The Thirty-First Annual Conference on Innovative Applications of Artificial Intelligence (IAAI-19)

Jan. 31, 2019

Early Detection of Vacant Parking Spaces Using Dashcam Videos

Ming-Che Wu and **Mei-Chen Yeh**

Dept. of Computer Science and Information Engineering
National Taiwan Normal University

Where to Park?

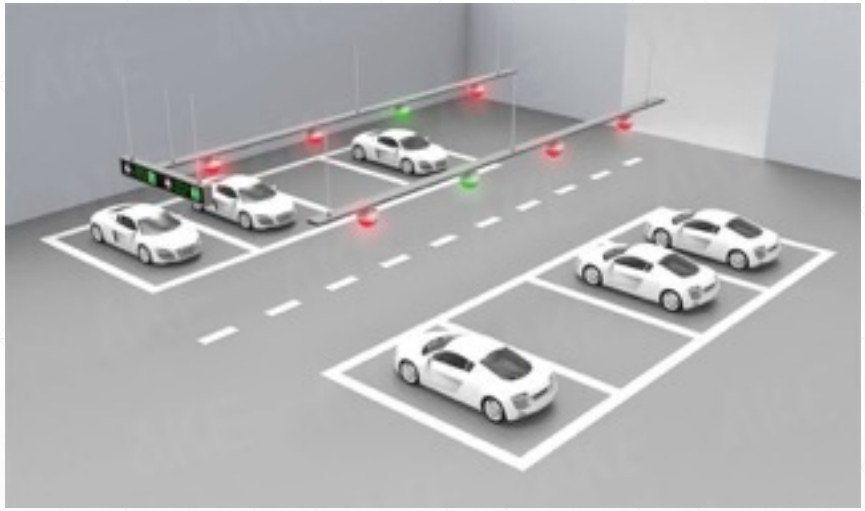
- No. of registered vehicles: 811,465
- No. of parking spaces: 752,773

臺北市政府交通局
Department of Transportation,
Taipei City Government

Taipei City
(Statistics on 2018/05)

Existing parking guidance systems

Sensors



- Costly
- Limited in indoor parking lots

Surveillance cameras

- Assuming the availability of a camera monitoring system

Detecting parking spaces using **dashcam videos**

Early detection is important

- Predict the availability of a parking space when it is partially observed or even completely occluded

1st second | 2nd second | 3rd second | 4th second | 5th second

Alarm the availability of the parking lot!

Early detection of parking space using dashcam videos

Outline

- Introduction
- **Approach**
  - **Dataset**
  - **Problem formulation**
  - **Network architectures**
- Experiments
- Conclusion

Neural network approaches

- Supervised learning setting
- Two architectures
  1. 3D convolutional neural network (3D CNN)
  2. Recurrent neural network with long short-term memory cells (LSTM)

Dataset

- 5,800 annotated 5-seconds driver's view videos
  - Collected using a dashboard camera
  - Collected at 22 different places (10 indoor and 12 outdoor) during 5 months
  - Containing parallel, perpendicular, angled and on-street parking spaces
- **Empty**: the fifth second of the clip captures a vacant parking spot
- **Full**: the clip contains no parking spot

Empty (outdoor)

(1)

(2)

(3)

(4)

(5)

Full (outdoor)

(1)

(2)

(3)

(4)

(5)

Empty (indoor)

(1)

(2)

(3)

(4)

(5)

Full (indoor)

(1)

(2)

(3)

(4)

(5)

Problem formulation

- Given a one-second video clip, predict whether a parking space is available "ahead."
- Input: $D = \{(X_1, y_1), (X_2, y_2), \ldots, (X_n, y_n)\}$
  - $X_i = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m]$
  - y_i = 0 (full) or 1 (empty)
- Output: $f(\boldsymbol{x}_i, \boldsymbol{\theta})$

Goal: Detect a vacant parking space as early as possible given the observations $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_t)$ I $t < m$ before reaching the spot at frame m

Training the detector θ

- The detection should be more reliable when the driver is approaching an empty parking space.
- Minimize the detection loss (2 terms):

1

$$L_{conf}(X_i) = \frac{1}{m-1}\sum_{t=2}^{m}\left(\max\left(0, l\times\left(f(x_{t-1})-f(x_t)\right)\right)\right)^2$$

$$l = \begin{cases} 1 & \text{if } y_i = 1 \\ -1 & \text{if } y_i = 0. \end{cases}$$

Training the detector θ

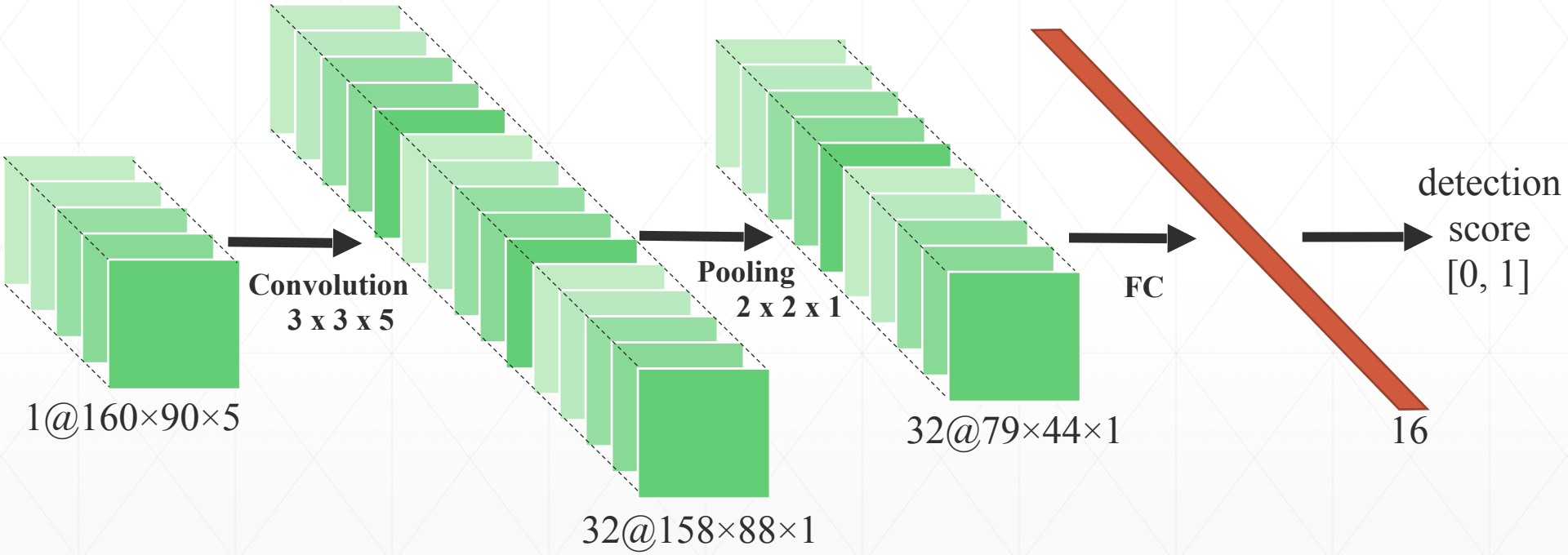
- The detection should be more reliable when the driver is approaching an empty parking space.
- Minimize the detection loss (2 terms):

