UNIVERSITY OF COMPUTER STUDIES, YANGON

Transfer Learning Based Object Detection and Classification

Ei Phyu Myint, Thin Lai Lai Thein

University of Computer Studies, Yangon, Myanmar

Contents

Introduction

- What: Which part of the knowledge can be transferred from the source to the target.
- When: Utilizing transfer learning makes the target task performance (or) results to be improved.
- How: Ways of transferring the knowledge across domains (or) tasks can be identified.

Objectives

- To classify objects using transfer learning technique
- To detect objects using region proposal network
- To view the detection and classification results via web browser
- To make comparisons of transfer learning techniques using the proposed dataset

Related Work(1)

- **Co-Tuning for Transfer Learning [1]**
  - The relationship between source categories and target categories is firstly learned.
  - One-hot target labels are translated into probabilistic source labels which collaboratively supervise the fine-tuning process.
  - It is called “Co-Tuning” because both the ground-truth y_t and probabilistic y_s estimated from the category relationship is used to fine-tune the full pre-trained network.

Related Work(2)

- **DELTA: Deep Learning Transfer using Feature Map with Attention for Convolutional Networks [2]**
  - The discriminative features from outer layer outputs are selected through re-weighting the feature maps with a novel supervised attention mechanism.
  - The distance between source and target networks is characterized using their outer layer outputs through paying attention to discriminative parts of feature maps.
  - This distance is incorporated as the regularization term of the loss function.

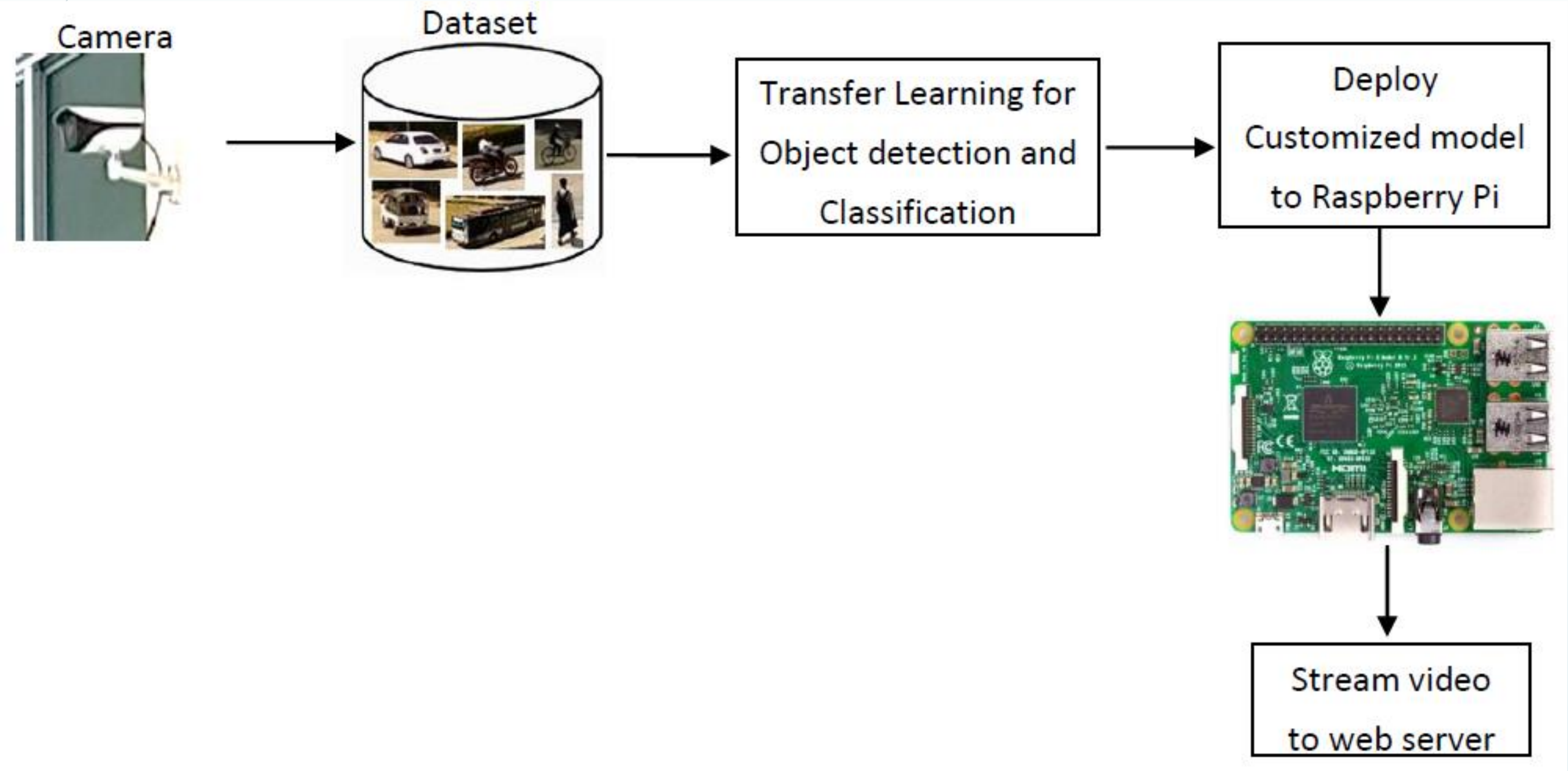
Related Work(3)

