氏　　名	SOOKSATRA, Sorn		
学位の種類	博士(情報科学)		
学位記番号	博情第 451 号		
学位授与年月日	令和 3 年 3 月 24 日		
論文題目	Vehicle Density Estimation Using Density-Aware Attention Network for Traffic Surveillance System		
論文審査委員	主査	吉高淳夫	北陸先端科学技術大学院大学　准教授
		赤木正人	同　教授
		鵜木祐史	同　教授
		岡田将吾	同　准教授
		Toshiaki Kondo	Sirindhorn International Institute of Technology, Thammasat University　准教授

論文の内容の要旨

The surveillance system is widely deployed in several applications for observing the characteristic of target density because it has an advantage in low installation and maintenance costs. In a traffic surveillance system, this characteristic is usually applied in a traffic light control system (TLCS) from the number of vehicles. It helps to control the period of traffic lights and prevent any traffic accidents from vehicles. Besides, the traffic characteristic is useful information for road users to avoid crowded regions, including vehicles, people, and others. Since this system utilizes computer vision-based techniques, it is sensitive to outdoor environments (e.g., lighting conditions, traffic viewpoints, and so on). In addition, hardware with low performance usually is deployed in the surveillance system. It means that the system should be compact and requires low computational cost for operation. Therefore, density estimation cannot be operated properly under any circumstances without concerning these conditions.

Recently, there are several related studies on density estimation. It is well known that density is originally estimated by detecting and counting their regions in the input image. The counting target regions were classified by their visual features (appearance and motion features). Instead of relying on a detection-based approach, recent studies focused on end-to-end learning methods relied on regression-based approaches. These approaches concern holistic features which are visual features extracted from whole input images predicted by a prediction model. The holistic features are utilized for mapping dense regions into density maps which represent the change of object density, where their ground truth was prepared by a convolution between Gaussian distribution and target coordinates corresponding to the target sizes. Nowadays, the task of vehicle counting via a regression-based approach is riddled with a scale-aware model to handle scaling problems. Prediction models are usually designed by several stacked CNNs in parallel, which are called multi-column network architectures. In short, one stack of CNN is used to predict density maps on a specific size of the vehicle. However, these techniques have problems related to practical application as follows:

- With the variation of camera viewpoints, the target sizes can be calculated accurately resulting in the misclassification for density map formulation in the ground truth of an estimation network.
- Recent regression models with multi-column network architectures have a large number of model parameters and high computational costs using several network architectures for various target sizes.

The purpose of this study is to examine the only one stack of CNN, which is called a single-column network architecture, and reduce computational costs and model parameters while keeping similar counting accuracy to multi-column network architectures. It also should not suffer from a scaling problem by avoiding the utilization of target sizes. It found that the target size can be categorized by their local densities which are related to distances of their neighbor target for preparing the ground truth. The proposed prediction model chooses to investigate the connections between feature maps from different layers, where holistic features of small and large objects can be extracted from feature maps in shallow and deep layers, respectively. The connection can be done by skip connections to integrate feature maps from shallow layers with another in deeper layers. This process, which is called forward connections, can recover the holistic features of small objects. On the other hand, the backward connections are introduced by extracting feature maps from deep layers to combine with the shallower layer. It can be expected that information on a deep layer can help to optimize the shallow layer, where the performance in an earlier stage can be improved. Considering the quality of a density map, feature maps in every layer should have the same resolution to prevent information loss from adjusting their resolutions. Then, pooling layers are replaced with a dilated convolutional layer. Therefore, the contributions of this dissertation can be summarized as follows:

- Instead of relying on the target size, the proposed density map utilizes average distances among target samples where it is designed for visualizing the difference pattern of various vehicle densities (high and low density regions).
- To reduce computation cost, a single column network architecture for vehicle density estimation is designed by including skip connections and dilated convolutions to extract holistic features from intermediate convolutional layers and keep semantic information for density map estimation, respectively.

Since vehicle density estimation is mainly focused on this research, all models are evaluated by the common criteria for vehicle density accuracy. The state-of-the-art of regression-based approach was implemented for comparison. In addition, well-known CNNs with skip connections (e.g., U-net, Resnet, and Densenet) were also applied for analysis. The empirical results show that the proposed prediction model with backward connections achieved a vehicle density accuracy 92.47 % which is close to accuracy of state-of-the-art (93.33 %). From this result, the achievement is summarized as follows:

- In the density map configuration, the target size estimated by an average distance of the

target is insensitive to camera viewpoints.

- From the point of view of counting accuracy, the proposed method with a single-column network achieves a promising result which is close to the vehicle counting accuracy from a multi-column network or network with a large number of model parameters.

Moreover, the proposed network satisfies with the minimum requirement of TLCS and it is effective to reduce under-counting errors. However, there is a room for improvement in other datasets consisting of overlapping target regions or crowd counting datasets.

Keyword: surveillance system, vehicle density estimation, regression model, skip connection, dilated convolution, traffic analysis.

論文審査の結果の要旨

監視カメラ映像の解析は現代社会をより安全、安心にする要素技術として重要である。交通監視カメラ映像の解析は、道路の渋滞や交通量の把握に必要であり、これらの情報を基に交通信号等の制御や車両への交通情報提供なども行われており、解析性能の高性能化、高精度化が望まれている。また、深層学習を適用した画像認識技術は近年目覚ましい進化を遂げており、物体の検出、識別等の精度を従来手法と比較して飛躍的に向上させた。SOOKSATRA, Sorn 君の研究は交通監視カメラ映像を入力とした車両密度の推定問題に対して深層学習手法の視点から取り組んだものである。

これまでにも画像中に映る同一種で多数存在する人や物等の数を計測､あるいは推定する研究は行われており、主として群衆を対象としてその数を計測、推定する研究がおこなわれてきた。その手法については、個々の対象物を検出してその数を直接的に計測する手法から、深層学習を適用し、群を形成している画像の事例を多数用意してそれを学習させ、対象物の密度分布を推定する手法へとその主流が変化してきた。

深層学習により、群をなす対象物の密度分布を推定する既存手法を、交通監視カメラ映像を入力とした車両密度分布の推定問題に適用する場合、群衆（人）を対象とした問題と比較して、車両の大きさに多様性があり、かつ物体間の重なりによる遮蔽の影響が大きいこと、照明環境の変化が大きく、これらの要因に依らず精度を安定化させることが容易でないことが挙げられる。また､深層学習の適用により物体の大きさに関する多様性の問題を解決するために対象物の大きさのクラス毎に検出器を用意し並列化するとパラメタが増大するという課題がある。これらの課題に対して、本論文では後ろ向きのスキップコネクションを導入し、学習パラメタの伝播に関しては同構成のマスタ・スレーブネットワークを用意して転送するという手法を導入し解決した。

以上、本論文は、対象物の大きさや、遮蔽、照明条件などに関して多様な条件下における映像中の車両密度分布を既存手法と比してより高精度に推定する手法を確立したものであり､群衆な

どの他物体の密度分布推定問題にも適用可能であることも示したものである。これらの成果は学術的に貢献するところが大きいと判断する。よって審査の結果、博士（情報科学）の学位論文として十分な価値があるものと認めた。