Parking Space Status Inference upon a Deep CNN and Multi-task Contrastive Network with Spatial Transform

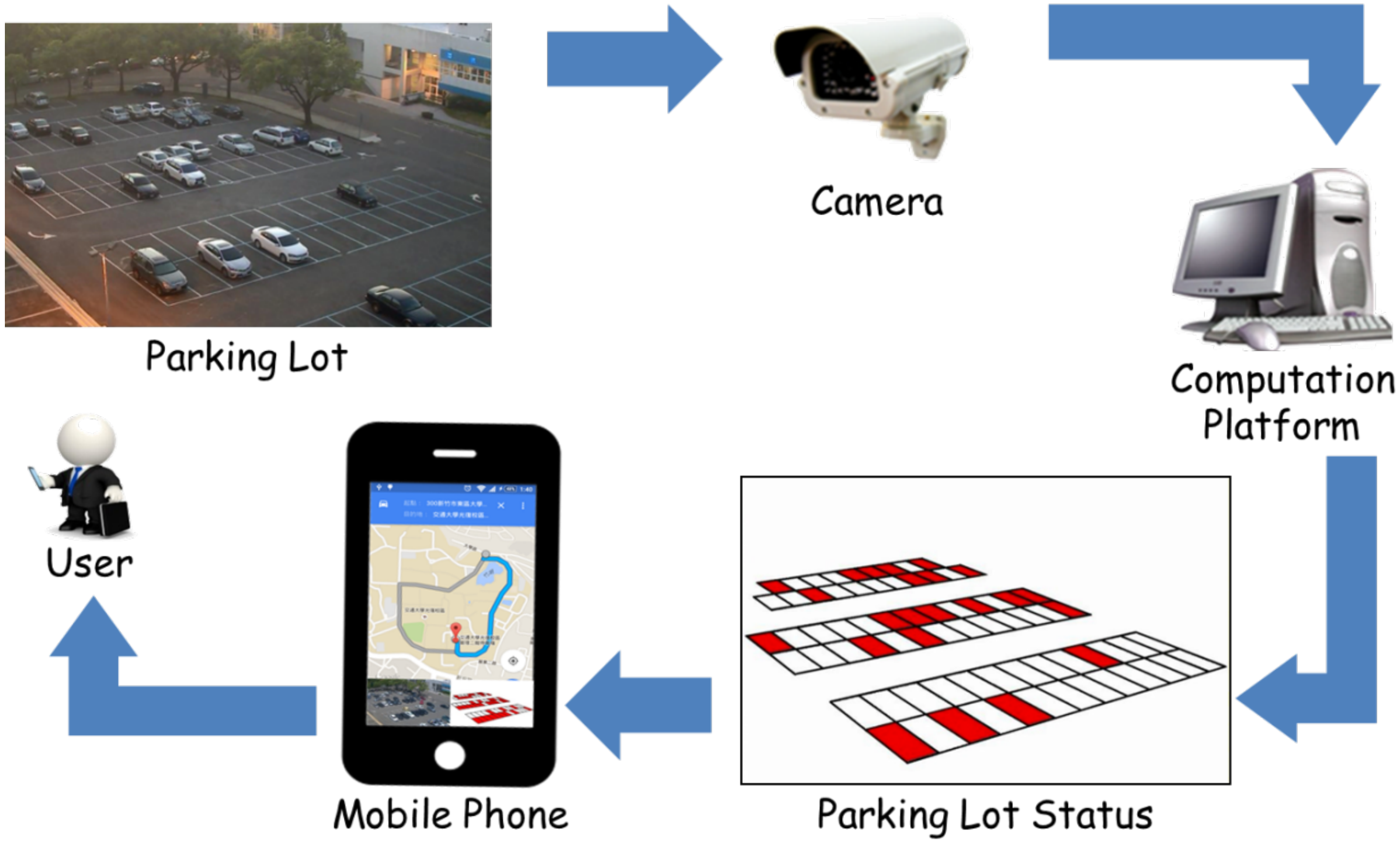
黃敬群
Ching-Chun Huang

- Hoang Tran Vu and Ching-Chun Huang

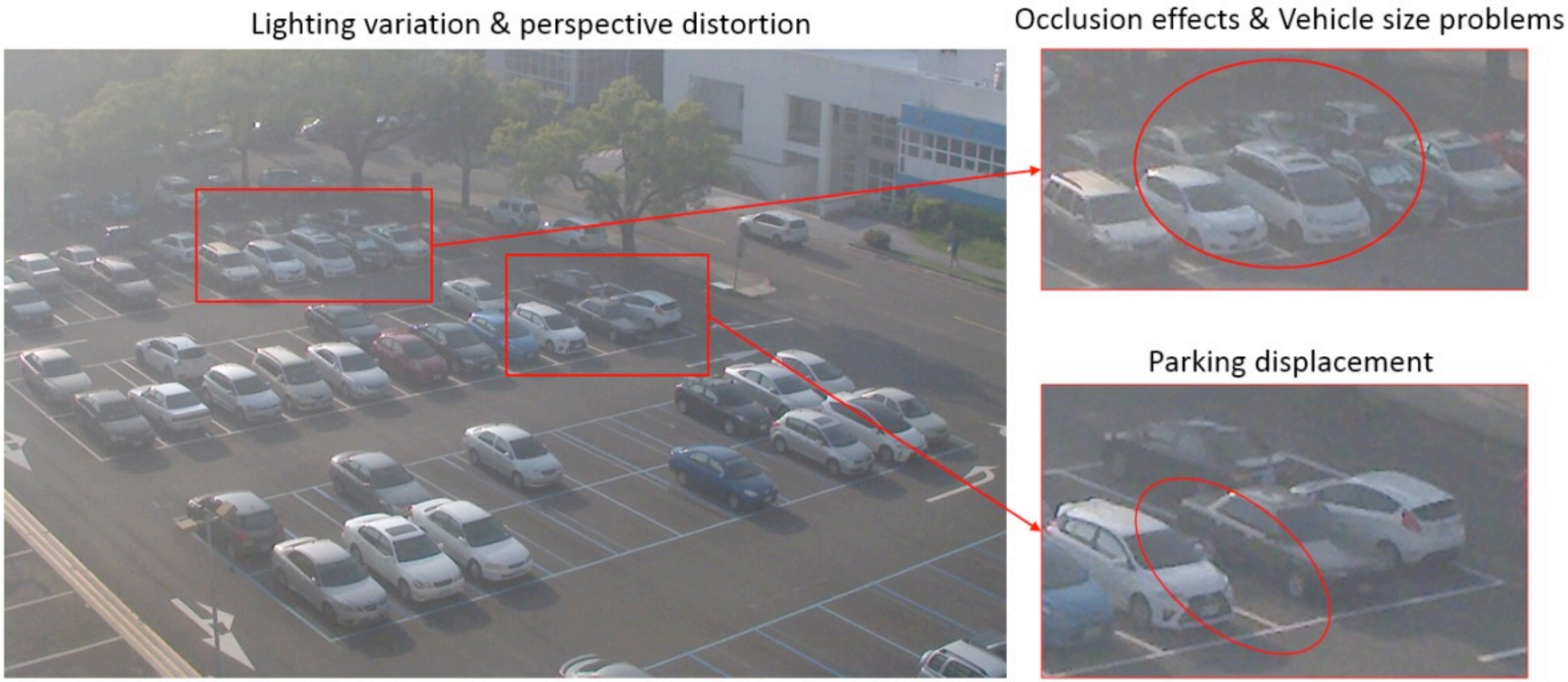
IEEE Transactions on Circuits and Systems for Video Technology

Goal

- Introducing a CNN-based deep learning framework for parking space detection system



Challenges


Lighting variation & perspective distortion
Occlusion effects & Vehicle size problems
Parking displacement

