FY2024/2024年度

Robotics for Infrastructure Management Team
Takayuki Okatani
インフラ管理ロボット技術チーム　岡谷 貴之

RIKEN

Center for Advanced Intelligence Project

当チームの目標と研究内容

目標と背景

- 目標：橋梁などインフラ構造物の維持管理をAI・ロボット技術で効率化・高度化
- 技術課題①：画像等を自動撮影可能なロボット ← 世界中で進む自律UAVを活用
- 技術課題②：画像等を元に**対象物の状況を理解し，説明できるAI**の開発

⇩

- マルチモーダルAI（GPT4vなど）が大きく進展し，実社会応用への適用が期待

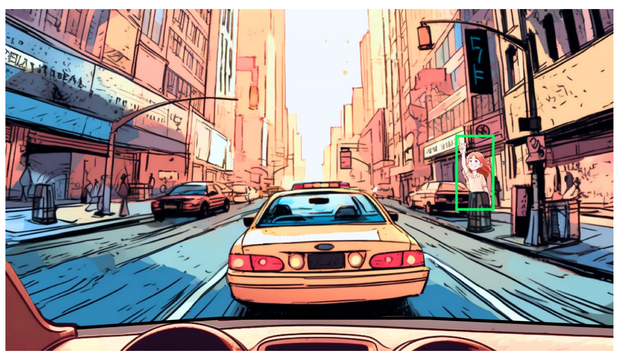
研究成果（2023年発表）

(本ポスターの内容)

- マルチモーダルAIによる自動車運転危険予測 [Charoenpitaks+, IEEE Transactions on Intelligent Vehicles誌]
- 無教師ドメイン適応画像セグメンテーション [Wang+, Pattern Recognition Letters誌]
- 高ダイナミックレンジ画像生成 [Ye+, Computer Vision and Image Understanding誌]

マルチモーダルAIによる運転時危険予測

- 近い将来に発生しうる潜在的な危険の予測
  →視覚情報を用いた「仮説推論」が必要

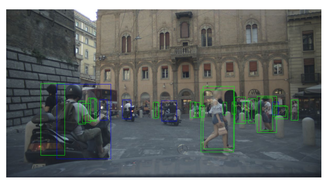


- 現在のマルチモーダルAIに，これを実現できるのか？
- 問題の単純化：視覚入力 = 単一の静止画像

データセットの構築

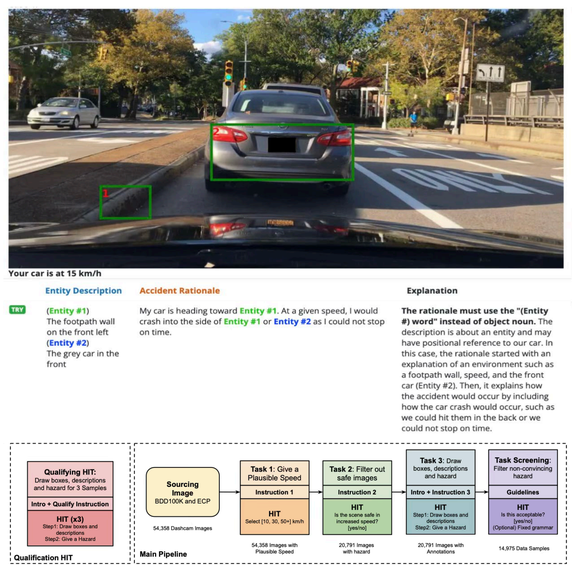
- 実際の事故のデータを収集するのは困難かつそれほど有用でない
  → 既存のドライブレコーダー画像データセットを流用，仮説上の危険をアノテーション

BDD100K [Yu+2020]