2

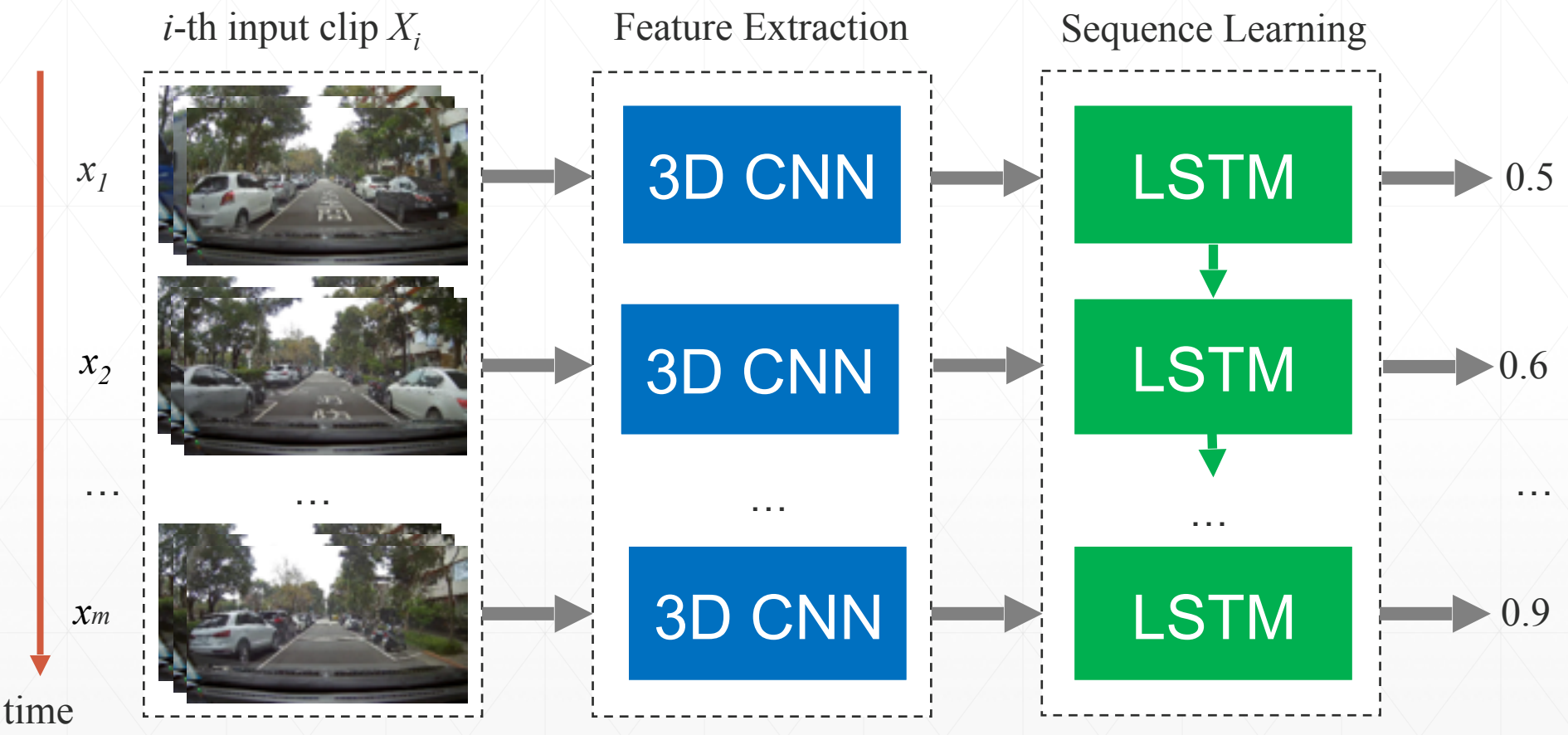
$$L_{acc}(X_i) = \frac{1}{m}\sum_{t=1}^{m}(y_i - f(x_t))^2$$

$$\sum_{X_i \in D}\left(L_{conf}(X_i) + \alpha \cdot L_{acc}(X_i)\right)$$

Network architecture: 3D CNN



Network architecture: 3D CNN+LSTM



Outline

- Introduction
- Approach
  - Dataset
  - Problem formulation
  - Network architectures
- **Experiments**
- Conclusion

