ITS World Congress
Bordeaux, France
5 to 9 October
2015

Recognizing Potential Traffic Risks through Logic-based Deep Scene Understanding

Naoya Inoue[*1], Sosuke Kobayashi[*1], Yasutaka Kuriya[*2] and Kentaro Inui[*1]
*1: Tohoku University
*2: DENSO CORPORATION

TOWARDS INTELLIGENT MOBILITY
Better use of space

Organised by

Co-organised by

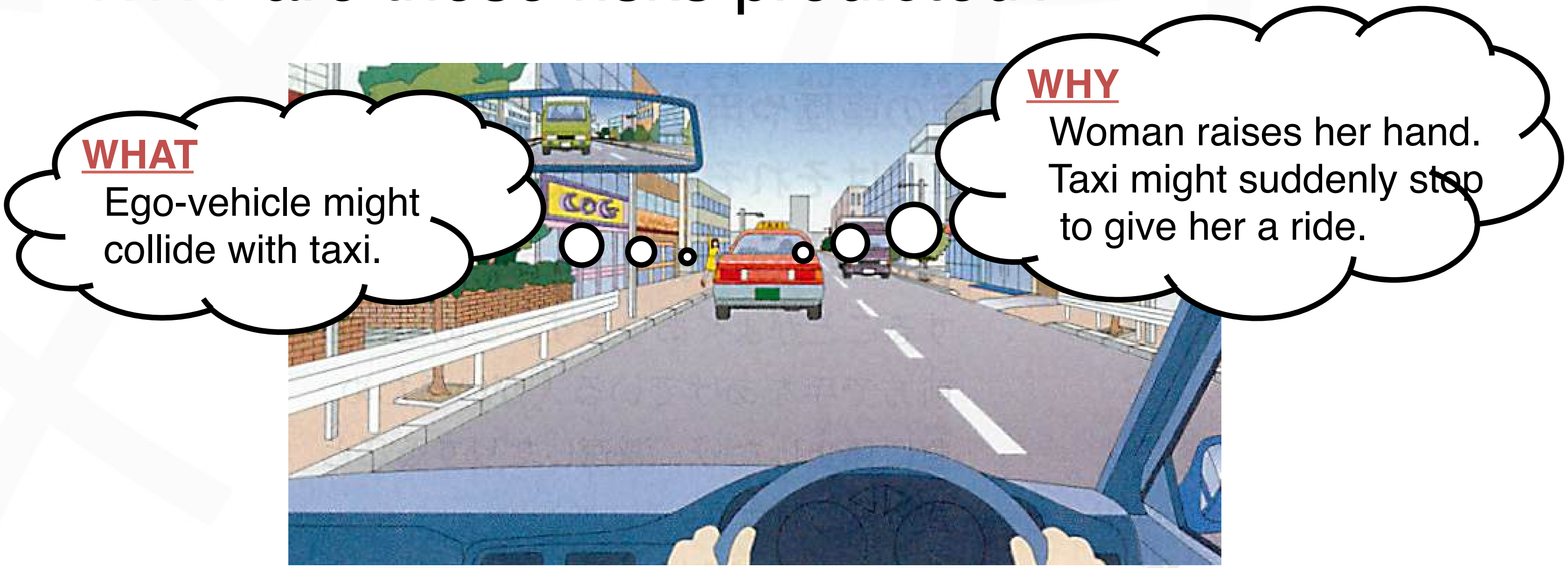
Hosted by

On behalf of

Supported by

Research goal

- Automatic deep understanding of traffic risks
  - WHAT risks can be predicted?
  - WHY are these risks predicted?



[Chubu-Nippon-Driver-School 1999]

  - App.: ADASs, automated driving, etc.

Challenges

General framework for risk prediction [Rendon-Velez, TMCE2008]:

Perception → **Analysis** → Decision → Action

- Perception: pretty advanced!
- Analysis: physics simulation-based approach is explored (e.g., [Broadhurst et al. IV2005]), but:
  - ☹ Not good at long-term prediction: prediction of behavior of traffic agents only depend on physical info. (velocity, position, etc.)
  - ☹ No qualitative explanation of risks: trajectories are explanation; but not sufficient in some situation

Key idea:

knowledge-based commonsense reasoning

X raises X's hand
=> X want to take a taxi

P want to take a taxi?

X see Y want to take a taxi
=> X give Y a ride

T give P a ride?

X give Y a ride
=> X stop

T stop?

