol{w}_i = 1.$$



(a) Gradient dominance (b) Gradient conflict (c) Imbalanced convergence (d) Imbalanced loss scale

- Surprisingly, we find some metrics serve as good indicators of performance!



(a) Gradient magnitude similarity (b) Condition number (c) Std of relative loss

- Given high performance -> optimal MTO metric values, we hypothesize that the reverse also holds: optimize MTO metric -> high performance. If so, we can localize the optimal task weights by optimizing the key metric value.
- We propose AutoScale, an efficient and practical two-stage pipeline that partitions a single training run into two phases.

$$\underbrace{w^*}_{\text{Optimal weight set}} = \arg\min_{w} \mathbb{E}[\mathbf{F}(w|\underbrace{\{\mathcal{G}\},\{\mathcal{L}\}}_{\text{Given set of gradient and loss}})], \quad \text{s.t.} \quad \sum_{i=1}^{K} w_i = \underbrace{K}_{\text{K loss terms in total}},$$

- Here we test three cost function considering gradient magnitude similarity, loss similarity, and condition number.

WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM