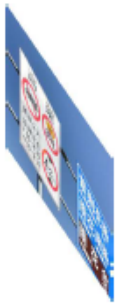
Tsinghua-Tencent 100K-CNN for Traffic Sign Detection & Classification

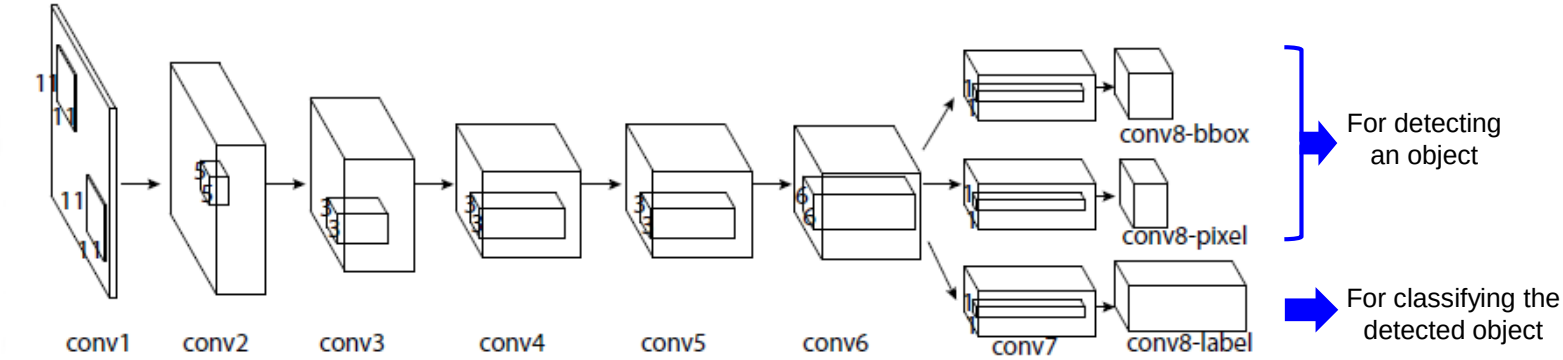
July 17 2017

Jungwon Kang

Introduction to Tsinghua Tencent 100K-CNN

- **TT100K-CNN**
  - *State-of-the-art* CNN for detecting & classifying an one-type object with multiple classes such as *traffic signs*
  - A variant of AlexNet, which branches the network into 3 streams producing outputs for generating bounding boxes with classes
  - Achieved accuracy 0.88 & recall 0.91 on the TT100K dataset, while Fast R-CNN produced accuracy 0.50 & recall 0.56.

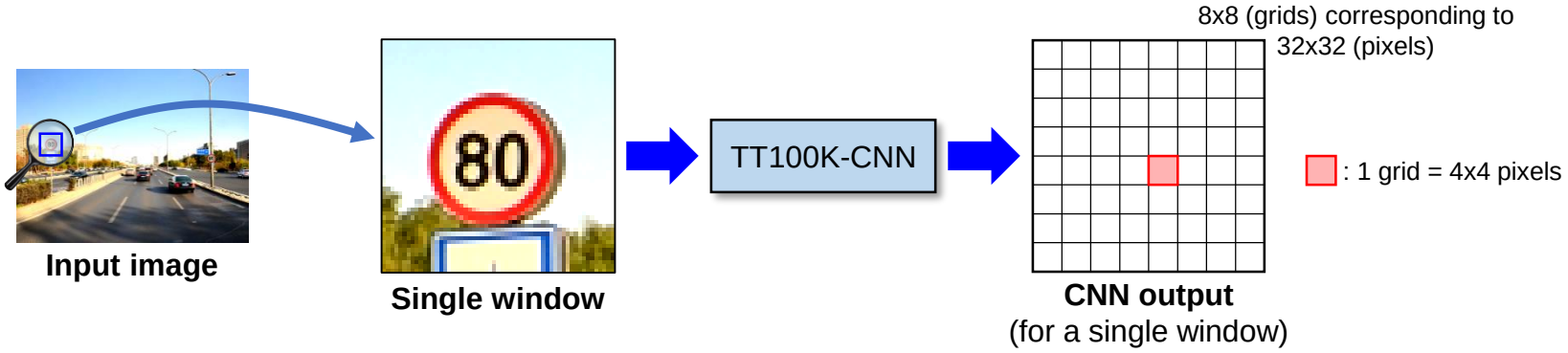




Z. Zhu et al, “Traffic-Sign Detection and Classification in the Wild,” CVPR 2016
(Available in http://cg.cs.tsinghua.edu.cn/traffic-sign/)

TT100K-CNN (1/3)

- **Single Window Feedforwarding in TT100K-CNN**
  - For a single window, TT100K-CNN produces an output of 8x8 grids, where each grid has 3 type output of a value, 4 dim vector, 220 dim vector.

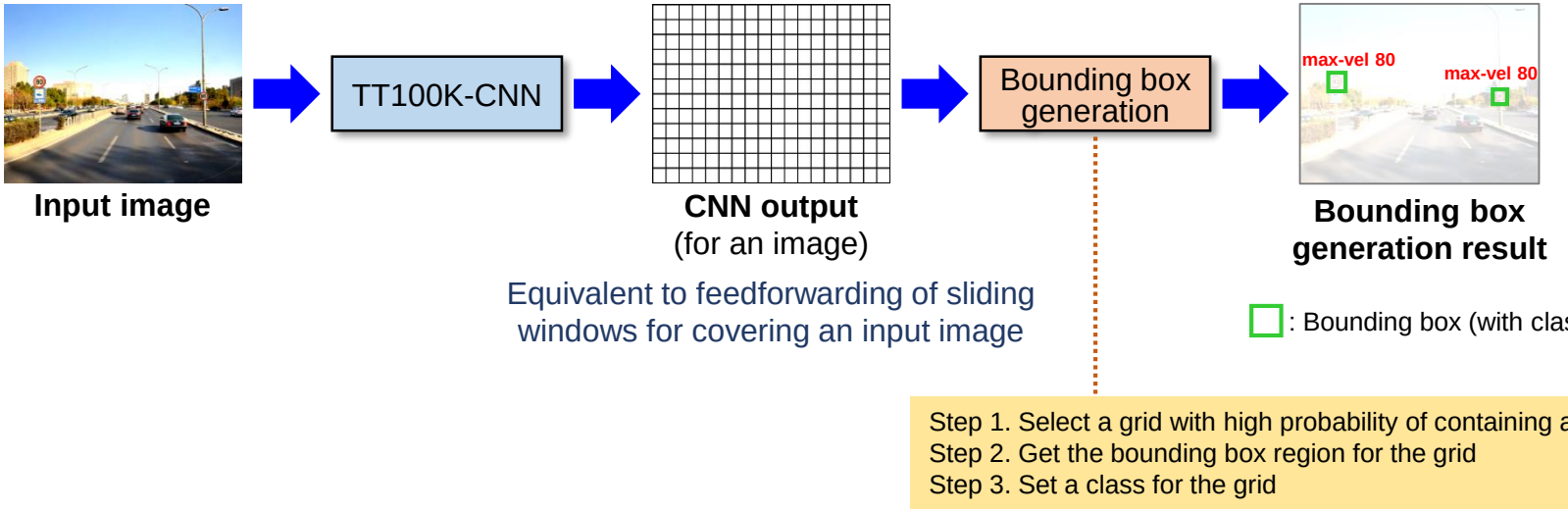


A value	Probability of containing an object
4 dim. vector	Bounding box of the object, i.e. x, y coordinate of left-top & right-bottom points of the box
220 dim. vector	Probability of each class

