

卒業論文 2020 年度 (令和 2 年度)

# ドライブレコーダーを用いた渋滞推定手法の提案

指導教員

慶應義塾大学環境情報学部

村井 純

楠本 博之

中村 修

高汐 一紀

Rodney D. Van Meter III

植原 啓介

三次 仁

中澤 仁

武田 圭史

慶應義塾大学 環境情報学部

李 広耀

*koyo@ht.sfc.keio.ac.jp*

## 卒業論文要旨 2019 年度 (令和元年度)

### ドライブレコーダーを利用した渋滞推定システム

#### 論文要旨

日本における渋滞情報は主に道路に設置されているセンサーや監視カメラの映像等からリアルタイムに算出されており、そのような情報は自動車のナビゲーションシステム等で使われている。しかしそのようなシステムは一部の道路に搭載されておらず、情報に限りがある。この研究では、近年のドライブレコーダー搭載数の増加や映像技術向上を生かし、ドライブレコーダーから得られる単眼カメラ映像から渋滞推定を行うことで、全ての道路での渋滞推定を可能にすることを目指す。実装にはドライブレコーダーから得られる単眼カメラ映像を用いて、距離推定ライブラリ及び物体検出システムを利用し、先行車との距離をもとに渋滞しているか否かを推定する。

#### キーワード

ドライブレコーダー, 渋滞, 深層学習, 画像処理

慶應義塾大学 環境情報学部

李 広耀

## **Abstract of Bachelor's Thesis Academic Year 2020**

### **Congestion Estimation System Using a Drive Recorder**

#### **Abstract**

Congestion information in Japan is mainly calculated in real time from road sensors and surveillance camera images, and such information is used in car navigation systems. However, such systems are not installed on some roads and information is limited. This study aims to take advantage of the recent increase in the number of drive recorders and improvements in video technology to estimate traffic congestion on all roads by using monocular camera images obtained from drive recorders. The implementation uses monocular camera video from drive recorders, a distance estimation library and an object detection system to estimate whether or not a vehicle is congested based on the distance from the vehicle ahead.

#### **Keywords**

**Keio University Faculty of Environment and Information Studies.  
Koyo Ri**

# 目次

第 1 章	はじめに	1
1.1	背景 . . . . .	1
1.2	目的 . . . . .	1
1.3	ドライブレコーダーについて . . . . .	2
1.4	渋滞の定義 . . . . .	2
第 2 章	現在の渋滞に関する問題	3
2.1	渋滞についての問題点 . . . . .	3
第 3 章	関連研究	4
3.1	ドライブレコーダーを用いた研究 . . . . .	4
3.2	ドライブレコーダー以外の車載カメラを用いた研究について . . . . .	4
3.3	渋滞推定の研究について . . . . .	4
3.4	カメラを用いた距離推定の研究 . . . . .	5
3.5	物体検出システムについて . . . . .	5
第 4 章	アプローチ	6
4.1	研究手法 . . . . .	6
4.2	ネットワークアーキテクチャ . . . . .	7
4.2.1	自動車を検出するフェーズ . . . . .	7
4.2.2	距離を推定するフェーズ . . . . .	7
4.3	渋滞の評価について . . . . .	7
第 5 章	実験	9
5.1	データーセット . . . . .	9
5.2	予備実験 . . . . .	9
5.2.1	予備実験 . . . . .	9
5.2.2	予備実験 - 評価 . . . . .	9
5.3	本実験 . . . . .	11
第 6 章	評価	12
6.1	考察 . . . . .	12
第 7 章	結論	13



# 図目次

5.1	FCRN-DepthPrediction を用いた予備実験 . . . . .	10
5.2	Google Tensorflow を用いた予備実験 . . . . .	10

## 表目次

# 第 1 章

## はじめに

本章では、はじめに本研究における背景とその目的等について述べる

### 1.1 背景

日本はアメリカや中国などの道路と比較して国土が狭く、人口密度が高い。それに伴って車道が狭く、渋滞が起こりやすくなっている。そのため、日々テレビやラジオのニュースで渋滞情報が報道されている。

日本では、そのような渋滞情報は、VICS 等の企業が道路に設置されているセンサーや車道付近に置かれているカメラの映像、および自動車に搭載されているセンサーや GPS の情報などをもとに算出されている。そのような情報をもとに自動車のナビゲーションシステム等では到着予想時刻が概算される。また、そのようなデータは Google 社が提供している Google Map にも提供されており、Web 上でリアルタイムのデータをいつでも閲覧することが可能である。しかし、この渋滞情報は国道や高速道路といった、日常的に交通量が多い道路には設置されているが、そうではない、細い道や一方通行といった道では渋滞情報がない。仮にそのような場所に車が密集すると、予想しない渋滞に巻き込まれてしまうことが起こり得る。