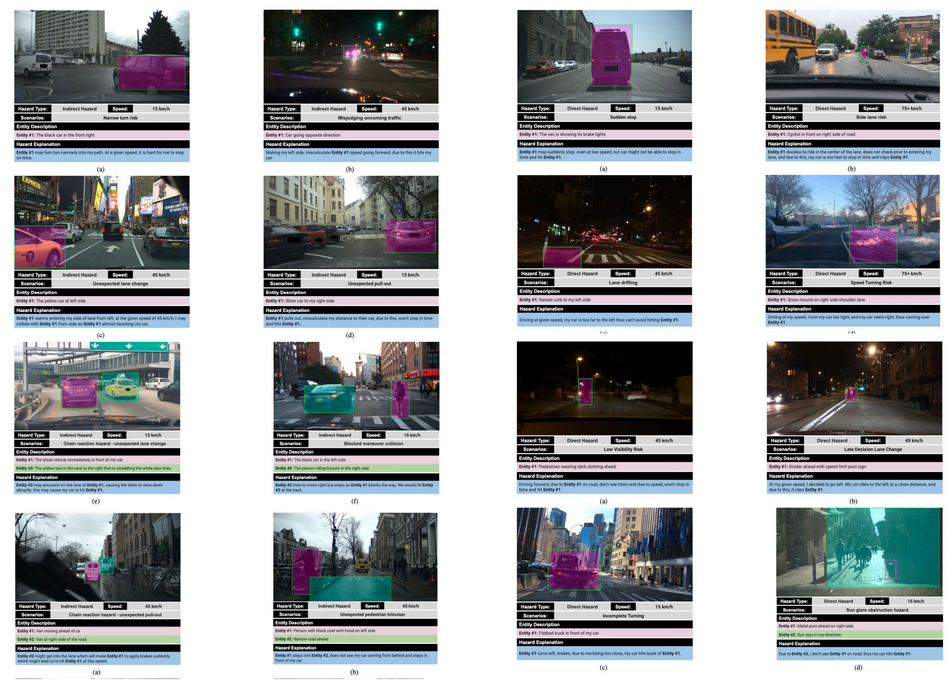


ECP(EuroCity Persons) [Braun+2019]

amazon mechanical turk

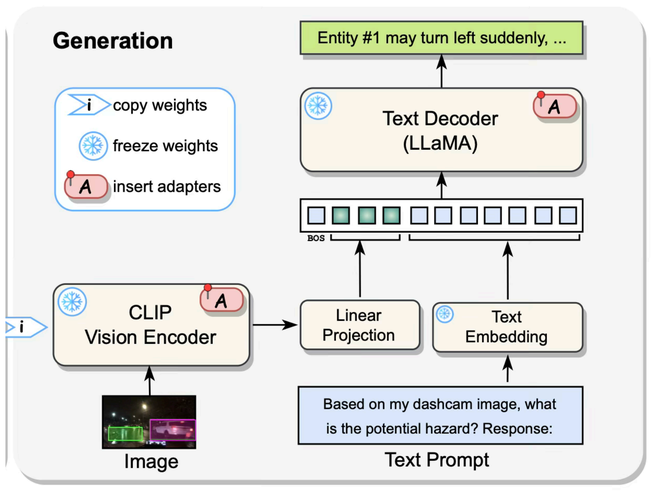


Driving Hazard Prediction and Reasoningデータセット



タスクの定義とモデルの設計

- タスク: 入力画像に対する危険説明文の生成
  - 入力: 画像+物体が色付きのバウンディングボックスで指定
- モデル: VLM = 事前学習済みLLM + 画像エンコーダ
  - 使用LLM: 「LLaMA-2 7B」
  - 学習: LLMおよび画像エンコーダに挿入したアダプタのみを学習



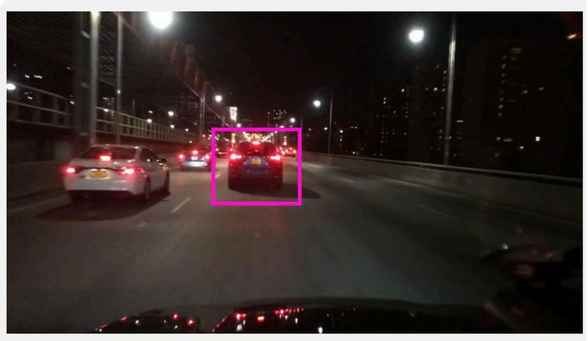
実験結果

- ある程度危険予測を実行できることが確認された
  - 提案手法は，GPT4を含む一般的なマルチモーダルAIを上回る
- 改善の余地も残す

Text Decoder	Generation Task B4 ↑	R ↑	C ↑	S ↑	GPT-4 ↑
-	-	-	-	-	-
BERT	12.6	32.9	34.9	30.3	39.34
OPT 6.7B	**18.7**	**42.7**	38.9	35.4	50.52
LLaMA-2 7B	14.9	36.9	34.5	30.9	56.2
GPT-4	0.3	19.0	0.9	7.2	50.0
LLaMA-2 7B	16.9	39.5	**49.1**	**39.6**	**58.5**