CNN output in one grid

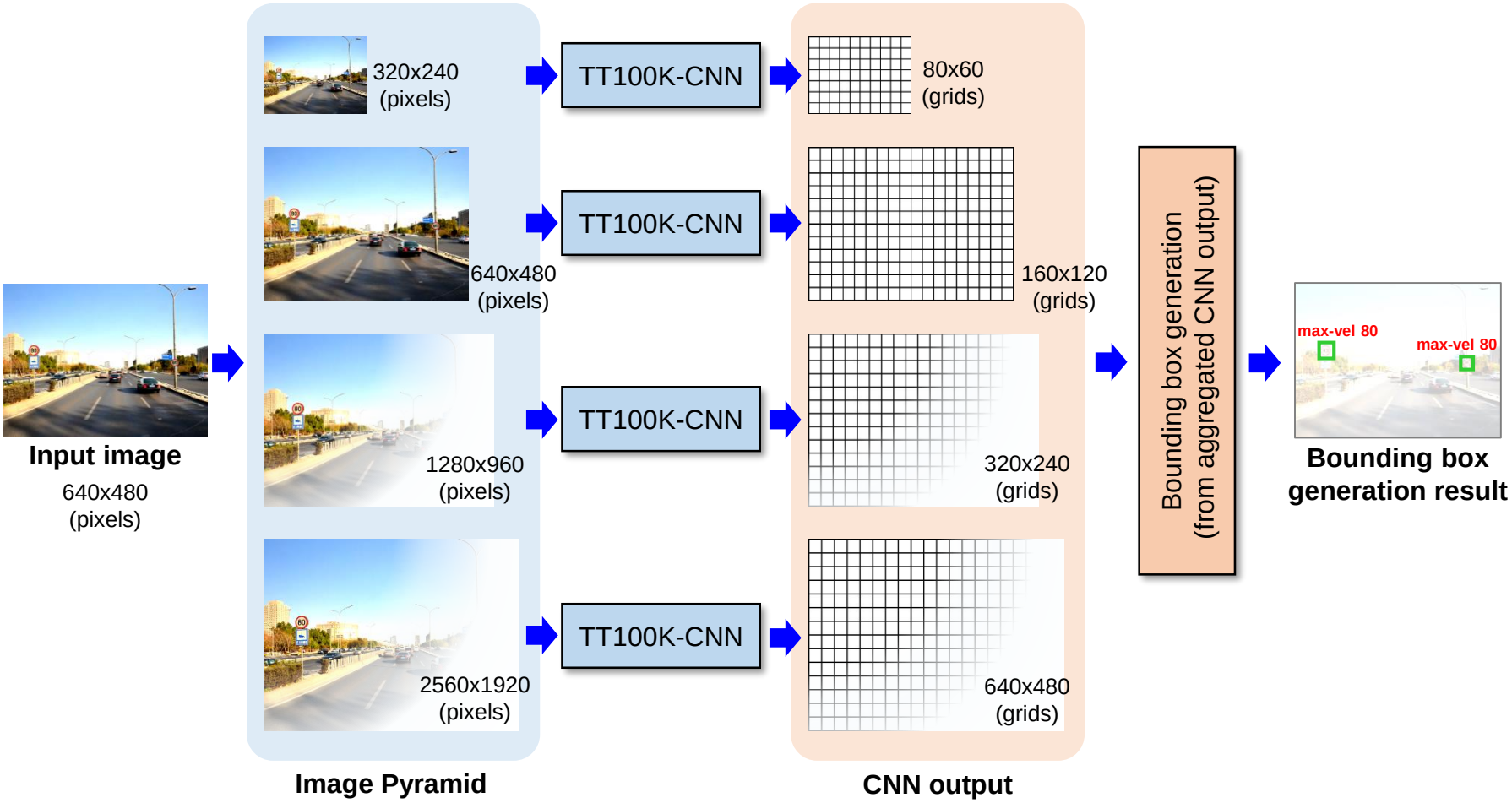
TT100K-CNN (2/3)

- **Bounding Box Generation Using Sliding Windows in TT100K-CNN**
  - Because TT100K-CNN is a fully convolutional neural network, simply feedforwarding of an input image in the network produces an output equivalent to (computationally intensive) feedforwarding of sliding windows for covering a whole region of the input image.
  - From the CNN output, we generate bounding boxes with their classes through three steps.



TT100K-CNN (3/3)

- **Handling Multi-Scale Using Image Pyramid**



Result on Thales Dataset

- **Testing out TT100K-CNN on the ‘Century Park to Yard’ Dataset**
  - Many traffic signs look blurry.
  - Used 4 staged image pyramid with an 2048x2048 input image
  - Took 27s per image on a laptop (with NVIDIA Geforce GTX 1050)

Summary

- Checked the feasibility of applying TT100K-CNN to real traffic sign detection & classification.
- TT100K-CNN surprisingly detected even very small & blurry traffic signs.
- However, long computation time is a bottleneck of TT100-CNN.