また近年、煽り運転等のマナーの悪い運転が報道されるようになり、それに伴って各自動車へのドライブレコーダー搭載数が年々上昇している。ドライブレコーダーは走行中の記録を撮ることにより事故の時の証拠を残すことができると同時に、事故等の証拠にもなり、警視庁も交通安全の面で、ウェブサイトにおいて取り付を推奨している。元々、煽り運転等の危険運転はドライブレコーダーが登場する前から発生していたが、近年社会問題として取り上げられているのはドライブレコーダーやスマートフォンなど、映像技術の向上により人々が気軽に高画質な映像を録画、記録することが可能になったからである。それに加えて近年の SNS の普及により、一般の人でも気軽に意見、画像、動画を発信できるようになった。そのような背景もあり、煽り運転のような危険運転が容易に可視化され、ドライブレコーダーや、スマートフォンで撮影した映像がニュースでも取り上げられるようになった。このような社会的変化を考えれば、ドライブレコーダーを単純な映像記録装置として扱うのでは不十分である。さまざまな機能をつけることで、より利用価値を高める方がより効果的な使い方だといえる。

### 1.2 目的

この研究の目指すところは、入力された画像および動画から、走行している場所が渋滞しているか否かを判断することにある。しかし、このシステムが目指すものは、その判断、評価のプロセスの次のフェーズとし



て、渋滞している場合、その実際に渋滞している現場の画像及び映像をサーバーにアップロードすること、さらにその後のフェーズとして、ドライバー同士でそのような渋滞情報を共有することを実装することである。交通の安全上、ドライバーは運転中にカーナビ等の電子機器を操作することは禁止されているため、システムが渋滞だと判断した場合、自動でそのデータをアップロードする必要がある。また、その渋滞データはオープンな情報としてドライバーに共有されることが必要である。以上の2つのフェーズを最終的に加えることで、このシステムの本来目指しているものは完成する。このシステムが実現すると、現在の VICS 等の渋滞、交通量を検知、測定するシステムでは賄えないような、細い道、住宅街の道、一方通行など、さまざまな道路で突発的に発生した渋滞情報を共有することができる。

### 1.3 ドライブレコーダーについて

現在、一般的なドライブレコーダーは単なる記録装置として活用されているが、ドライブレコーダーの映像からは先行車や道路の状態、歩行者など、さまざまなデータを取得することが可能であり、このようなデータを集めることで様々な事象を可視化することができることが考えられる。この研究ではそのようなドライブレコーダーを使用して渋滞を推定することで、VICS や Google Map、そして GPS の情報だけでは見ることができない道路の渋滞の現場を可視化することを目的とする。VICS では車両感知器 (Vehicle Detectors) と光学式車両感知器の2つが主に使われ、それぞれ通過車両台数や渋滞情報を自動的に感知し交通管制センターへ送信する役割と、通過車両を感知するとともに車載装置との双方向の通信を行う役割を持っている。さらに、VICS は主に国産の自動車メーカーに取り付けられているカーナビゲーションシステムで主に利用されているため、外国産の車など、自動車によってはカーナビゲーションシステムが VICS に対応していない例もある。そして、VICS は日本でのみ使われているシステムであるのに対し、ドライブレコーダーはどのような自動車でも追加で取り付けることが可能であり、渋滞推定を行うことができれば、日本に限らず世界中の道路で渋滞推定を行うことができる汎用性がある。

### 1.4 渋滞の定義

渋滞についての研究を行うためには、まず渋滞とはどのような状態か定義する必要がある。普段我々がニュース等で情報を得ている渋滞情報として、日本道路交通情報センター (JATIC) は、高速道路では、時速 40km 以下で低速走行あるいは停止発進を繰り返す車列が、1km 以上かつ 15 分以上継続した状態であるとし、一般道では、時速 10km 以下で低速走行している状態が渋滞であると定義している。

この研究では正確な速度をカメラ映像から割り出すことは不可能なため、それらの定義とは異なる独自の定義を使用する。

## 第 2 章

# 現在の渋滞に関する問題

ここでは本研究をするにあたって、本研究が意識する問題について述べる。

### 2.1 渋滞についての問題点

## 第 3 章

# 関連研究

### 3.1 ドライブレコーダーを用いた研究

自動車にカメラ及びドライブレコーダーを取り付けて行なった研究として、まず道路のひび割れや塗装の剥がれといったものを検出する研究 [1] がある。画像処理技術が向上したことで、道路における障害物やひび割れといったものの検出が可能となった。しかし、それらの不動で固定的なものとは異なり、渋滞というものは時間と共に発生したり解消したり、また規模も大きくなったり小さくなったりと変化するものである。そのような変化するものに対しての画像認識の対応は困難なものだった。

ドライブレコーダーと機械学習を組み合わせたものとしては、Japan Taxi 株式会社が既にタクシーに高性能なドライブレコーダーをつけて、走行中の道路の状況等をリアルタイムに収集し解析するというサービスを行なっている。しかし、そこで行われているのは未だ歩行者の検出や道路工事の検出などであり、そこから発展して渋滞を推定するという旨の実現には至っていない。各タクシーから集められた情報をもとに人間の手で渋滞かどうか判断しているのが、この研究によってその人間の判断を自動化する、あるいは事前に渋滞かどうか疑わしい現場の映像に判断材料を絞ることができる点でこの研究には貢献できる点があると言える。

### 3.2 ドライブレコーダー以外の車載カメラを用いた研究について

ドライブレコーダー以外の車載カメラの研究について、三上ら [2] は、ゴミ収集車にカメラ取り付けるとことで、ゴミ収集車にどれほどのゴミ袋が収集されたかを調べる研究を行った。ゴミ清掃車は、日々決まったルートを通り、決まった場所でゴミを収集している。その収集されるゴミの量を数値として可視化することで住民の生活の変化を可視化する研究である。自動車にカメラを取り付けて、映像情報からデーターを可視化するという目的の点で、可視化するデーターが異なるとはいえ、三上らの研究はこの研究と共通点がある。また、その研究には私の研究と同様の YOLO と呼ばれる物体検出システムが使用されており、その物体検出システムが検出したゴミ袋の量を計測していた。

このように、物体検出システムを使用することでデーターの可視化が可能であると言える。

### 3.3 渋滞推定の研究について

ドライブレコーダー以外の方法でカメラから渋滞を測定する方法として、進藤ら [3] は、対向車の交通量を測定することで渋滞を推定する研究を行なった。この方法は自動車 2 台を走行させ、その 2 台の横を通過した

対向車の数から渋滞しているか否かを判断するものである。しかし、この方法では測定するための自動車は常に道路の右側車線を走行しなければならないと同時に、対向車道がない道路では渋滞を推定することができないという問題がある。

### 3.4 カメラを用いた距離推定の研究

車間距離を推定するフェーズに関して、動画から距離を推定する方法はいくつか存在する。まず最も簡単かつ確実な方法は複眼カメラと三角関数を利用した方法である。同じ対象を移したときに2つの視点からそれぞれのカメラから対象の距離を測定する方法である。しかし、この研究の対象はドライブレコーダーであり、1つの自動車に2つのドライブレコーダーを搭載することは現実的ではなく、複眼のドライブレコーダーは一般的に搭載されていないためこの方法は利用できない。そのため、この研究では単眼カメラの映像から車間距離を推定する必要がある。単眼カメラから物体との距離を推定する研究としては、カメラ情報に加えて各種距離センサーの情報を利用することで実現する方法が一般的であった。

しかし、近年では Deep Learning 技術の向上により、各種センサーといった教師データのない、純粋に画像のみでの距離推定プログラムが実現している。Zhou ら [4] はセンサー情報等の教師データのないカメラ映像から奥行きを推定するシステムを開発した。このシステムは、車載動画等の連続的な画像の集まりであるデータセットを元に映像における視界の変化を元に奥行きを推定するものであった。これ以降、カメラ映像における奥行きの研究はこの Zhou らの研究を元に性能向上がなされ、Yin ら [5] などによってアップデートがなされた。

中でも、Google の研究チームの Casser らが行なった研究である "Struct2Depth" [6] というシステムは上記の研究と比較して精度がより向上されている。この研究ではその Struct2Depth を利用し、車間距離を推定する。

