Intelligent Automation & Soft Computing
DOI: 10.32604/iasc.2023.033286
Article

Tech Science Press

Guided Dropout: Improving Deep Networks Without Increased Computation

Yifeng Liu[1], Yangyang Li[1,*], Zhongxiong Xu[1], Xiaohan Liu[1], Haiyong Xie[2] and Huacheng Zeng[3]

[1]National Engineering Research Center for Risk Perception and Prevention (NERC-RPP), CAEIT, Beijing, 100041, China
[2]University of Science and Technology of China, Hefei, Anhui, 230026, China
[3]Department of Computer Science and Engineering, Michigan State University, East Lansing, MI, 48824, USA
*Corresponding Author: Yangyang Li. Email: liyangyang@cetc.com.cn
Received: 13 June 2022; Accepted: 26 October 2022

Abstract: Deep convolution neural networks are going deeper and deeper. However, the complexity of models is prone to overfitting in training. Dropout, one of the crucial tricks, prevents units from co-adapting too much by randomly dropping neurons during training. It effectively improves the performance of deep networks but ignores the importance of the differences between neurons. To optimize this issue, this paper presents a new dropout method called guided dropout, which selects the neurons to switch off according to the differences between the convolution kernel and preserves the informative neurons. It uses an unsupervised clustering algorithm to cluster similar neurons in each hidden layer, and dropout uses a certain probability within each cluster. Thereby this would preserve the hidden layer neurons with different roles while maintaining the model's scarcity and generalization, which effectively improves the role of the hidden layer neurons in learning the features. We evaluated our approach compared with two standard dropout networks on three well-established public object detection datasets. Experimental results on multiple datasets show that the method proposed in this paper has been improved on false positives, precision-recall curve and average precision without increasing the amount of computation. It can be seen that the increased performance of guided dropout is thanks to shallow learning in the networks. The concept of guided dropout would be beneficial to the other vision tasks.

Keywords: Neural network; guided dropout; object detection; shallow learning

1 Introduction

In recent years, deep learning technics [1–3] have greatly improved the performance of neural networks. Dropout [4] is an effective trick of deep learning for reducing overfitting in neural networks that works by randomly switching off 50% of nodes in a network during a training epoch. It has also been previously tested on large well-known datasets such as ImageNet [5]. The standard dropout randomly sets hidden neuron activities to zero with a certain probability during training.

So why does dropout switch off the neurons randomly? Actually the hidden layer neurons with different attributes play the different roles, while some neurons have similar attributes. Hence, the different roles of

neurons should be regarded as a priori to decide whether to switch off neurons, rather than treat them equally. We propose a method called "guided dropout", whose main idea is to cluster similar neurons in each hidden layer and dropout using a certain probability within each cluster. It helps to make the hidden layer neurons with different roles to be well preserved while maintaining the scarcity and generalization of the model. At the same time, it will increase the drop probability for similar neurons in the hidden layers. This is an idea of deep model guided by shallow model.

The avoidance of handcrafted features engineering may have both advantages and shortcomings. An appropriate orientation plays a vital role in daily life, as well as in deep learning. The use of shallow learning for proper guidance is conducive to deep learning to get faster convergence and improve training models' accuracy. This paper focuses on using shallow learning model in dropout to set some neurons to zero with selective probability according to the difference of neurons.

In this paper, the proposed guided dropout uses the results obtained from shallow learning to reduce the certain drop probability of the hidden layer neurons that contribute differently to the object detection task [6–9]. Human detection [10] is a challenging task in computer vision and it would be more challenging than general object detection. Our experiments were conducted using the three human detection datasets. To test the effectiveness of guided dropout against that of the standard dropout on more challenging datasets, we apply the guided dropout to the faster regions with convolutional neural network features (Faster R-CNN) [11]. Our method has achieved better results on three datasets without increasing the amount of computation.

This paper organizes as follows. Section 2 outlines the related works on dropout methods. Section 3 goes into detail on guided dropout. Sections 4 and 5 outline the experiment results and conclusions.

2 Related Works

Recently some other alternative dropout methods have been proposed. These include a method of adaptive dropout [12] proposed by Ba et al. where they modify the probability of a node being switched off using a separate network, which allows the network to find the optimum drop rate for any given node. Another dropout method [13] proposed by Duyck et al. applies optimization methods to dropout, such as simulated annealing. These alternative dropout methods improve the performance but ignore the features of the neurons.

Dropout has also been used innovatively by Wan et al. to create a new method called DropConnect [14], where a random subset of the weights on a node is randomly set to zero instead of all of a node's weights. Barrow et al. [15] present the Selective Dropout method for improving the effectiveness of the dropout method over the same training period by selecting neurons to be dropped through several statistical values.