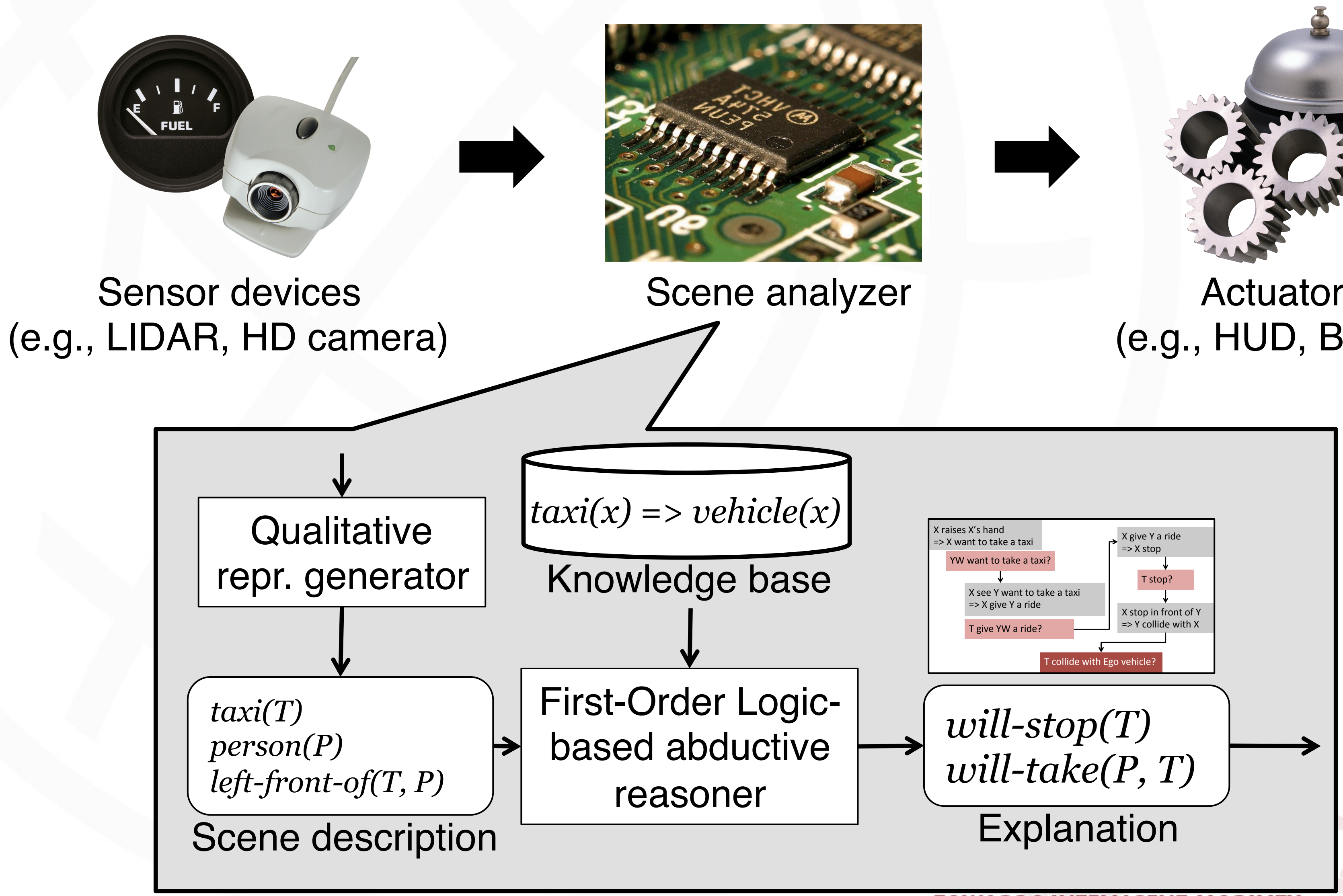
X stop in front of Y
=> Y collide with X

T collide with Ego vehicle?