### 3.5 物体検出システムについて

物体検出を行うライブラリは様々あり、Fast R-CNNs[7]、Mask R-CNN[8]、SSD[9] といった手法が挙げられる。それぞれ CNN と呼ばれる畳み込みネットワークを利用し物体検出を行なっている点は同じだが、その CNN の層や物体検出を行う際の手法がわずかながらに異なっている。この研究では物体検出手法として YOLO(You Only Look Once)[10] と呼ばれる物体検出手法を用いる。この手法は上記の3つの手法と比較して畳み込みネットワークやニューラルネットワークがシンプルな構造になっているにもかかわらず、高い検出率を誇っている。また、構造が比較的シンプルなため演算処理にかかるスピードが速く、リアルタイムでの処理にも適している。この研究では、その YOLO が Microsoft COCO dataset を機械学習したものを利用し、ドライブレコーダーに写っている自動車、トラック、バスを主に検出させ、距離推定及びその結果からシステムの評価を行う。

## 第 4 章

# アプローチ

本研究ではドライブレコーダーの映像から自動車を検出、車間距離を推定し、渋滞しているか否かを判断するシステムを提案する。本章では、自動車の検出と車間距離を推定する手法とそのネットワーク構造について述べる。その次に、実験と評価について述べる。

### 4.1 研究手法

この研究では、汎用性を重視するためにドライブレコーダーから得られる情報のみを利用する。渋滞の定義の項でも述べたが、渋滞を推定するためには速度が重要だが、ドライブレコーダーと自動車の速度を同期させるためには専用の取り付け工事が必要であり、また車の種類や自動車メーカーによって内部構造は異なっている。また、映像から速度を推定する方法も手法の一つとして考えられるが、映像から速度を出すためには相対距離の算出が必要であり、リアルタイムの情報が求められているこの研究ではそのような相対速度を計算する処理は不向きなと同時に、映像から正確な速度情報を算出することは非常に困難である。また、速度センサーを搭載して速度を求める方法も考えられる。そのような手法に使われるのは加速度センサーが考えられるが、現状加速度センサーが搭載されている市販のドライブレコーダーは存在しないため、状況が限られてしまい、汎用性が担保できない問題が発生してしまう。また、加速度センサーが搭載されているスマートフォンを自動車に取り付けて速度推定を行う方法も考えられる。一般的に、速度は加速度を積分することでその時々速度を推定することができるが、不良積分が多く発生してしまう恐れがあると同時に、物体検出、速度推定に加えてそのような加速度センサーデータから速度を計算すると機械が処理するものが増え、映像の処理における FPS が低下する恐れがある。この FPS の低下はリアルタイムでの処理を目指すこの研究において重大な欠陥となってしまう。

さらに、GPS を用いて渋滞を推定する方法もあるが、ドライブレコーダー等の映像記録装置は実際に映像を取得できる利点があるのに対して、GPS は位置情報データーを取得するため、得られる情報に限りがある。その上、GPS のみで渋滞情報を取得する方法は例えば GPS を搭載した電子機器を複数自動車に持ち込むと、実際には渋滞していないのにもかかわらず、渋滞していると出力してしまうような事例がある [出典] ため、実際に渋滞の現場を確認することができるという点でドライブレコーダーを使う方法が最も汎用性が高い。それに加えて、速度だけで判断してしまうと、例えば自動車が停止したとき、渋滞のために停止したのか、信号のために停止したのか、あるいは駐車したために停止したのか、判断することができない、という問題が発生してしまう。近年では自動運転技術等の向上により、障害物に近づくとドライバーに警告するようなセンサーを搭載した高性能な自動車も存在するが、この研究ではより汎用性を重視するため、ドライブレコーダーの情報

のみで渋滞の推定を行う。この研究ではドライブレコーダーからの車間距離を推定することで渋滞を推定することを目的としているが、この車間距離の推定をすることでいずれは煽り運転の推定や警告などドライバーにとって有益な情報提供ができる研究に繋げることが可能だと考えられる。車間距離を推定することができれば、安全運転を評価することも可能になる。世界の情勢として、自動運転への移行に舵がとられているが、日本では法整備の問題等で完全な自動運転への移行は未だに時間がかかると予想されている。ドライブレコーダーの映像から得られる情報は、ドライバーの見ている景色とほぼ同じだとすると、ドライブレコーダーを使って渋滞だけではなく、道路標識や信号等も検出することが可能である。それらの検出をもとに、ドライバーがきちんと道路標識に従って、一時停止等していたか、十分な車間距離を守って走行できていたか、というようなドライバーが安全運転だったか否かを評価することができる。これは他のセンサーや GPS のデータだけを用いる手法では難しく、ドライブレコーダーを活用する利点であると言える。さらに、先行車の動きをトラッキング、分析することで煽り運転の検出、警告することも可能である。