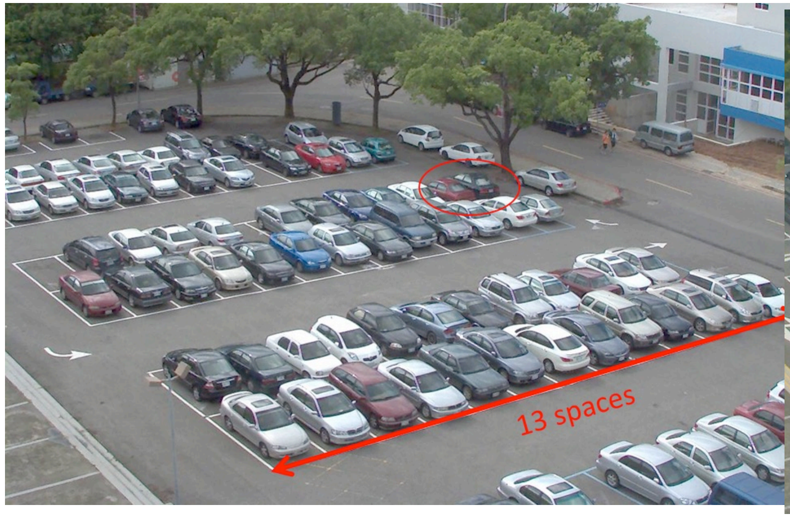
Challenges

- Difference of space size

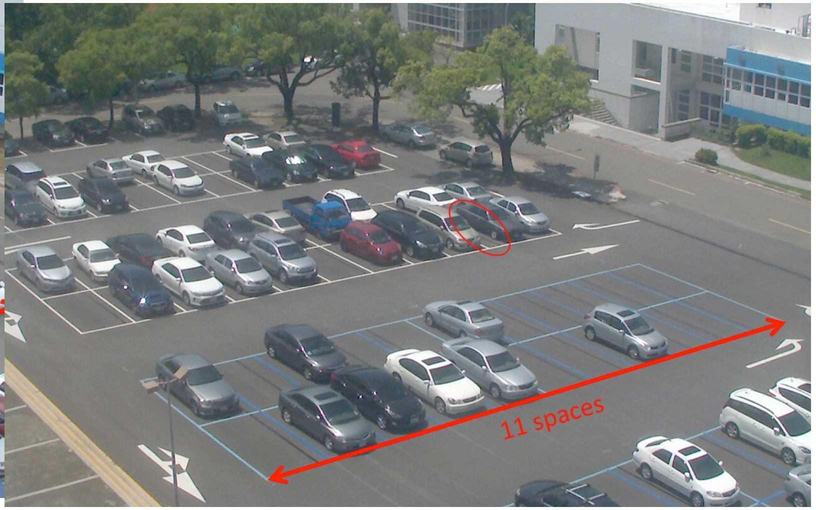
Occlusion problem becomes severer

Parking displacement becomes uncontrollable

Small size



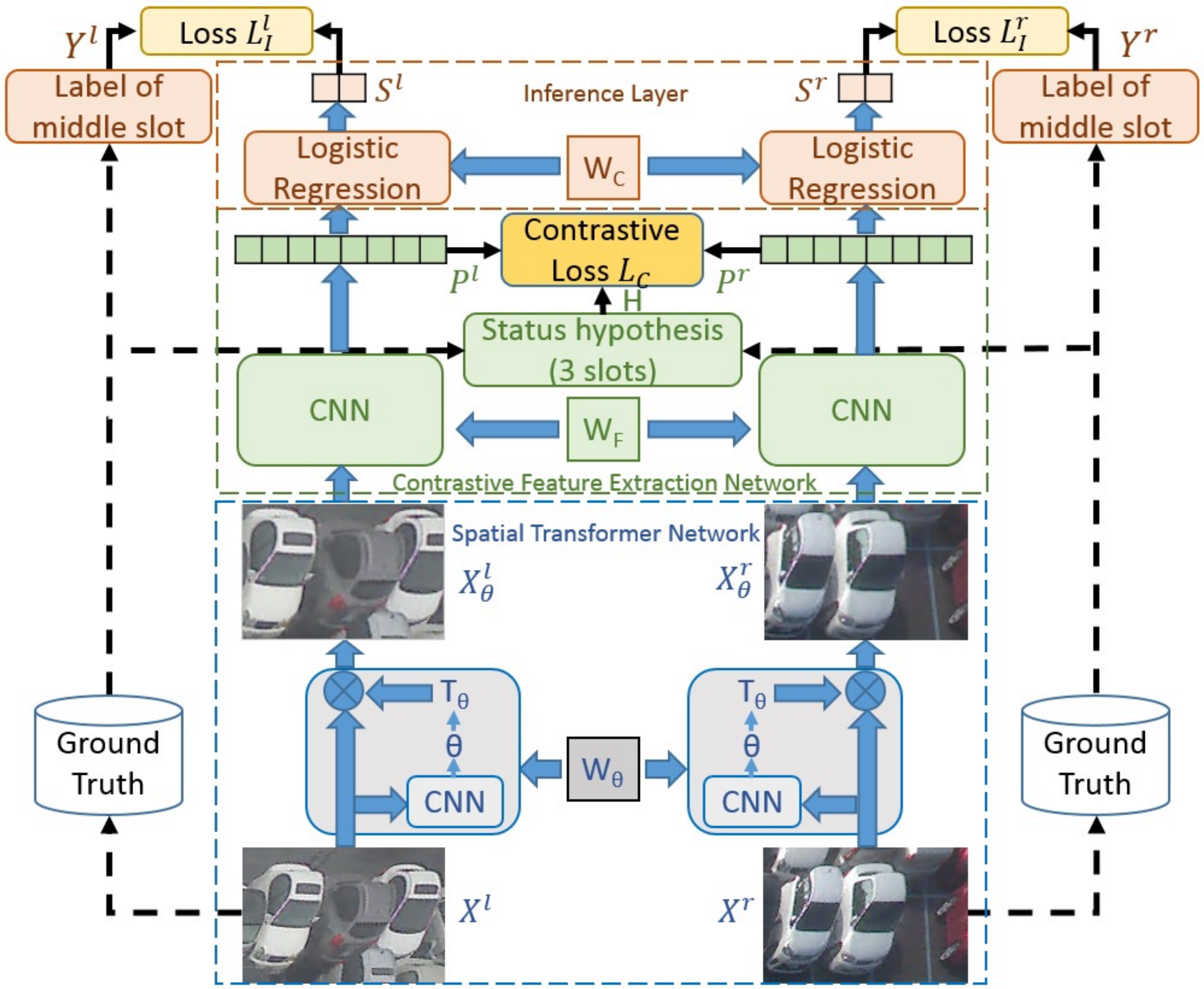
Medium size



Large size

Proposed method

- Three main parts:
  - Spatial Transformer network (STN)