More than ten years ago, Dalal et al. [16] proposed the histograms of oriented gradients (HOG) descriptor to obtain the robust human feature and built the Institut National de Recherche en Informatique et en Automatique (INRIA) dataset. Although the problem of human detection has been well studied, it remains a challenging study due to various changes in many factors. Recently, a few works based on a shallow model for human detection focusing on developing performance and accelerating computation have achieved inspiring results. Nam et al. [17] proposed the aggregate channel features (ACF) method, which uses the pyramid method to calculate the characteristics of each channel and dramatically accelerates the detection without sacrificing performance. Liu et al. [18] proposed that the approach uses binarized normed gradients to generate a small set of estimation proposals efficiently and formulates segmentation by weighted aggregation and perceptual hash into a joint descriptor to improve the detection performance significantly. During the same year, the ACF improved the local decorrelation channel features (LDCF) [19] method to enhance the performance further and achieved the best results on the shallow model.

In the aspect of deep models, Ren Shaoqing and He Kaiming et al. improved Faster R-CNN based on the deep model and a series of subsequent improvements [20–23] have significantly improved the detection performance. From then on, object detection enters the real-time and high-performance epoch.

3 Approach

During a training epoch, the standard dropout method leads the neurons in the hidden layer to switch off with a certain probability of increasing the sparseness of the network. Each training sample can provide gradients for a different, randomly sampled architecture so that the final neural network efficiently represents a considerable ensemble of neural networks with good generalization capability. In the dropout method, a thinned network is sampled from the complete set of possible networks with a certain probability for each mini-batch. Gradient descent is then applied to the thinned network. This is an embodiment of the ideas of the "ensemble model" [24] and "average model," which enhances the robustness of the model.

In fact, a large number of hidden layer neurons in a deep network usually have similar attributes. In other words, the hidden layer neurons with similar attributes can be approximated by some types of transformation. Hidden layer neurons with different attributes play the more critical roles. For example, for image recognition, generally, it is necessary to describe convolution kernel types of features such as edges [25], angles, and planes. As shown in Fig. 1, a plurality of similar convolution kernels corresponds to the edge convolution kernels. These similar convolution kernels may produce similar feature maps.

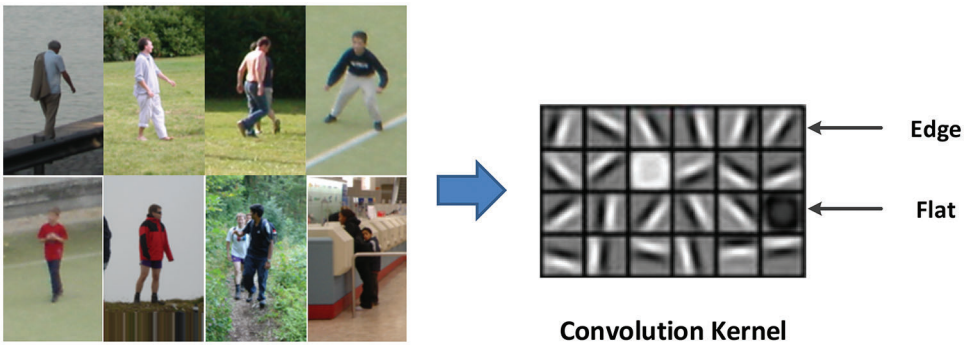

Figure 1: Different types of convolution kernels

Therefore, it is not the optimal mode that all neurons in the dropout are randomly set to 0 with a certain probability. The mode should be set to zero selectively according to the difference in the hidden layer's neurons. In other words, the probability that the hidden layer neurons with different roles are set to zero should be smaller than the hidden output neurons of similar effects.

This paper proposes a new alternative dropout method based on unsupervised shallow learning. The main idea is first to use an unsupervised clustering algorithm to cluster similar neurons in each hidden layer and dropout using a certain probability within each cluster. This method would make the hidden layer neurons with different roles to be well preserved while maintaining the scarcity and generalization of the model.

Consider a neural network with L hidden layers. Let $l \in \{1, ..., L\}$ index the hidden layers of the network. Let $z^{(l)}$ denote the vector of inputs into layer l, $y^{(l)}$ denote the vector of outputs from layer l. $W^{(l)}$ and $b^{(l)}$ are the weights and biases at layer l.

Different from the standard dropout, as shown in Fig. 2, for any layer l, $r_k^{(l)}$ is a vector of independent Bernoulli random variables in every cluster C_k each of which has a probability p of 1. This vector is sampled and multiplied element-wise with the outputs of that layer, $y^{(l)}$, to create the thinned outputs $\tilde{y}^{(l)}$ in every cluster C_k.