結果の例

Hazard Generation Example: Speeding & Braking

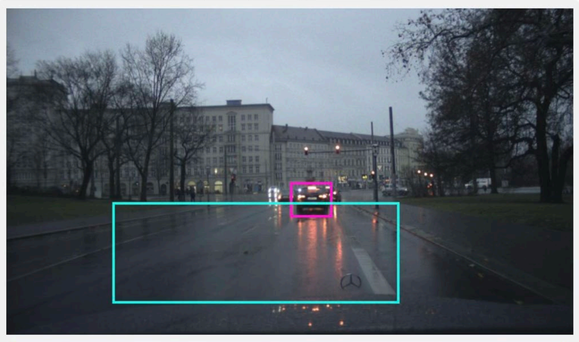


Annotation:
Entity #1 brakes, due to speed, my car won't stop in time and crash into the back of **Entity #1**

Our:
Entity #1 brakes, does not factor my speed, due to speed, can't stop in time and rear end **Entity #1**

GPT-4V:
Our car, traveling at a speed of 45 km/h, could potentially be involved in a rear-end collision with **Entity #1** if it were to suddenly brake and we did not have sufficient time to react and stop.

Hazard Generation Example: Speeding & Braking



Annotation:
Entity #1 brakes, due to my speed and **Entity #2**, my car won't stop in time and rear ends **Entity #1**

Our:
Entity #1 brakes, due to speed and **Entity #**2, my car hits **Entity #1**

GPT-4V:
In the hypothetical scenario, traveling at a speed of 75+ km/h in wet road conditions, our car could potentially lose traction and control, leading to an accident with **Entity #1** (Magenta box: the car directly ahead). The likelihood of a collision increases if **Entity #1** were to make a sudden stop and our car's stopping distance were insufficient due to the high speed and slippery road surface.

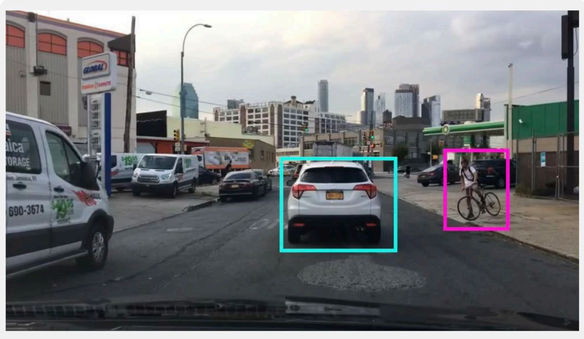
Hazard Generation Example: Merging Maneuver

Annotation:
Entity #1 pulls out from parking spot, does not check mirror or factor my approaching speed, due to this, I can't stop in time and hits **Entity #1**

Our:
Entity #1 pulls out from parked position, does not check for oncoming traffic, due to this will hit **Entity #1**

GPT-4V:
As our car was traveling at 45 km/h, **Entity #1**, highlighted by the magenta box, made a sudden stop to avoid an obstacle not visible in the image, causing our car to react too late and resulting in a rear-end collision with **Entity #1**.

Hazard Generation Example: Pedestrian



Annotation:
Entity #1 decides to go behind **Entity #2** to cross the street, misjudges my speed, won't be able to stop in time and hits **Entity #1**

Our:
Entity #1 may try to go around **Entity #2** by coming into my lane. I may not be able to stop in time and hit **Entity #1**

GPT-4V:
As our car approached the intersection at 45 km/h, **Entity #2** (Cyan box) unexpectedly changed lanes without signaling, causing our driver to swerve to avoid a collision. This sudden maneuver led our car into the path of **Entity #1** (Magenta box), a cyclist crossing the street, resulting in an accident.