  - Contrastive Feature Extraction Network (CFEN)
  - Status Inference Layer



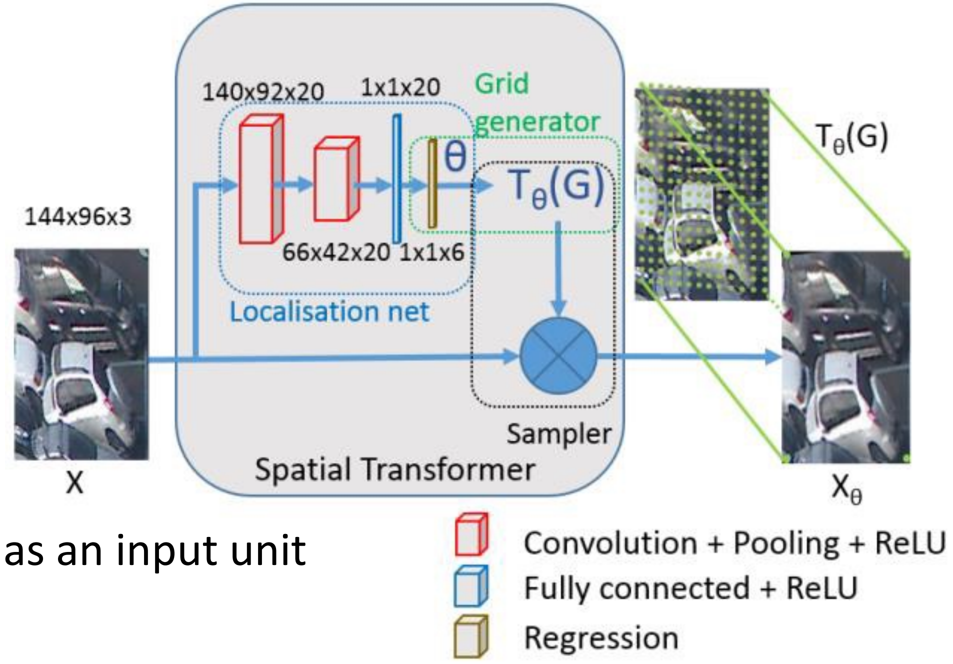
3-space as an input unit 3-space as an input unit

Spatial Transformer network

- Reducing the variations from perspective distortion, parking displacement, and vehicle size.
  - Using a spatial transformer network (STN) [25]

