Parking Space Detection Based on A Multi-task Deep Convolutional Network with Spatial Transform

Hoang Tran Vu and Ching-Chun Huang

Department of Electrical Engineering, Chung Cheng University, Taiwan

Speaker: Hoang Tran Vu

ACM Applied Computing and Multimedia Lab
應用運算與多媒體實驗室

Outline

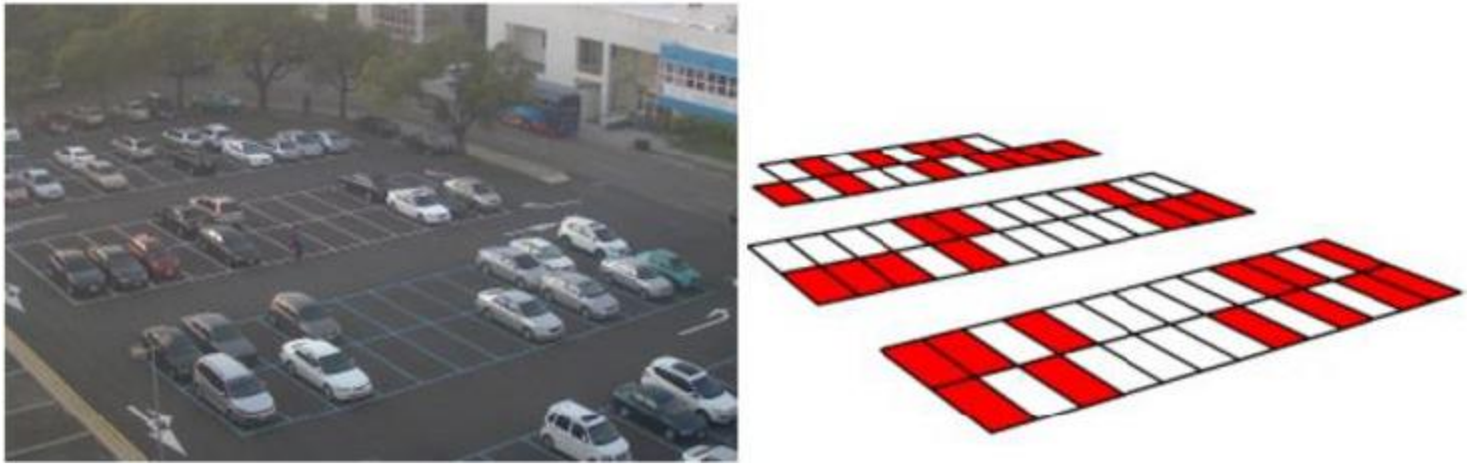
- Goal
- Challenges
- Proposed method
- Results and Discussions
- Conclusions
- Future works

Outline

- Goal
- Challenges
- Proposed method
- Results and Discussions
- Conclusions
- Future works

Goal

- Using a CNN-based deep learning framework to infer the parking status

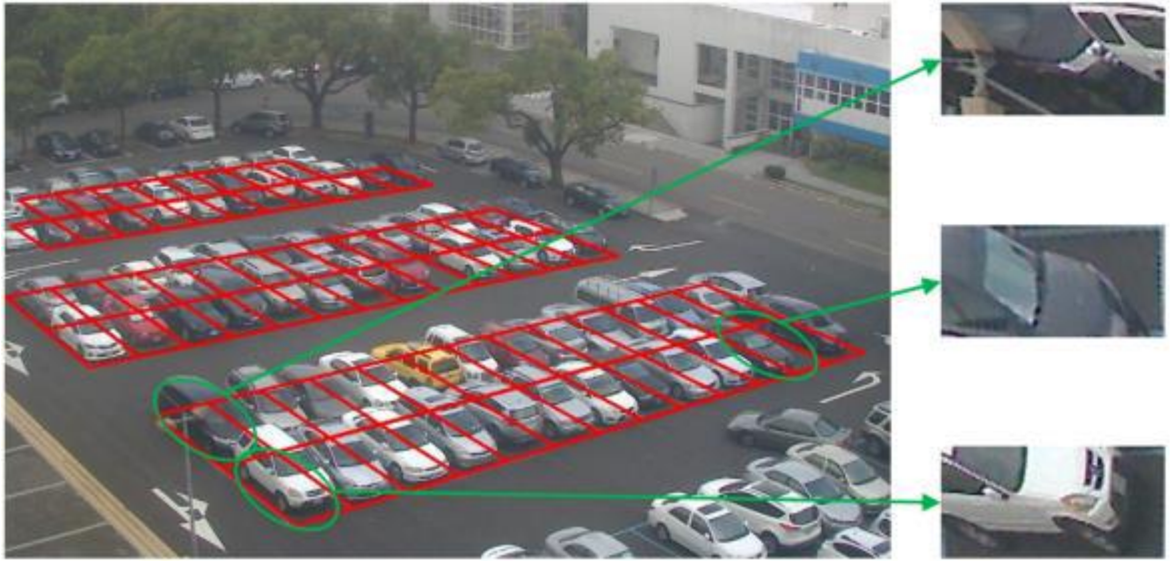


Outline

- Goal
- Challenges
- Proposed method
- Results and Discussions
- Conclusions
- Future works

Challenges

- Inter-object occlusion and perspective distortions



Challenges

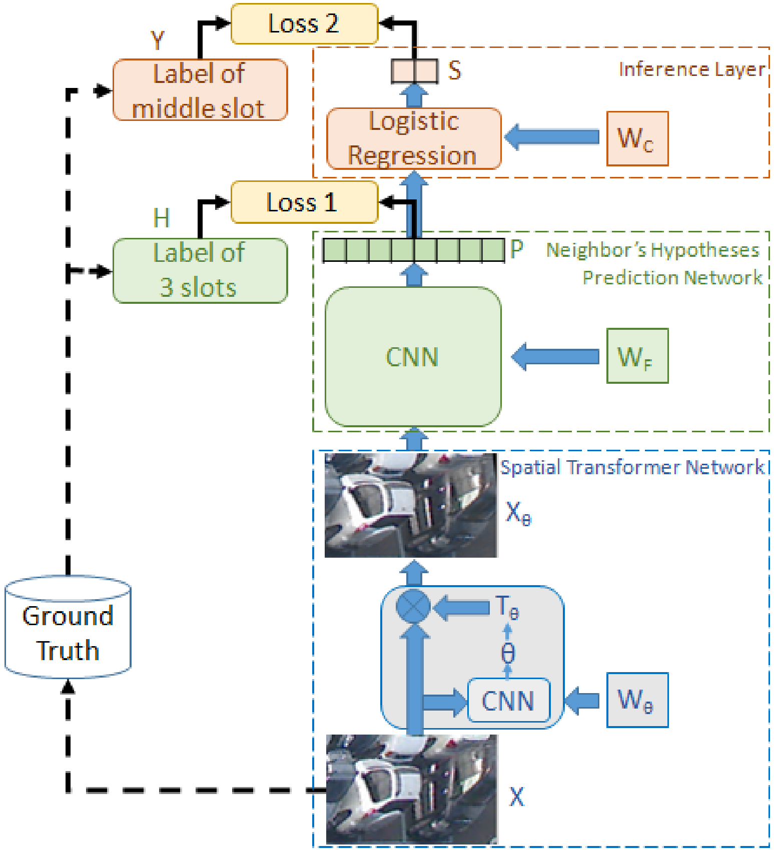
- Non-unified vehicle size and uncontrollable parking displacement.

Outline

- Goal
- Challenges
- Proposed method
- Results and Discussions
- Conclusions
- Future works

Proposed method

- Three main parts:
  - Spatial Transformer network (STN)
  - Neighbor’s Hypotheses Prediction Network (NHPN)
  - Inference layer



Outline

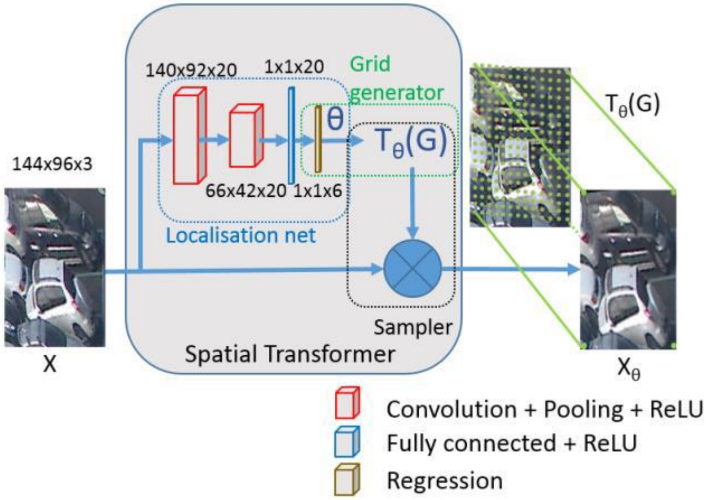
- Goal
- Challenges
- Proposed method
  - Spatial Transformer network
  - Neighbor’s Hypotheses Prediction Network
  - Inference layer
- Results and Discussions
- Conclusions
- Future works

