Small datasets for fruit detection with transfer learning*

Dan Dai
LIAT, University of Lincoln
DDai@lincoln.ac.uk

Junfeng Gao
LIAT, University of Lincoln
JuGao@lincoln.ac.uk

Simon Parsons
L-CAS, University of Lincoln
Sparsons@lincoln.ac.uk

Elizabeth Sklar
LIAT, University of Lincoln
ESklar@lincoln.ac.uk

***Abstract*—A common approach to the problem of fruit detection in images is to design a deep learning network and train a model to locate objects, using bounding boxes to identify regions containing fruit. However, this requires sufficient data and presents challenges for small datasets. Transfer learning, which acquires knowledge from a source domain and brings that to a new target domain, can produce improved performance in the target domain. The work discussed in this paper shows the application of transfer learning for fruit detection with small datasets and presents analysis between the number of training images in source and target domains. This investigation is based on three datasets: two containing tomatoes and one containing strawberries. Experimental results indicate that transfer learning can enhance prediction with limited data.**

***Index Terms*—fruit detection, limited datasets, transfer learning, target domain**

I. Introduction

Within the task domain of plant phenotyping, fruit detection is a difficult problem, particularly when trying to identify objects of interest in small image datasets. *Deep learning* is a common approach, using multi-layered *Convolutional Neural Networks (CNNs)* to obtain feature maps, but these networks require sufficient numbers of training examples in order to produce accurate results. *Transfer learning* enables reusing knowledge acquired previously from other tasks or applications and could greatly improve the performance of learning by avoiding various expensive efforts [7].

Recently, several deep learning architectures have been developed from the basic *Region-based CNN (R-CNN)* [4], including *Faster R-CNN* [9], *YOLO* [8] and *Single Shot MultiBox Detector (SSD)* [6]. Most of the machine learning approaches to fruit detection apply these Faster or Mask R-CNN methods [2], [13]. In contrast, transfer learning approaches applied to the agriculture domain mainly focus on identifying plant species [5], classifying pests [12] or diseases [1].

The general principle underlying transfer learning is to take a model trained from data in a *source* domain and adjust this model to a new dataset in a *target* domain. Research in this area has explored the impact of the size of the source dataset and number of labelled examples on the results [1], [2]; but little work has studied these properties in the target domain. The work presented here asks the following questions: Is the size of the source and/or target training sets associated with the accuracy of detection? Is it possible to get the ideal performance in the target domain without carrying out training on large amounts of annotated data (source domain)?

As we already have some knowledge learned from the source domain, therefore it can be saving model training time and resources consumed for the task.

II. Methods

This section presents the basic SSD [6] framework we applied for strawberry and tomato detection, then introduces the transfer learning methods employed. We analyse the relationship between the training dataset size in the source domain and the number of labels in the target domain with respect to the results obtained.

SSD is a one-stage detection system; it eliminates proposal generation and subsequent pixels or feature re-sampling stages, then encapsulates all computation in a single network. This model contains multi-scale feature maps and convolutional predictors for detection, sets default bounding boxes and aspect ratios and allows for different default bounding box shapes in several feature maps to discretize the space of predicted bounding boxes efficiently.

The experiments presented here explore the application of transfer learning for detecting fruits in images. The features learned from a source dataset are transferred to two different new target datasets, each of which may not contain enough training data, due to a paucity of examples or labels. We analyse our results by comparing the accuracy values when transferring from the source to each target, investigating the relationship between these metrics and sizes of the source and target training sets.

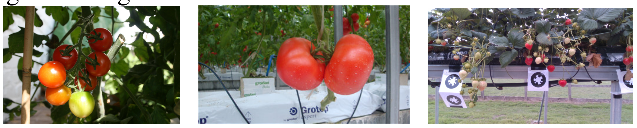
(a) T1 (b) T2 (c) SR

Fig. 1. Tomato and strawberry examples in our datasets.

III. Experimental setup

a) Datasets and Training Parameters: Our experimental data is comprised of three datasets: two different types of tomato (T1, T2) and strawberry (SR). These datasets are collected under different conditions: the strawberry images are all from a polytunnel, whereas the tomato images in T1 are from a garden, showing different growth stages, and the images in T2 are from the Internet.The backgrounds, lighting conditions and other factors differ, so there is some diversity across the data sets. Detailed information about our datasets is shown in Table I and Fig. 1. In our SSD model, the backbone is VGG-16 [11] and is pre-trained with ImageNet [3]. The batch size is 4 and the learning rate is $1e-4$ with the SGD optimizer [10] setting the momentum to 0.9.

TABLE I
NUMBERS OF IMAGES IN OUR DATASETS

Dataset	*Total*	*Training*	*Testing*
T1	496	396 (80%)	100 (20%)
T2	73	59 (80%)	14 (20%)
SR	124	99 (80%)	25 (20%)

b) Transfer learning between tomato and strawberry data sets: Our goal is to quantify the influences of the numbers of training images in the source (T1) and target domains (SR or T2). The number of training images are randomly partitioned as follows: into four parts for the source domain, T1[1]; into twelve parts for SR[2]; and into seven parts for T2[3]. The number of images in each test set (SR and T2) is almost 20% of each total. Table II shows the results of transferring the T1 model to different sized datasets of SR and T2 images. Results are also shown graphically in Fig 2. For comparison, we also trained SSD models with the SR and T2 datasets (3000 iterations). The mAP values we obtained for these models are 0.354 and 0.789, respectively.