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

(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



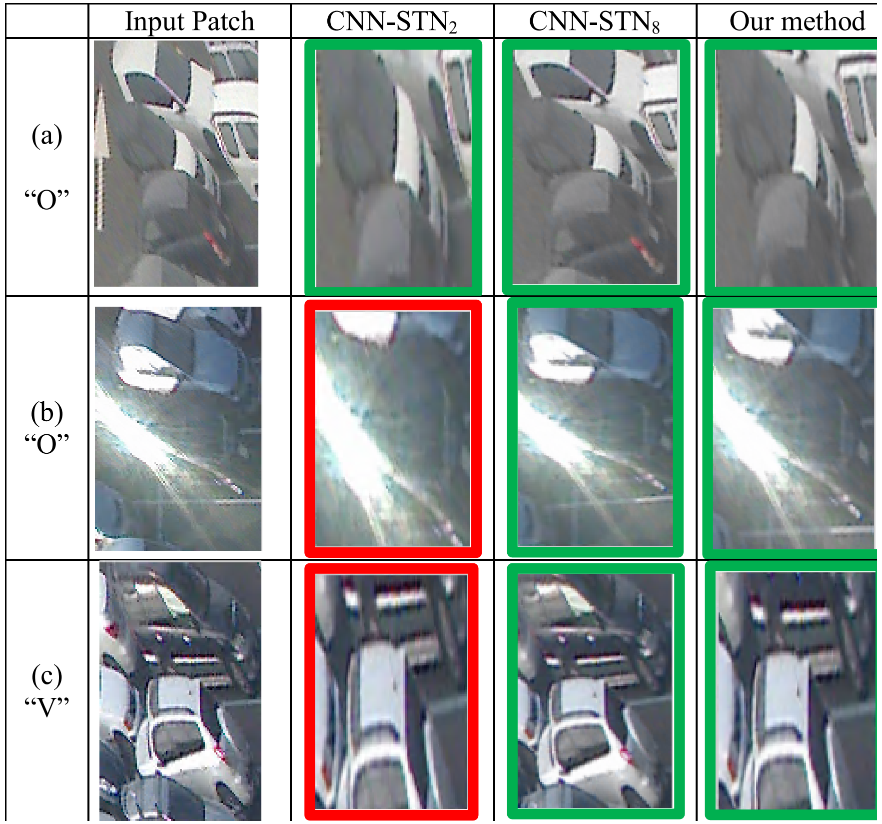
3-space as an input unit

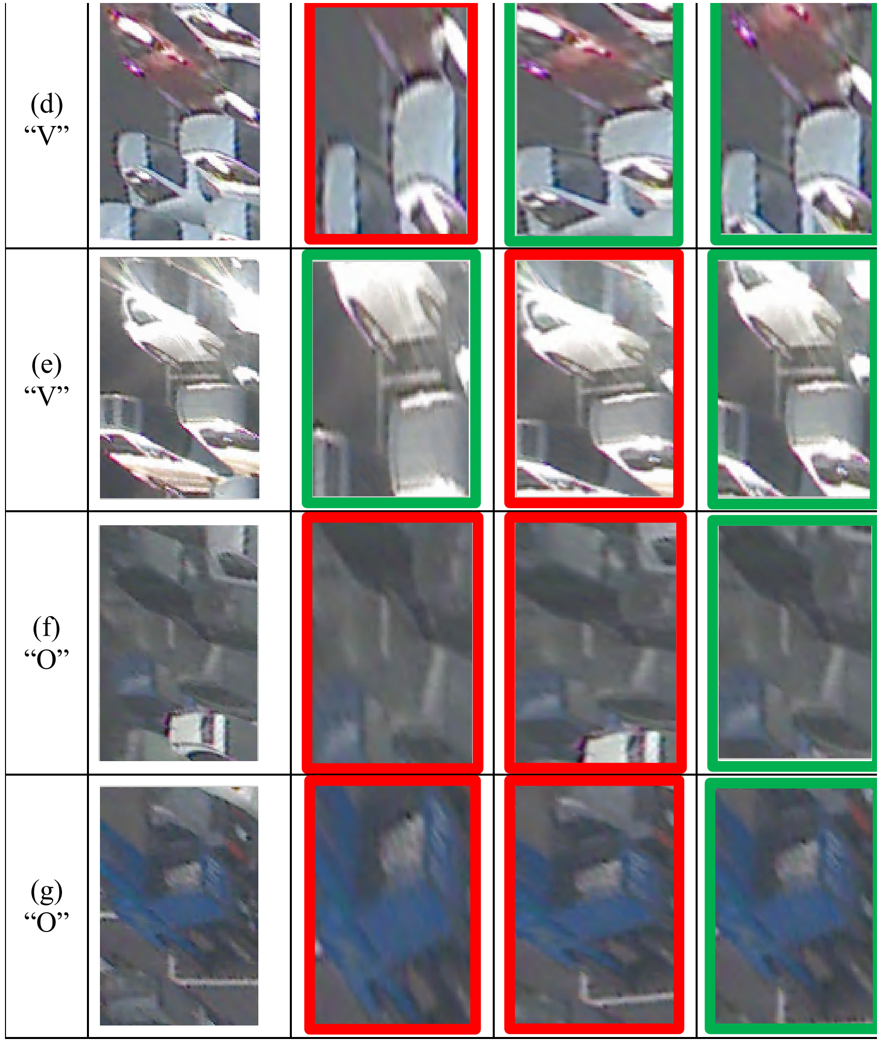
[25] Max Jaderberg, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu, "Spatial transformer networks," *in Advances in Neural Information Processing Systems*, pages 2017–2025, 2015.