Experiment: Setup

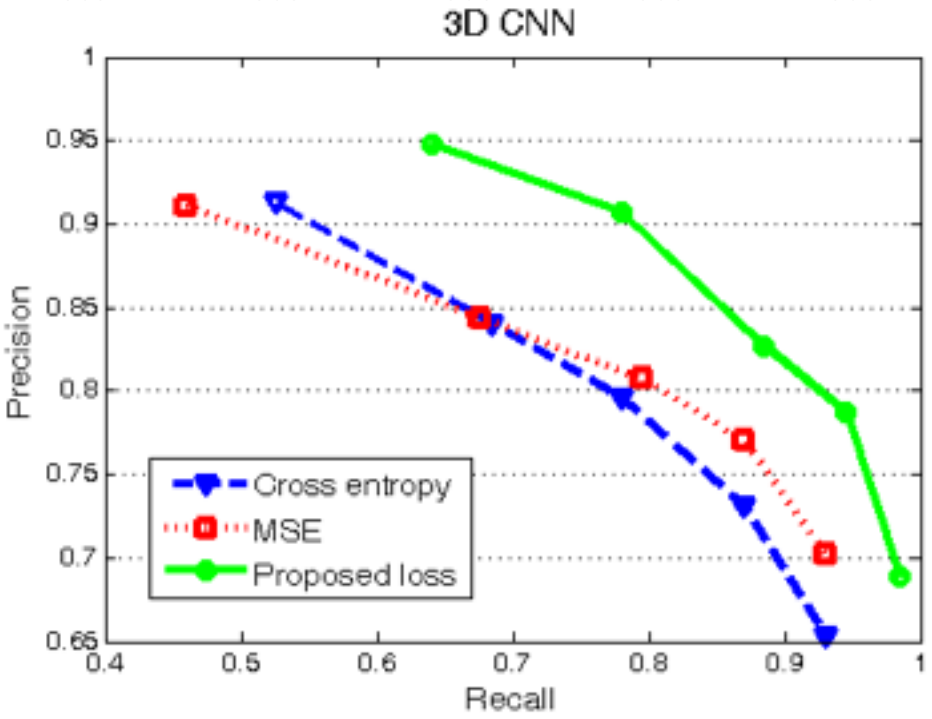
- Data (5,800 video clips):
  - Training: 5,000
  - Validation: 400
  - Testing: 400

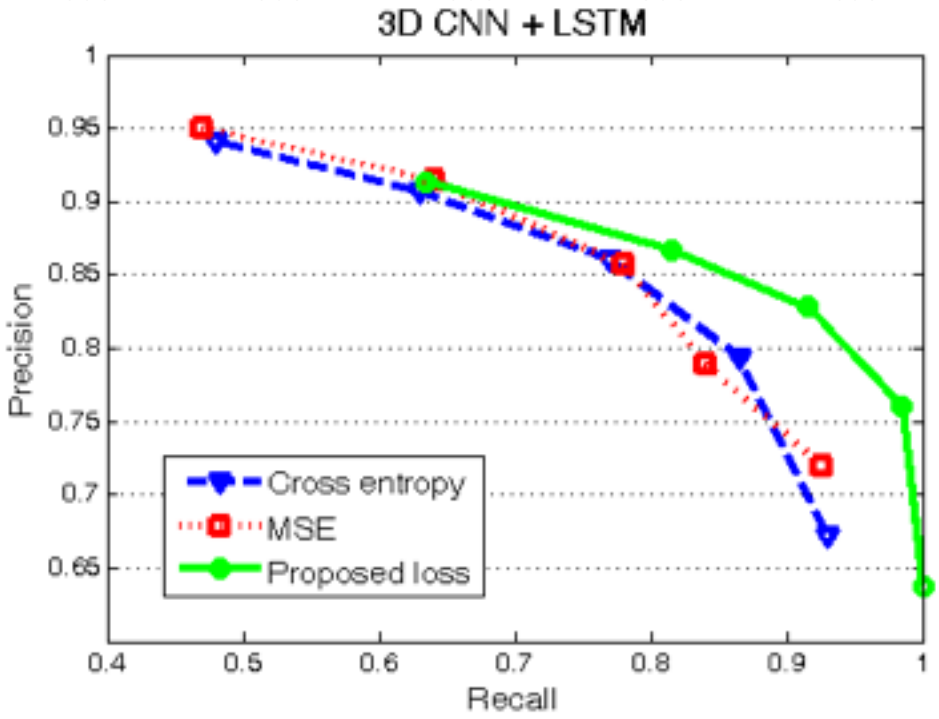
  Training and testing clips capture completely different places.
- Baseline models: 3D CNN and 3D CNN +LSTM models with the cross entropy loss and the MSE loss

Experiment: Measurement

- On the basis of the correctness of classifying a 5-s video clip
- Precision
- Recall
- Classification accuracy