## 4.2 ネットワークアーキテクチャ

本研究はドライブレコーダーの映像から自動車を検出するフェーズと同じく映像から先行車との車間距離を推定するフェーズの2つのフェーズから構成されている。

### 4.2.1 自動車を検出するフェーズ

まず、入力された動画から車を検出するフェーズでは、画像認識技術を利用して、車を検出している。物体検出を行うライブラリは様々あり、Fast R-CNNs、Mask R-CNNs、SSD といった手法が挙げられる。それぞれ CNN と呼ばれる畳み込みネットワークを利用し物体検出を行なっている点は同じだが、その CNN の層や物体検出を行う際の手法がわずかながらに異なっている。この研究では物体検出手法として YOLO(You Only Look Once) と呼ばれる物体検出手法を用いる。このシステムは物体を検出するとその物体を Bounding Box(BBOX) と呼ばれる四角形で囲むことでその物体の映像における位置を示すものである。YOLO は上記の3つの手法と比較して畳み込みネットワークやニューラルネットワークがシンプルな構造になっているにもかかわらず、高い検出率を誇っている。また、構造が比較的シンプルなため演算処理にかかるスピードが速く、リアルタイムでの処理にも適している。この研究では、その YOLO が Microsoft COCO dataset を学習したものを利用する。

### 4.2.2 距離を推定するフェーズ

本研究では Google Tensorflow が開発した Struct2Depth システムを利用する。このシステムは入力された映像から奥行きを推定し、カメラとの距離に近いほど明るい色に、遠いほど暗い色に色分けすることによって奥行きを可視化する。本研究ではそのシステムを利用し、自動車が検出された BBOX の場所と同期させることでその明るさから先行車との車間距離を推定する。

## 4.3 渋滞の評価について

ここでは渋滞の評価方法について述べる。背景の項でも述べた通り、日本において渋滞の定義は自動車の運行スピードに依存しているが、この研究においては正確なスピードを取得するのは困難なため、先行車との車

間距離を推定することで、その推定された距離を元に渋滞しているかどうかを判断する。

## 第 5 章

# 実験

ここでは、本研究の実験が行なった実験について述べる。

### 5.1 データセット

ここでは、本実験に用いたデータセットについて述べる。本実験ではドライブレコーダーの映像を用いる。そのドライブレコーダーの映像についてはインターネット上の動画投稿サイト YouTube に投稿されているドライブレコーダーの車載動画を用いる。その中でも主観的に渋滞していると思われる動画をピックアップし、データセットとする。

### 5.2 予備実験

ここでは、本研究が実験を行うにあたってその準備実験の試行として行なった実験について述べる。

#### 5.2.1 予備実験

本研究ではドライブレコーダーから得られる動画を用いて実験、評価を行うが、その予備実験として動画から切り出された画像を用いて本研究のシステムの実験を行った。この予備実験は、本研究を行うにあたって、距離推定プログラムあるいは深度推定プログラムの候補が2つあり、そのどちらが相応しいかを定める実験である。一つ目は FCRN-DepthPrediction[?] というシステムであり、もう一つは struct2depth である。その結果を図??とに示す。図??のように、このシステムは画像を上下二段に切り分け、上段では入力された映像の高さを半分にした画像、下段はその画像を用いて奥行き推定を行ったものである。また、上段で自動車とトラックを検出し、下段の同じ位置に BBOX を出力している。下段の BBOX 内に出力されているパラメータは、下段 BOX 内の RGB 値のそれぞれの値を平均した数値である。

#### 5.2.2 予備実験 - 評価

ここでは、本研究における予備実験の評価について述べる。図 5.1 と図 5.2 を比較すると、前者は深度推定を行なった結果のみを出力するのに対し、後者は元の画像を圧縮し上下二段二分けて出力されていることがわかる。また、前者は奥にある物体が明るく色分けされているが、カメラから近いところにある道路やトラックなど、ほぼ暗い青色で塗り分けられているだけであり、物体の識別が難しい。それに対して後者はカメラから