TABLE II
RESULTS OF TRAINING ON SR AND T2 AND SELECTED MAP RESULTS FOR TRANSFER LEARNING FROM DIFFERENT-SIZED T1 DATASETS TO SR AND T2 (BEST PERFORMANCE IN EACH ROW IN BOLD)

Training/test	mAP	Training/test	mAP
(SR) (99/25)	0.354	(T2) (59/14)	0.789

Source (T1) (training/test)	Target	0	10	50	60(59)[1]	90	Avg
62(49/13)	SR	0.076	0.299	**0.393**	0.383	0.380	0.3363
	T2	0.779	0.739	**0.796**	0.788	– [2]	0.7714
124(99/25)	SR	0.108	0.274	0.350	**0.394**	0.338	0.3154
	T2	0.798	0.794	**0.813**	0.764	–[2]	0.7904
248(198/50)	SR	0.067	0.334	0.359	**0.401**	0.381	**0.3368**
	T2	0.827	0.764	0.827	**0.829**	–[2]	**0.7961**
496(396/100)	SR	0.052	0.304	0.303	0.378	**0.380**	0.3240
	T2	0.810	0.759	**0.838**	0.811	–[2]	0.7933
Avg	SR	0.0758	0.3027	0.3513	**0.3890**	0.3698	
	T2	0.8035	0.7640	**0.8186**	0.7980	–[2]	

[1] 60(59) means 60 training images for SR and 59 for T2.

[2] – refers to the fact that T2 has fewer training images (i.e. <90)

If we use the source model without any re-training (i.e. number of target training images is 0), as the number of training images in T1 increases, fruit detection performance in SR decreases. This is because of the feature difference between strawberry and tomato: with more source data training, features learned by the model are more related to tomatoes. In contrast, detection accuracy for T2 improves as the source dataset size increases. We also find that T2 provides better detection results if we do not use any images to re-train the source model.

Examining the relation between the numbers of training images in the source and target datasets, the best average performance is with T1 $= 248$. For T2, the average results are better than training on the target domain only ($0.7961 > 0.789$). As the size of the source training dataset increases, the detection results in the target domain seem to reach a saturation state. This suggests that we don't need to train and label large amounts of data in the target domain in order to get high performance, thus saving model training time and resources consumed for the task. Indeed, judging from the current results, using a target dataset that is almost half the size of the source dataset achieves high detection performance.

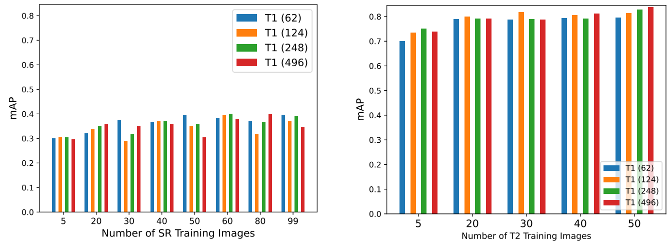

Fig. 2. Results of transfer learning from T1 to SR (left) and T2 (right). mAP is shown for different numbers of images.

IV. SUMMARY AND NEXT STEPS

We applied transfer learning to fruit detection in limited datasets and analysed the impact of the number of the training images in the source and target domains. Next, we will consider how to reduce the features distribution differences between the source and target domains to improve detection performance, discuss and explain the effects of transfer from small samples to large data sets.

REFERENCES

[1] J. G. A. Barbedo. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Computers and Electronics in Agriculture*, 153, 2018.

[2] S. Bargoti and J. Underwood. Deep fruit detection in orchards. In *IEEE Intl Conf on Robotics and Automation (ICRA)*, 2017.

[3] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *IEEE Conf on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[4] R. Girshick and J. Donahue. Rich feature hierarchies for accurate object detection and semantic segmentation. In *IEEE Conf on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[5] A. Kaya, A. S. Keceli, C. Catal, H. Y. Yalic, H. Temucin, and B. Tekinerdogan. Analysis of transfer learning for deep neural network based plant classification models. *Computers and Electronics in Agriculture*, 158, 2019.

[6] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. SSD: Single shot multibox detector. In *European Conference on Computer Vision (ECCV)*. Springer, 2016.

[7] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Trans on Knowledge and Data Engineering*, 22(10), 2009.

[8] J. Redmon and A. Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[9] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, 2015.

[10] H. Robbins and S. Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3), 1951.

[11] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[12] K. Thenmozhi and U. S. Reddy. Crop pest classification based on deep convolutional neural network and transfer learning. *Computers and Electronics in Agriculture*, 164, 2019.

[13] Y. Yu, K. Zhang, L. Yang, and D. Zhang. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Computers and Electronics in Agriculture*, 163, 2019.

[1] The four partitions of the T1 dataset each contain {62,124,248,496} images, respectively.

[2] The twelve partitions of the SR dataset each contain {0,5,10,20,30,40,50,60,70,80,90,99} images, respectively.

[3] The seven partitions of the T2 dataset each contain {0,5,10,20,40,50,59} images, respectively.