- **Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [3]**
  - A region proposal network (RPN) that shares full-image convolutional features with the detection network is introduced.
  - To generate high-quality region proposals which are used by Fast R-CNN for detection, the RPN is trained end-to-end.
  - RPN and Fast R-CNN are merged into a single network by sharing their convolutional features.
  - The RPN component tells the unified network where to look by using neural networks with “attention” mechanisms.

Motivation

- Fast and accurate object detection and classification systems are rising in demand.
- Traditional feature extraction techniques make us handcrafted to extract features from the image.
- Deep learning technique requires a very large amount of dataset.
- Obtaining large data sets gives us to make time consuming.
- Training is very expensive, both in time and resources in order to get large and effective deep learning models.

System Design Overview



Data Set

- The proposed dataset contains 3000 images.
- The dataset is split into training, validation and testing dataset which contains 1920, 600 and 480 images.
- The categories of class are as follows:

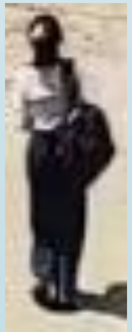

car


bus


truck


motorcycle


bicycle


person

Transfer Learning

- One of the most well-known transfer learning techniques is fine-tuning.
- The first issue of using fine-tuning is that pre-trained models are only partially transferred by discarding task-specific layers and fine-tuning bottom layers.
- The second issue is that by fixing weights in lower convolution layers and re-training weights in upper layers using data from the target domain, parameters of the target model may be driven far away from initial values, which also causes over-fitting.

Co-Tuning

- Suppose that f_0 is a pre-trained deep neural network, y_s is a source category and y_t is a target label to learn category relatioinship.
- The mapping y_s->y_t is firstly learned from $(f_0(x_t),y_t)$ pairs, where $f_0(x) \approx p(y_s|x)$ is a probability distribution over y_s.
- $P(y_s|y_t)$ can be computed from $p(y_t|y_s)$ by Bayes's rule .
- To be reflected with high fidelity of the probability output of source categories of $f_0(x)$, category relationship learning mentions a calibration procedure.

Co-Tuning (Cont.)

Proposed Dataset Class	Top 3 Similar ImageNet Class		
Bicycle	moped	book jacket, dust cover, dust jacket, dust wrapper	Tusker
Bus	minibus	trolleybus, trolley coach, trackless trolley	recreational vehicle, RV, R.V.
Car	moving van	minivan	sports car, sport car
Motorcycle	black grouse	moped	motor scooter, scooter
Person	bearskin, busby, shako	red-breasted merganser, Mergus serrator	abaya
Truck	moving van	amphibian, amphibious vehicle	garbage truck, dustcart

Behavior-based Regularization using Feature Maps with Attentions

- Behavioral Regularization: The distance between the network behaviors (the outer layer outputs, e.g., the feature maps produced by each layer) is considered to regularize the behavior of the networks.
- Feature Map Extraction from Convolution Layers: Each filter of the network from the target dataset is used to get the corresponding output.
- Weighting Feature Maps with Supervised Attention Models: The weights of features are characterized by the potential performance loss when removing these features from the network.