Results and Discussions

Transformed Patches based on Different Network Structures:

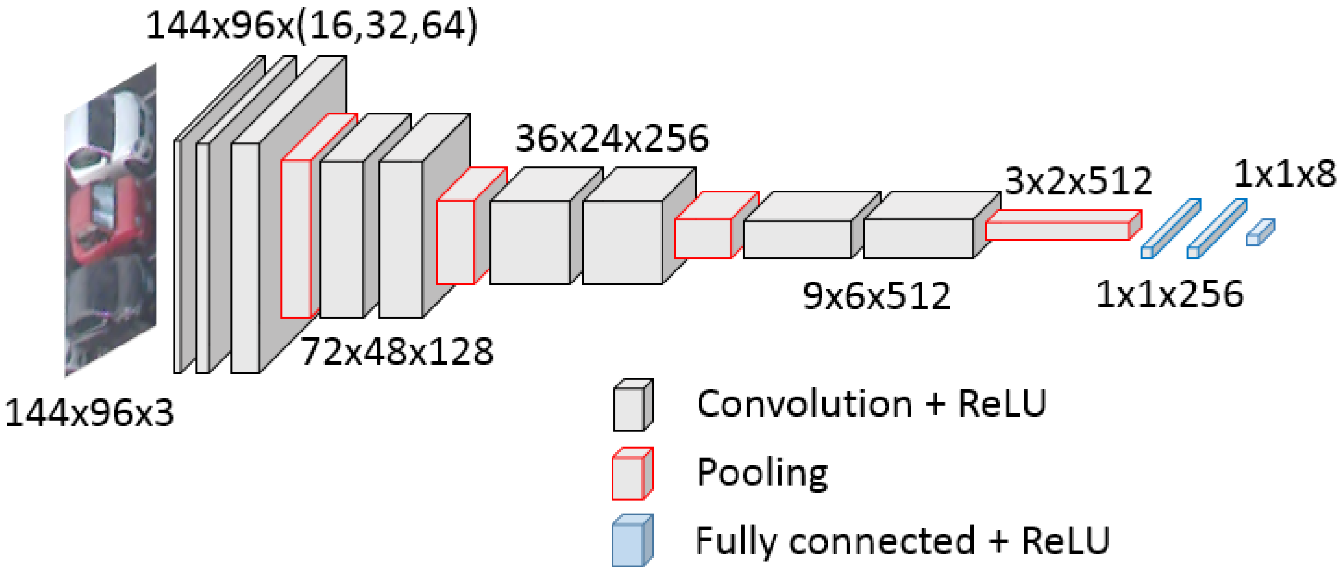
- Red boxes: false detection
- Green boxes: correct detection
- The true status of the target space: “V” is vacant and “O” is occupied.