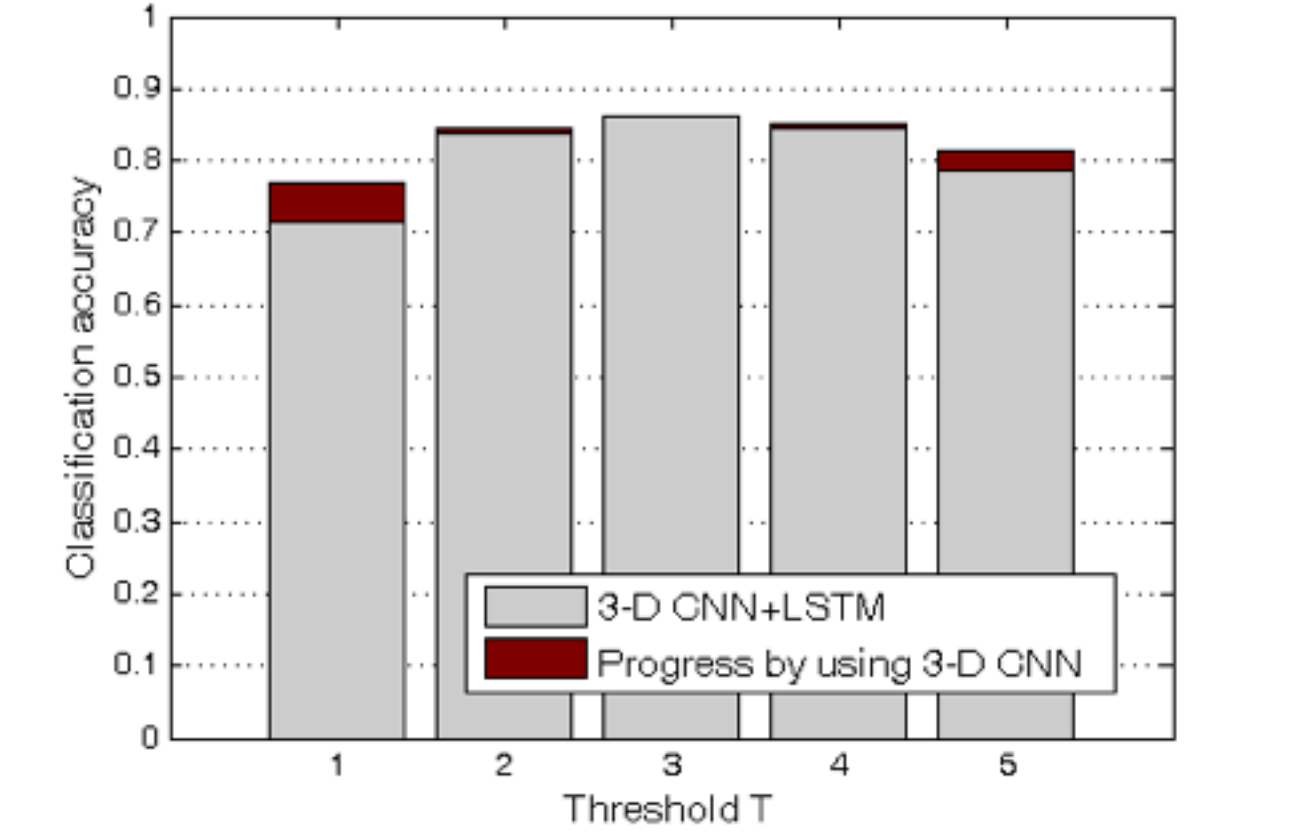
Experiment: Loss comparison





The proposed loss function consistently outperforms MSE and cross entropy.

Experiment: Model comparison



3D CNN slightly improves 3D CNN + LSTM

Experiment: Early detection

Time to Parking Lot	3 s	2 s	1 s	0 s
Cross Entropy	63.5%	68.5%	75.5%	82.0%
MSE	67.0%	72.5%	78.0%	83.0%
Ours	80.0%	87.5%	93.0%	97.0%

Our method can predict an empty parking spot 3 seconds before reaching it with 80% recall and 82.5% precision.

Outline

- Introduction
- Approach
  - Dataset
  - Problem formulation
  - Network architectures
- Experiments
- **Conclusion**

Conclusion

- Early detection of vacant parking spaces can be achieved using dashcam videos.
  - Introduced the first dashcam video dataset for studying early detection of parking spaces
  - Proposed a new loss function for early detection: constraining the subsequent detection scores can considerably improve the detection results
- To be implemented in a dashcam system and to extend the method for parking space localization

The Thirty-First Annual Conference on Innovative Applications of Artificial Intelligence (IAAI-19)

Jan. 31, 2019

Early Detection of Vacant Parking Spaces Using Dashcam Videos

More information:
http://www.csie.ntnu.edu.tw/~myeh