Explanation
(WHY)

Prediction result
(WHAT)

Overall architecture



Abduction (on First-Order Logic)

- Input B, O:
  - Background knowledge B: set of first-order logical formulae (e.g., ∀x child(x) ⇒ will-rush-out(x))
  - Observation O: set of first-order literals (e.g., {car(C), truck(T), left-of(T, C)})
- Output H^*:
  - Best explanation H^* (set of literals)

$$H^* = \arg\max_{H \in \mathcal{H}} score(H),$$

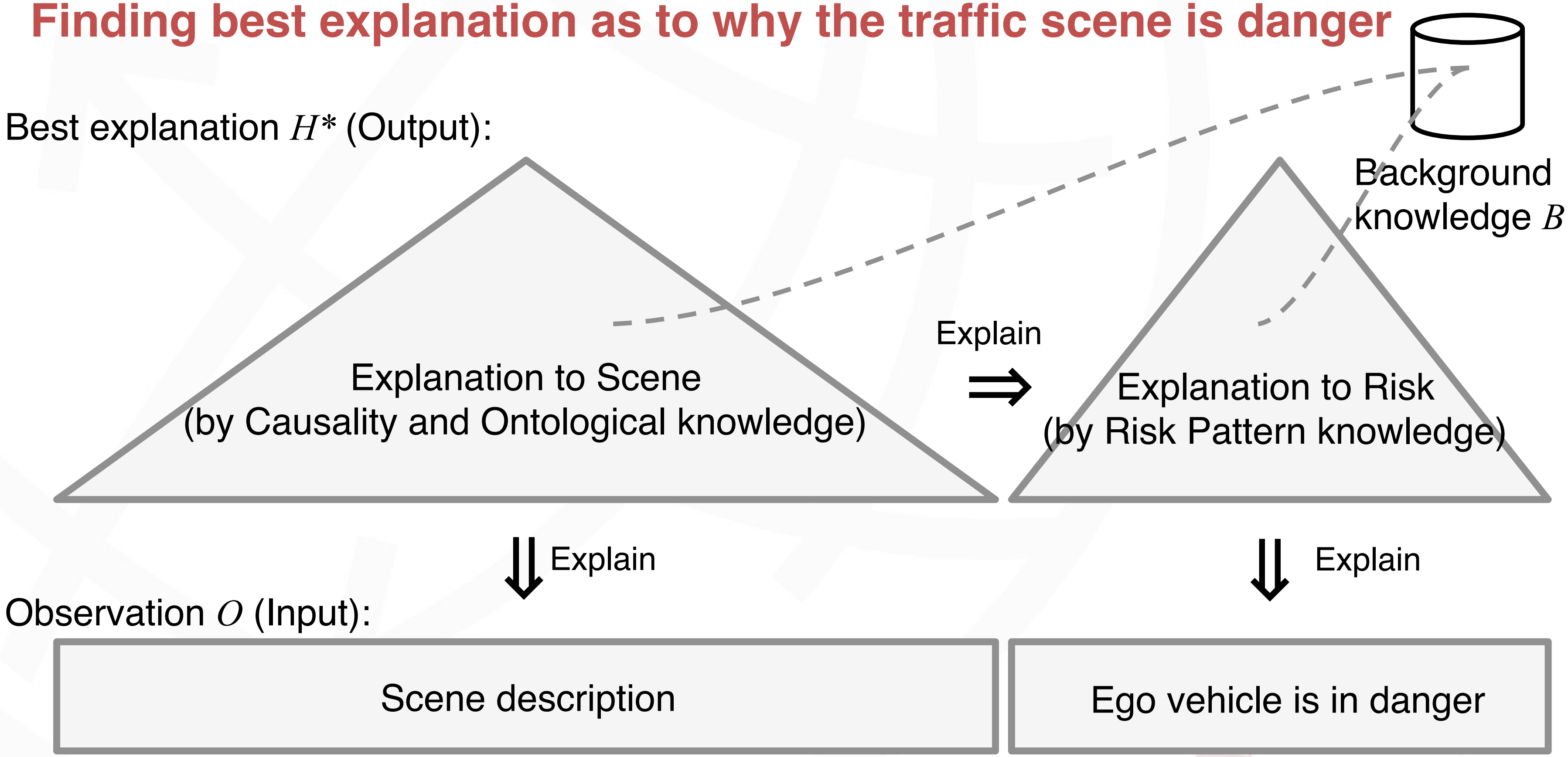
where:

$B \cup H \models O$ (H should entail O w.r.t. B)