Spatial Transformer network

- Reducing the variations from perspective distortion, parking displacement, and vehicle size.
  - Using a spatial transformer network (STN) [15]

$$\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix} = T_\theta(G_i) = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix}$$

(x^s, y^s): the source coordinate in the input image
T_θ : 2D affine transformation (6 parameters)
(x^t, y^t): the target coordinate in the transformed image



[15] Max Jaderberg, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu, "Spatial transformer networks," *in Advances in Neural Information Processing Systems*, pages 2017–2025, 2015.

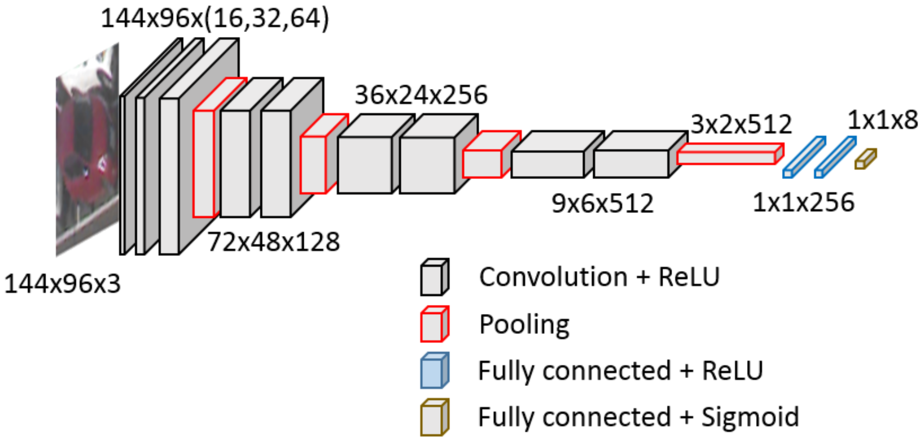
Outline

- Goal
- Challenges
- Proposed method
  - Spatial Transformer network
  - Neighbor’s Hypotheses Prediction Network
  - Inference layer
- Results and Discussions
- Conclusions
- Future works

Neighbor’s Hypotheses Prediction Network

- Solving the inter-occlusion problem.
  - Designing a CNN-based deep learning network to predict the status of a 3-space.

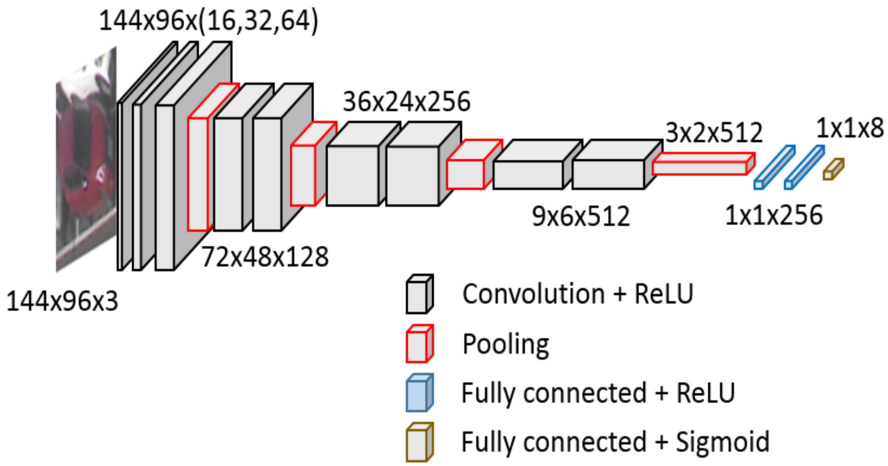
$$L_1(W_F, W_\theta, X_t, H) = -\frac{1}{8N}\sum_{n=1}^{N}\sum_{k=1}^{8}[h_n^k log p_n^k + (1-h_n^k)\log(1-p_n^k)]$$



W_F : NHPN parameters.
W_θ : STN parameters.
X_t : input training set.
H : corresponding status labels.
N : sample number.
h_n^k : the label of the k^{th} hypothesis of the n^{th} sample