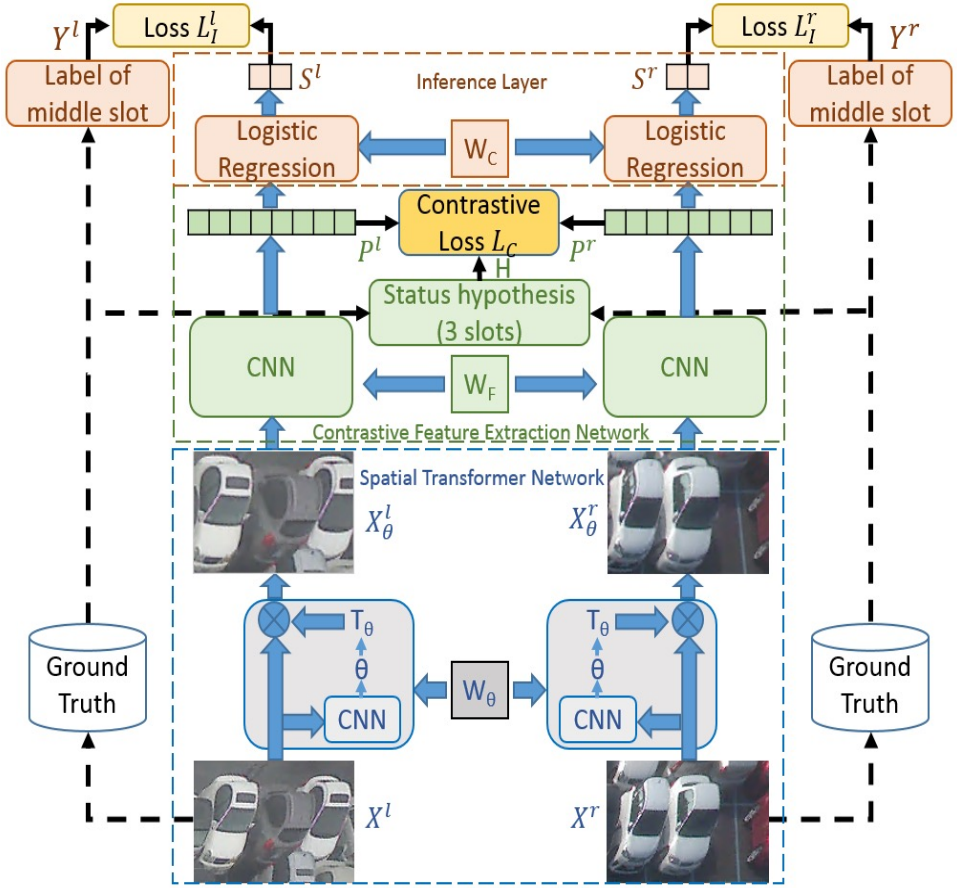




Contrastive Feature Extraction Network

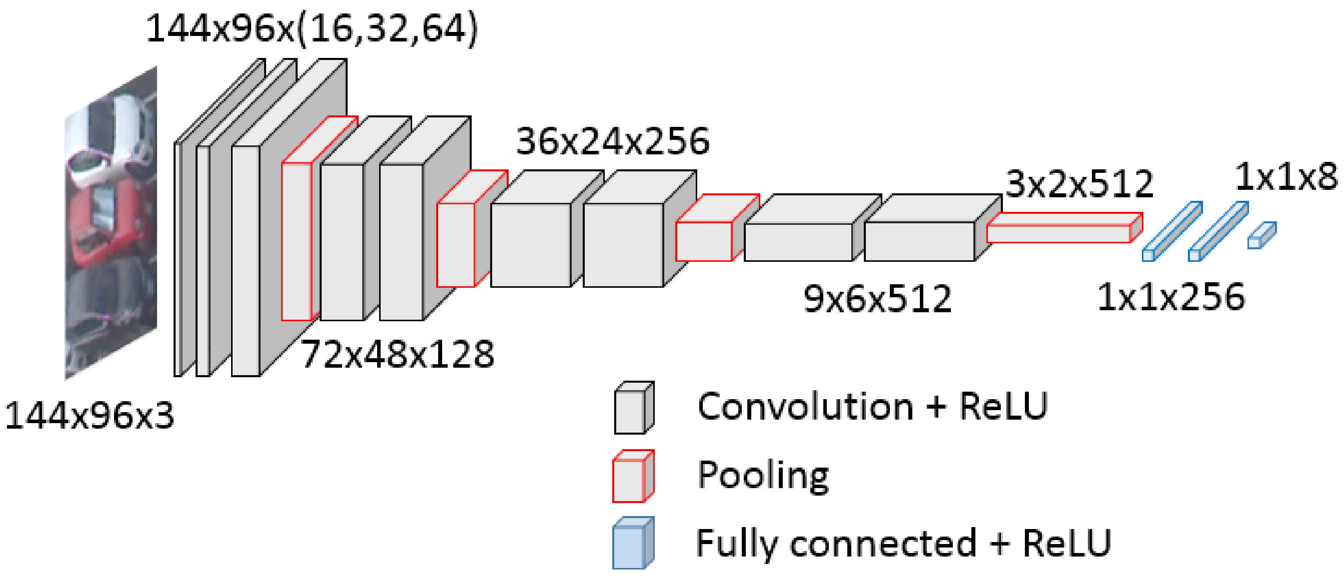
- Solving the inter-occlusion/lighting problem.
  - Designing a CNN-based deep learning network to learn the occlusion pattern within the transformed patch.