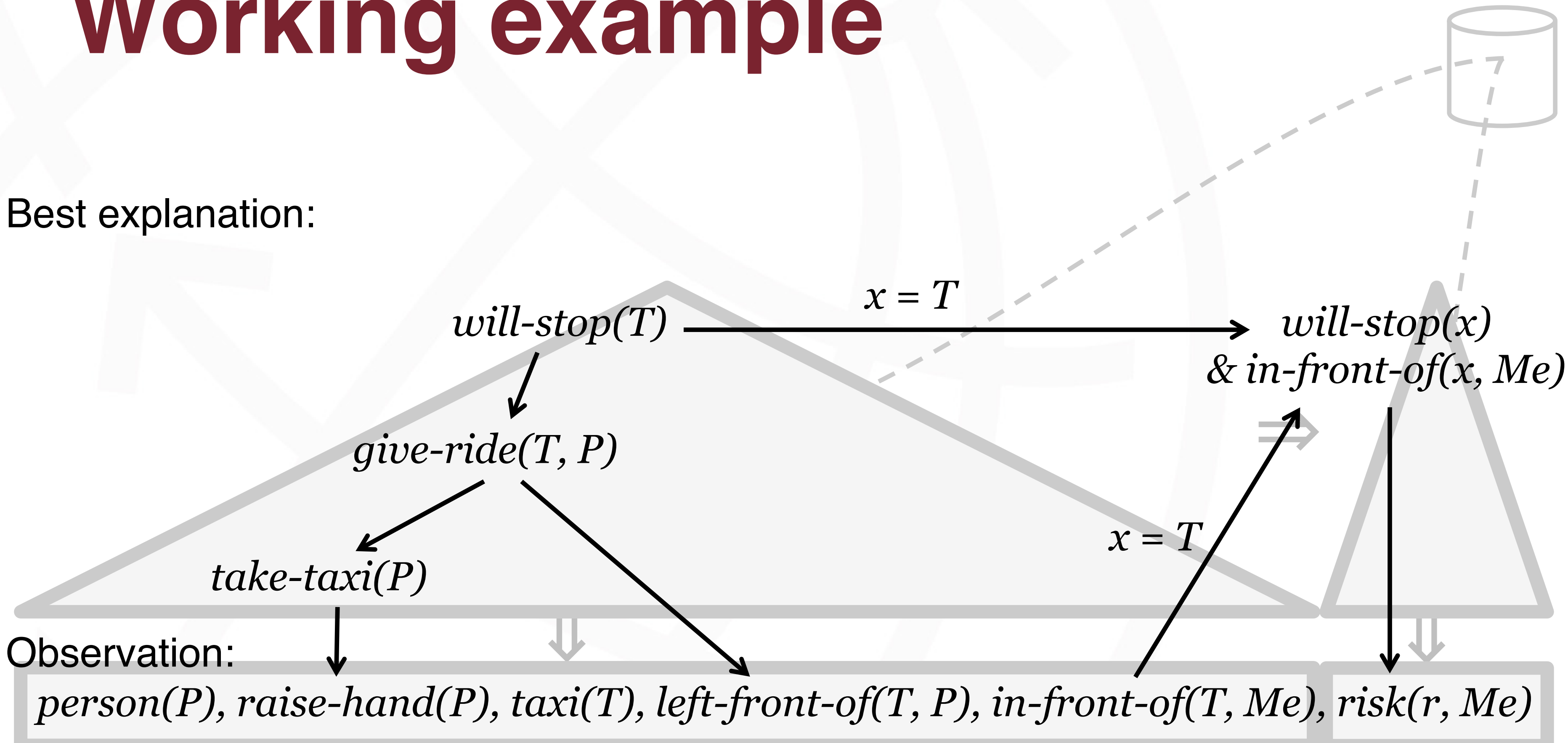
$B \cup H \not\models \perp$ (H should not contradict B)

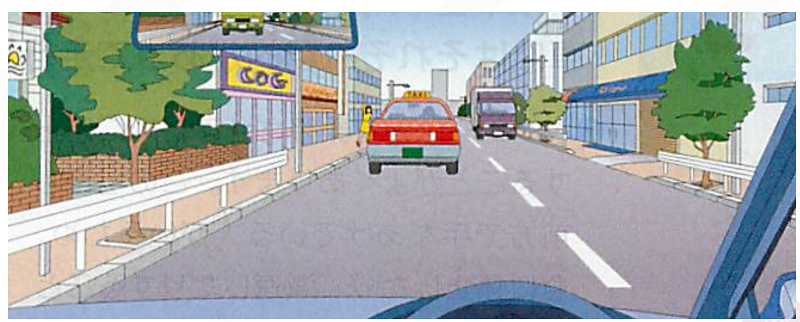
Risk prediction as Abduction

Similarly to Hobbs et al. 1993…
Risk prediction =
Finding best explanation as to why the traffic scene is danger