Neighbor's Hypotheses Prediction Network

- This network is designed with three properties.
  - Being determined by many stages separated by a pooling layer.
  - down-sampling the input image to a small size before applying fully connected layers for classification.
  - Increasing the number of kernels in the later layers



Outline

- Goal
- Challenges
- Proposed method
  - Spatial Transformer network
  - Neighbor’s Hypotheses Prediction Network
  - Inference layer
- Results and Discussions
- Conclusions
- Future works

Inference layer

- Inferring the status of the considered space.
  - Building a 2- class logistic regression model on the top of NHPN.

$$L_2(W_C, P, Y) = -\frac{1}{N}\sum_{n=1}^{N} \mathrm{y}_n \log(S_n^1) + (1-\mathrm{y}_n)\log(S_n^0)$$

W_C : inference layer parameters.
P : input training set.
Y : corresponding status labels.
N : sample number.
y_n : the label of the middle space of the n^{th} 3-space unit.
S_n^1 and S_n^0 : the occupied and vacant probabilities of the middle space.

Outline

- Goal
- Challenges
- Proposed method
- Results and Discussions
- Conclusions
- Future works

Results and Discussions

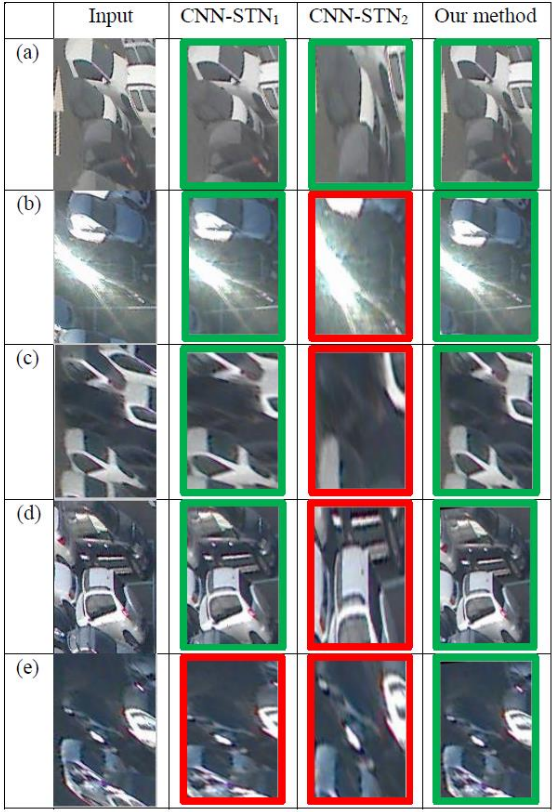
- The training and testing datasets are collected within 1 month.
  - 8277 images for training.
  - 525 images for testing.

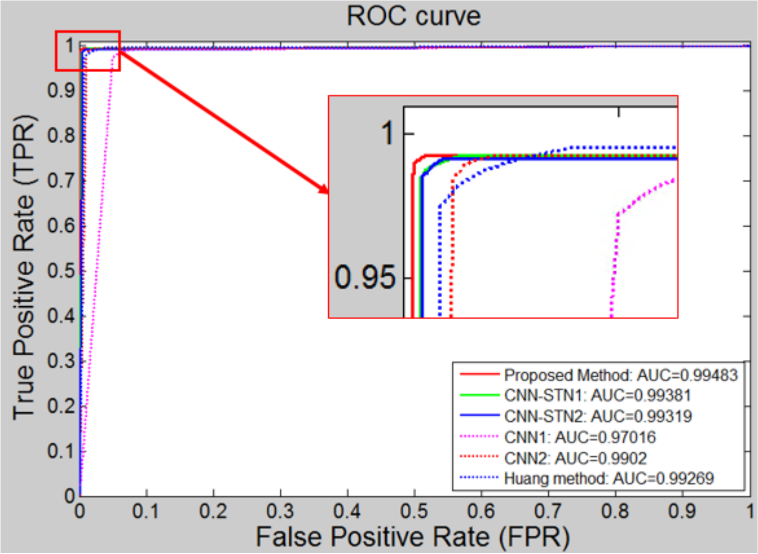
Huang's work [3] : one of state-of-art for hand-craft feature based methods
CNN_1 : CNN(considered space) + $L_2(.)$
CNN_2 : CNN(three spaces) + $L_2(.)$
$CNN\text{-}STN_1$: CNN(three spaces) + STN + $L_1(.)$
$CNN\text{-}STN_2$: CNN(three spaces) + STN + $L_2(.)$