Contrastive Feature Extraction Network

- This network is designed with the following 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
  - Applying FC layers to reduce the feature dimension

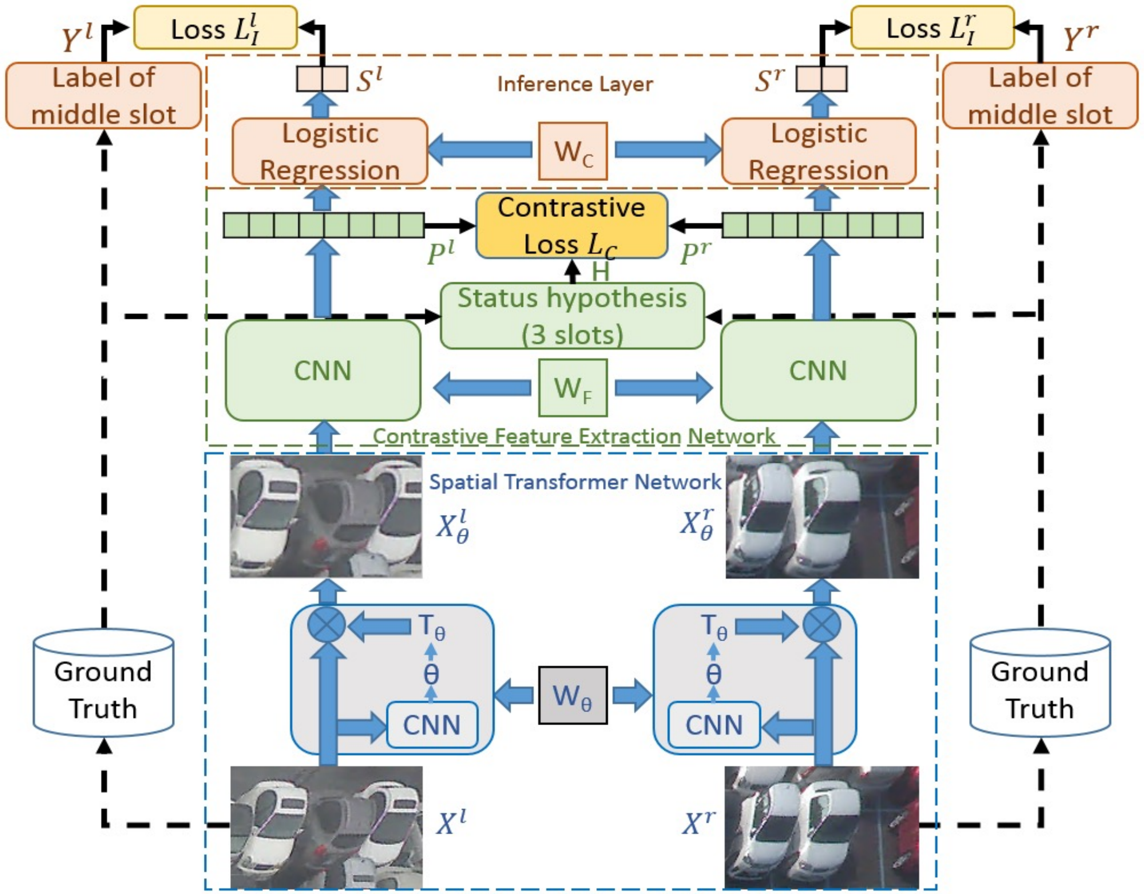


Inference layer

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

$$L_S(W_F, W_\theta, W_C|\{X_t, H\}) = -\frac{1}{N}\sum_{n=1}^{N}[\mathrm{y}_n \log(S_n^1) + (1-\mathrm{y}_n)\log(S_n^0)].$$

W_C : inference layer parameters.
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.



Results and Discussions

- Evaluation under Different Parking Space Sizes

Table VI. The Performance of Status Inference under Different Space Sizes

	ACC (%)			FPR (%)			FNR (%)		
Space Size	S	M	L	S	M	L	S	M	L
Huang's work [17]	98.44	99.61	99.69	1.28	0.5	0.40	1.73	0.31	0.25
CNN_2	96.78	98.26	99.00	6.66	1.7	1.12	1.36	1.78	0.91
CNN_8	98.71	99.53	99.82	1.29	0.64	0.17	1.29	0.28	0.20
Our method	**99.68**	**99.74**	**99.87**	**0.24**	**0.39**	**0.14**	**0.37**	**0.12**	**0.13**

Huang's work [17]: IEEE TCSVT 2016
CNN_2 : CNN(considered space) + $L_2(.)$
CNN_8 : CNN(three spaces) + $L_2(.)$

Results and Discussions

- The real-time camera view and detection results.