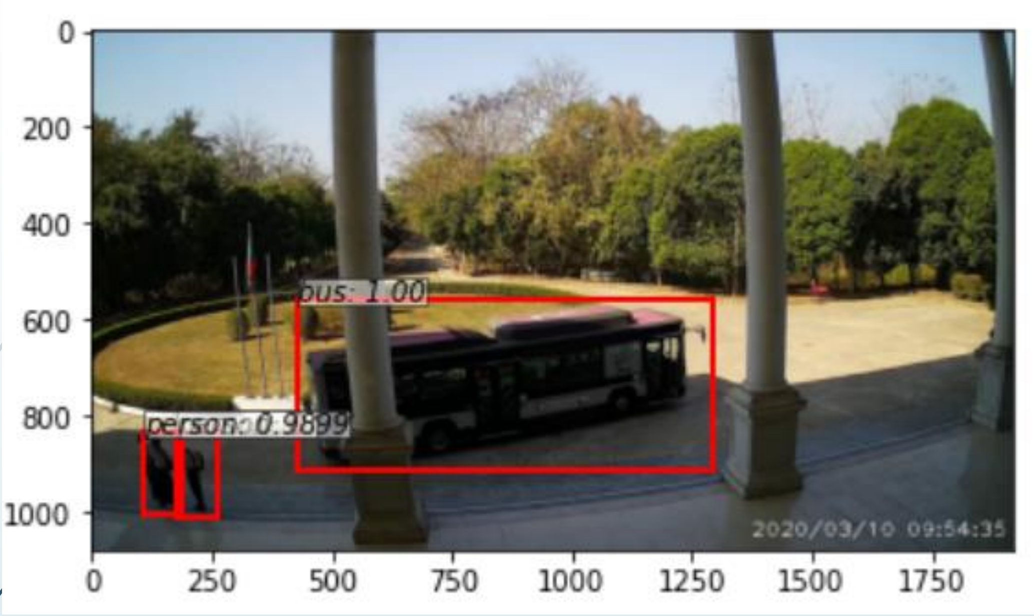
Region Proposal Network based Object Detection

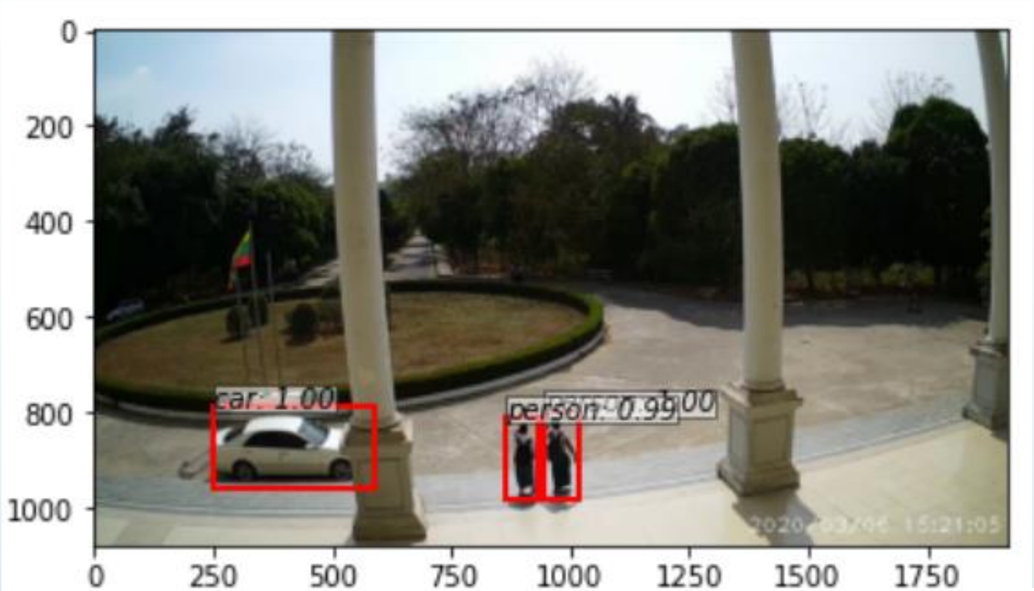
- A small network is slide over the convolutional feature map output by the last shared convolutional layer to generate region proposals.
- Multiple region proposals (the number of maximum possible proposals for each location) called anchors are simultaneously predicted at each sliding-window location.
- The region proposal network (RPN) is trained for the region proposal task.