Working example





Advantages of abductive modeling

- Commonsense reasoning is complex
  - Various kinds of interdependent inferences (e.g., inference of intention, existence of hidden objects, …) are involved
- Hard to find optimal setup of connecting several components for inferences…
- Abduction: declarative problem solving
  - Procedure is not needed to explicitly specify
  - Only knowledge base needs to be given

Knowledge representation

- All background knowledge/observation are written based on the following predicates

Type	Example	Description
Type of object	*taxi(x)*	x is taxi
Status of object	*icy(x)* *left-head-lamp-on(x)*	x is icy x 's head lamp is on
Relative position	*left-front-of(x,y)*	x is left front of y
Intention	*will-stop(x)*	x will stop
Risk	*risk(r, x)*	x is in danger

Inference rules

- 100+ rules are manually induced from textbook for risk prediction [Chubu-Nippon-Driver-School 1999]

Type	Example	Description
Causality	*large-vehicle(x) & in-front-of(Now, x, y)* $\Rightarrow$ *will-avoid(y)*	If large vehicle x is in front of y, then y will avoid it
Ontological	*bicycle(x)* $\Rightarrow$ *vehicle(x)* *car(x) & bicycle(x)* $\Rightarrow \perp$	- x is vehicle - car and bicycle are mutually exclusive concept
Risk pattern	*in-front-of(Now, x, y) & will-stop(x)* $\Rightarrow$ *risk(r, y)*	If x in front of y will stop, then y is in danger.

Evaluation

- Dataset
  - "Master of your driving", textbook for risk prediction [Chubu-Nippon-Driver-School 1999]
    - 93 problems
  - Web training materials for risk prediction:
    - 100 problems
  - Each problem contain: traffic scene (picture) and (2-3) expected traffic risks
- 10-fold cross validation
- Abductive inference engine:
  - Phillip [Yamamoto et al. IJMLC2014]
- Score function:
  - Weighted linear model + Soft Exact Confidence Weighted Learning [Wang et al. ICML2012]

Setting

- Evaluation measures
  - Precision = $\frac{\text{\# of problems where model predicts risk correctly}}{\text{\# of problems where model predicts risk}}$
  - Recall = $\frac{\text{\# of problems where model predicts risk correctly}}{\text{\# of all problems}}$
  - F-score = $\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$
- Compare with 3 baseline models:
  - random: naïve system that randomly chooses person/vehicle in traffic scene
  - majority: naïve system that says all people and vehicles in traffic scene would rush out
  - SVM: Ranking Support Vector Machines [Joachims KDD2003]-based risk prediction system

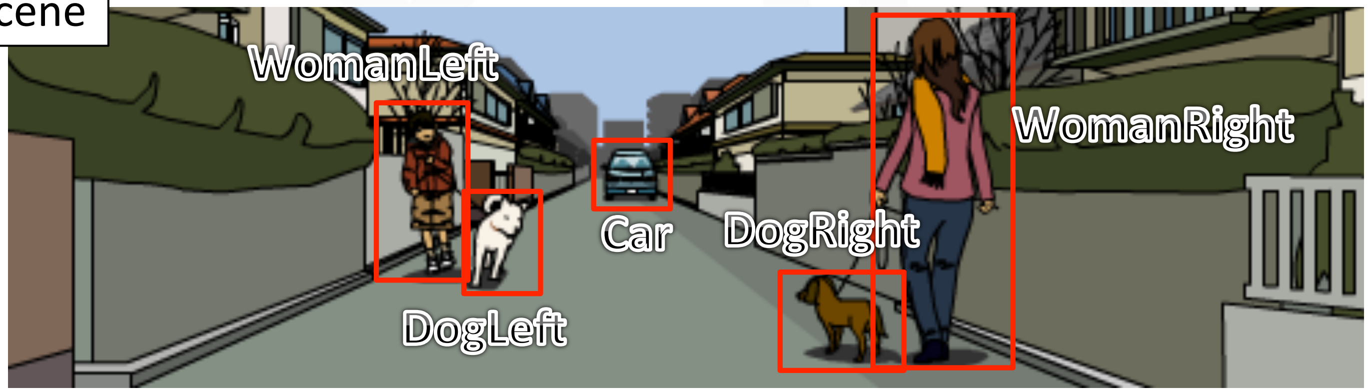
Results

Model	Precision@k	Recall@k	F
Baseline (random)	2.0 (2/100)	1.2 (2/161)	1.5
Baseline (majority)	22.6 (95/420)	**59.0 (95/161)**	32.7
Baseline (SVM, k=1)	30.0 (30/100)	18.6 (30/161)	23.1
Baseline (SVM, k=2)	30.5 (61/200)	37.9 (61/161)	33.9
Baseline (SVM, k=3)	28.0 (84/300)	52.2 (84/161)	**36.5**
Abduction (k=1)	**31.5 (39/124)**	24.2 (36/161)	27.4
Abduction (k=2)	30.3 (59/195)	36.6 (69/161)	33.1
Abduction (k=3)	22.5 (62/276)	38.5 (78/161)	28.4