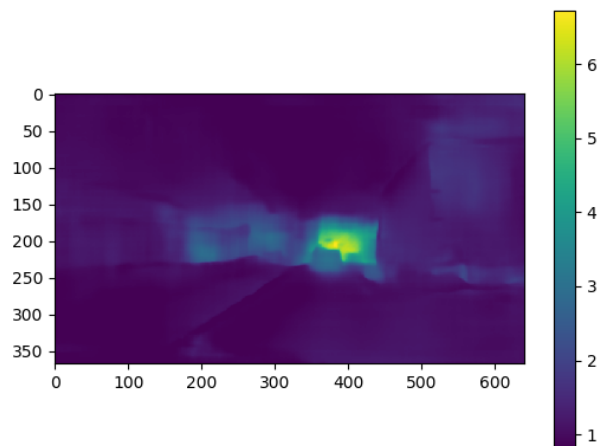


図 5.1 FCRN-DepthPrediction を用いた予備実験



図 5.2 Google Tensorflow を用いた予備実験

近い距離にある物体ほど明るく塗り分けられており、奥にある物体ほど暗い色になっている。前者と後者を比較した際、特に違いが顕著なのは後者のシステムの方が画面左側にあるトラックをはっきりと色分けできているという点である。そのため、本研究においては後者の Google Tensorflow を採用し、本実験を行うことにする。

また、図 5.2 を見ればわかる通り、物体検出システムと距離推定システムを用いることで先行車との大まかな車間距離を調べることが可能になることがわかる。図 5.2 においてシステムが検出した自動車をそれぞれ右から車 1、車 2 とおくと、下段の RGB 値のそれぞれの平均を見てみると、奥にある車 1 の方が手前にある車 2 よりも R,G の値が低く、B の値が高いことがわかる。この数値の違いを評価することで先行車との距離を推定し、走行スピードを推定し、渋滞しているかどうかを機械に判断させる。この予備実験を生かして、本実験で

は動画を用いた渋滞推定を行う。

### 5.3 本実験

予備実験にて、本研究で使用するシステムを決定した。この項では動画を用いて実際に渋滞を推定する実験について述べる。

## 第 6 章

# 評価

ここでは本研究が行った実験における評価について述べる.

### 6.1 考察

## 第 7 章

# 結論

このようにドライブレコーダーに限らず、現在映像記録装置としてのみ扱われている様々な機械は画像認識技術の向上や研究によって今までとは全く異なる使い方が考えられる。

# 謝辞

本研究を進めるにあたり、ご指導を頂きました慶應義塾大学環境情報学部教授中澤仁博士に深く感謝いたします。また、慶應義塾大学中澤研究室の諸先輩方には折りに振れ貴重なご助言を頂きました。特に慶應義塾大学大学院政策・メディア研究科陳寅特任助教、慶應義塾大学大学院政策・メディア研究科大越匡特任講師、には本論文を執筆するにあたってご指導頂きました。ここに深く感謝の意を表します最後に、前大学での1.5年間、そこから海外への半年の留学、そして1年間の浪人を経ての慶應義塾大学での4年間の生活を支えてくださった家族に深く感謝いたします。

2020年1月21日

李 広耀

## 参考文献

- [1] 全邦釘, 嶋本ゆり, 大窪和明, 三輪知寛, 大賀水田生. ディープラーニングおよび random forest によるコンクリートのひび割れ自動検出手法. 土木学会論文集 F3 (土木情報学), Vol. 73, No. 2, pp. L297–L307, 2017.
- [2] 三上量弘, 陳寅, 中澤仁ほか. Deepcounter: 深層学習を用いた細粒度なゴミ排出量データ収集手法. 情報処理学会論文誌, Vol. 61, No. 1, pp. 36–48, 2020.
- [3] 進藤瞭, 白石陽ほか. 車載カメラ画像を用いた対向車線の渋滞状況の把握手法. マルチメディア, 分散協調とモバイルシンポジウム 2013 論文集, Vol. 2013, pp. 139–146, 2013.
- [4] Tinghui Zhou, Matthew Brown, Noah Snavely, and David G Lowe. Unsupervised learning of depth and ego-motion from video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1851–1858, 2017.
- [5] Zhichao Yin and Jianping Shi. Geonet: Unsupervised learning of dense depth, optical flow and camera pose. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1983–1992, 2018.
- [6] Vincent Casser, Soeren Pirk, Reza Mahjourian, and Anelia Angelova. Depth prediction without the sensors: Leveraging structure for unsupervised learning from monocular videos. In *Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)*, 2019.
- [7] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pp. 91–99, 2015.
- [8] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.
- [9] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pp. 21–37. Springer, 2016.
- [10] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv*, 2018.
- [11] Iro Laina, Christian Rupprecht, Vasileios Belagiannis, Federico Tombari, and Nassir Navab. Deeper depth prediction with fully convolutional residual networks. In *3D Vision (3DV), 2016 Fourth International Conference on*, pp. 239–248. IEEE, 2016.