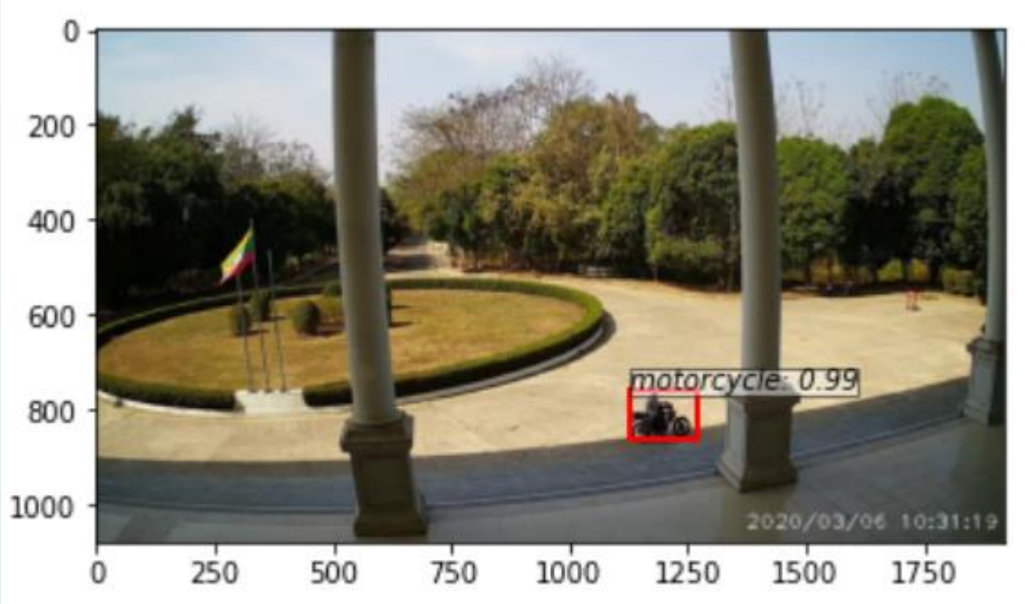
User Interface

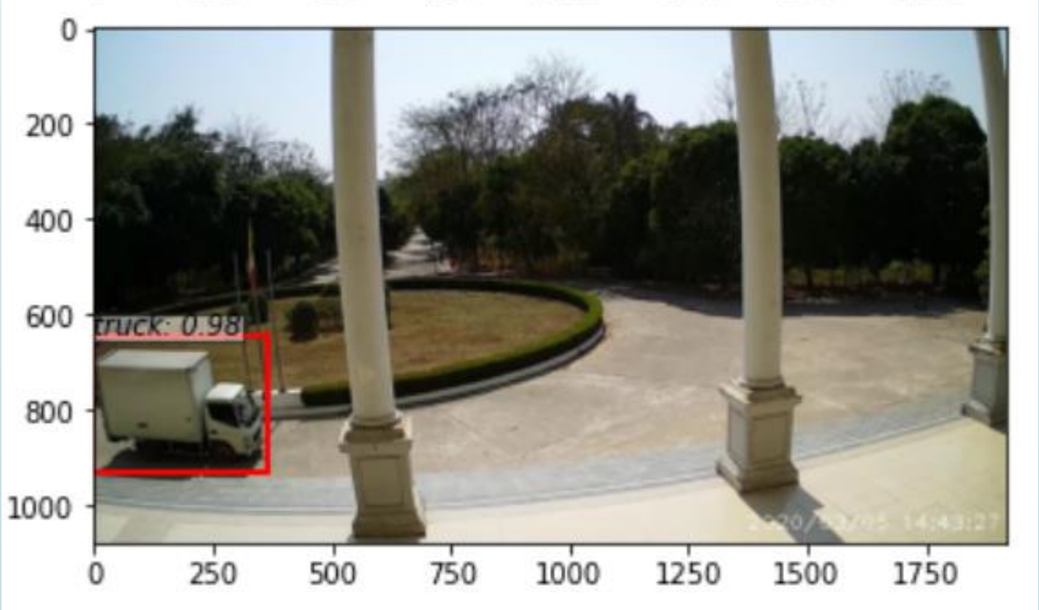
- The Flask web framework is used in order to view the resulted frame in web browser.
- Flask is a popular micro web framework written in the Python programming language.
- Flask is very lightweight, making it super easy to build basic web applications.

Sample Resulted Frames









Performance Evaluation

- The images in the proposed dataset are classified by using transfer learning techniques.
- The learning rate is 0.001, momentum is 0.9 and weight decay is 0.0005 and number of epoch is 100.
- The classification accuracy is as follows:

RenNet50	Fine-tuning(partially transfer)	Co-Tuning	Fine-tuning(fixing weights)	Behavior-based regularization
Proposed dataset	0.91	0.95	0.73	0.99

Conclusion

- This system is implemented using transfer learning technique and pre-trained resnet50 model based on ImageNet dataset is used to generate the customized model.
- A customized model can be built by using transfer learning even if there is a small data set.
- It also makes us to save training time.
- The processed information can also be viewed via web browser.

References

[1] You, Kaichao and Kou, Zhi and Long, Mingsheng and Wang, Jiamin; “Co-Tuning for Transfer Learning”, Advances in Neural Information Processing Systems, vol. 33, 2020.

[2] Li, Xingjian and Xiong, Haoyi and Wang, Hanchao and Rao, Yuxuan and Liu, Liping and Chen, Zeyu and Huan, Jun; “Delta: Deep learning transfer using feature map with attention for convolutional networks”, arXiv preprint arXiv:1901.09229, 2019.

[3] Ren, Shaoqing and He, Kaiming and Girshick, Ross and Sun Jian; “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, arXiv: 1506.01497, 2016.

Thank You