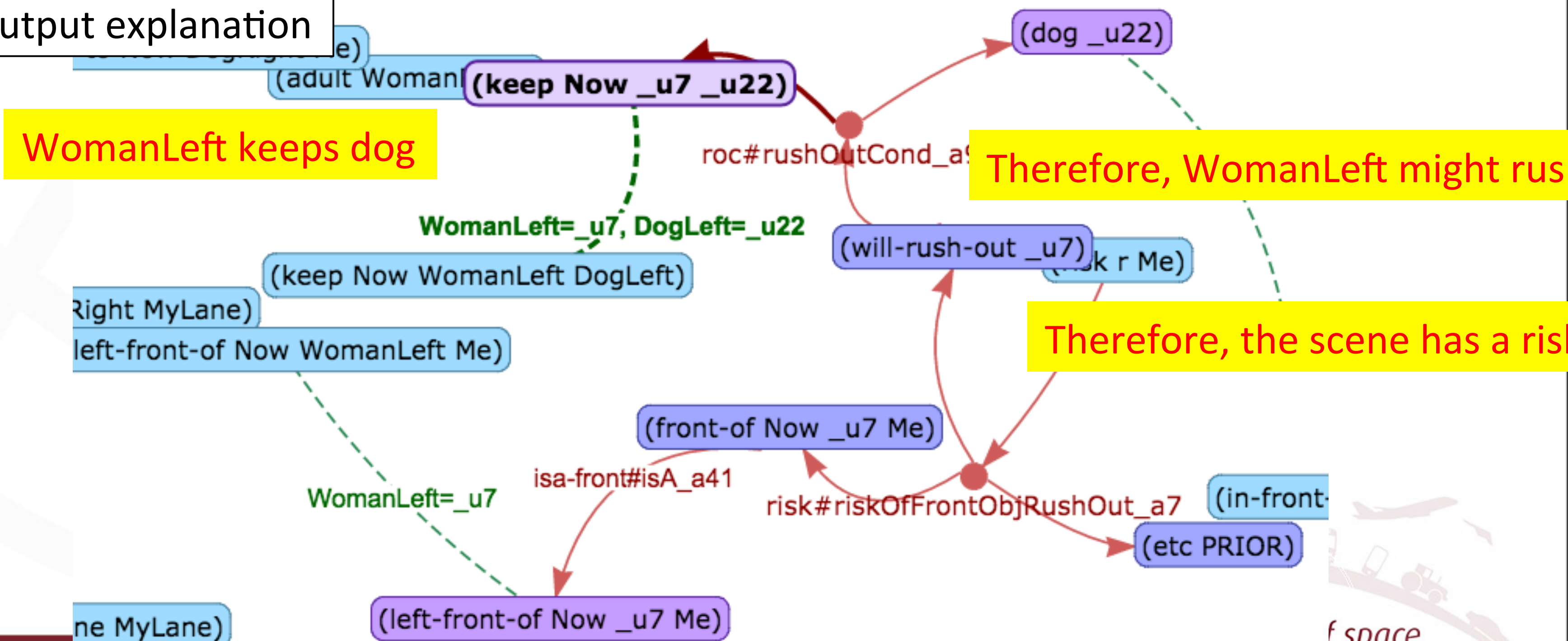
- Proposed model did not outperform baseline models very much
- Why:
  - Lack of physical information (e.g., precise position of pedestrians)
  - Knowledge base is not generalized well

Example inference result

Input scene



Output explanation



Summary

- Proposed abductive reasoning-based model for deep understanding of traffic scenes
- The experiment shows the potentiality of proposed model; still, there's a lot of room for improvement
- Future work
  - Integration of quantitative reasoning (e.g., physics-simulator) with current framework
  - Enrichment or generalization of knowledge base
  - Test with real sensor devices and actuator