	ACC	FPR	FNR
Huang's work [3]	98.44%	0.0128	0.0173
CNN_1	96.78%	0.0666	0.0136
CNN_2	98.71%	0.0129	0.0129
$CNN\text{-}STN_1$	99.01%	0.0057	0.0124
$CNN\text{-}STN_2$	98.98%	0.0057	0.0129
Proposed method	**99.25%**	**0.0029**	**0.0103**

[3] C.C. Huang, Hoang Tran Vu, "Vacant Parking Space Detection based on a Multi-layer Inference Framework," *IEEE Transactions on Circuit and Systems for Video Technology*, May 2016.

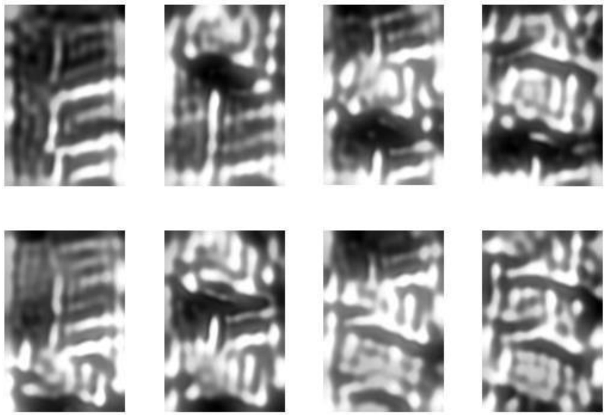
Results and Discussions





Results and Discussions

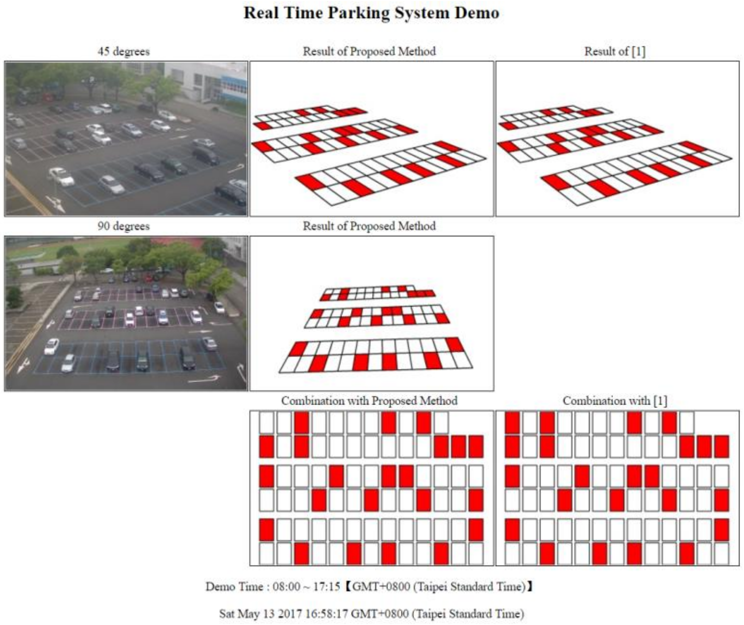
- Understanding what the network learned in the feature domain.
  - Generating the synthetic images [18] that cause high activation.



[18] J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, H. Lipson. "Understanding neural networks through deep visualization," *in ICML Deep Learning Workshop*, 2015.

Results and Discussions

- The real-time camera view and detection results.



[17] C. C. Huang. (2015). *Huang's Projects*. [Online]. Available at http://acm.ee.ccu.edu.tw:2017.

Outline

- Goal
- Challenges
- Proposed method
- Results and Discussions
- Conclusions
- Future works

Conclusions

- Proposing a deep convolutional network for parking space detection.
  - Addressing the practical challenges : lighting variations, parking displacement, non-unified car size.
  - Integrating a convolutional spatial transformer network (STN) to crop the local image area adaptively.
  - Adopting a multi-task loss function to handle the inter-object occlusion problems.

Outline

- Goal
- Challenges
- Proposed method
- Results and Discussions
- Conclusions
- Future works

Future works

- Enhancing the low contrast areas as well as learning the invariant color features.
- Using the transfer learning methods to transfer the information between different domains (different parking spaces or different viewing angles) efficiently.

Thanks for listening!