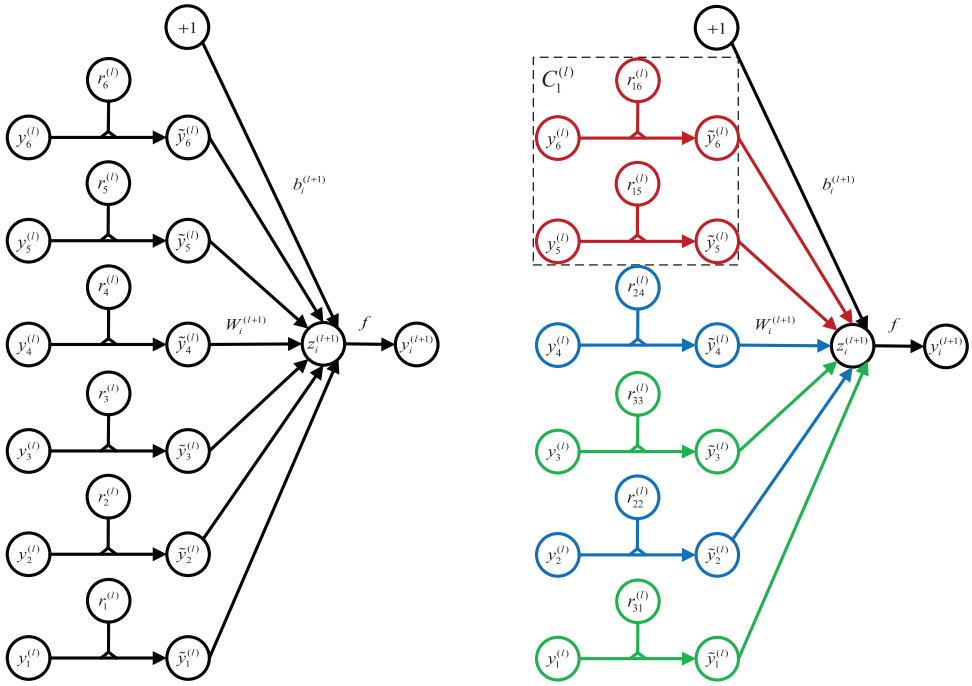

Figure 2: Comparison of the standard dropout and the guided dropout

The detailed algorithm is described in Table 1. K-means++ is used to cluster the nodes in the hidden layer. By evaluating the clustering results, the optimal cluster number is selected from 2 to 8. Within each cluster, the algorithm does dropout using a certain probability. We mainly provide a framework to demonstrate the effectiveness of guided dropout. It is possible to obtain better results by replacing the clustering algorithm and the number of clusters.

Table 1: Guided dropout algorithm

Algorithm 1 guided dropout
Input: all output nodes $\{y_i^l\}$ in a hidden layer l, drop probability p, $i \in [0,\ n]$ **Output:** non-zero output nodes $\{\tilde{y}_i^l\}$ of with guided dropout **Initialize:** $p = 0.5$, number of clusters $k \in [2,\ 8]$ **(I) for** k = 2 : 8 **do** 1) clustering the output nodes $\{y_i^l\}$ by k-means++ [a] and obtain the k clusters $\{C_k\}$, for $C_k = \{y_i^{l_k}\}$, $\sum C_k = \{y_i^l\}$ 2) Minimizes the validity index S_Dbw[b] with different k, obtain the optimal k value and optimal k clusters $\{C_k\}$, for k = arg min $S\_Dbw(k)$ **(II)** Set to zero the output of each hidden neuron in each cluster C_k with probability p and obtain the non-zero output nodes $\{\tilde{y}_i^l\}$.

Notes: [a] The k-means++ [26] improves both the speed and the accuracy of k-means.

[b] The definition of S_Dbw [27] indicates that both criteria of "good" clustering are properly combined, enable reliable evaluation of clustering results.

4 Experiments

4.1 Experiment Environment

The experiments were implemented on the workstation with Intel Core i7-6900k 3.6 GHz processor, 64 GB random access memory (RAM) and Nvidia Titan Xp 12GB. We implemented the proposed method on Python 2.7.12, Compute Unified Device Architecture (CUDA) 10, Compute Unified Deep Neural Network library (cuDNN) 7 and our modification of the Caffe library (https://github.com/rbgirshick/caffe-fast-rcnn/tree/0dcd397b29507b8314e252e850518c5695efbb83).

4.2 Experiment Details

We apply guided dropout to Faster R-CNN, called guided Faster R-CNN. For convenience, the guided Faster R-CNN proposed in this paper uses the ZF (Zeiler & Fergus) model [28] and the ImageNet 2012 pre-training model. The short side of the image is scaled to 600 pixels. The size of the minibatch is 128. The multi-scale anchor in the region proposal network (RPN) selects three areas which are $\{128^2, 256^2, 512^2\}$ and also select {1:1, 2:1, 4:1} three different length-width ratios. RPN produces Top-2000 candidate boxes for training. In all proposals anchor, we assign a positive label if an anchor that has an intersection-over-union (IoU) overlap higher than 0.7 with any ground-truth box. We assign a negative label to a non-positive anchor if it IoU ratio is lower than 0.3 for all ground-truth boxes. During testing, the RPN generates Top-300 proposals and the non-maximum suppression (NMS)'s IoU threshold is fixed at 0.7. Experiments in this article will use the above default parameters unless a special illustration is specified.

4.3 Experiment Results

The compared approaches are Faster R-CNN with standard dropout and LDCF.

In order to evaluate our guided dropout method comprehensively and objectively, we have made experiments on 3 well-established public object detection datasets, namely the INRIA person dataset, Eidgenössische Technische Hochschule Zürich (ETH) pedestrian dataset (Setup 1 (chariot Mk I)) [29], and the more challenging Caltech dataset [30]. The INRIA person dataset includes 1805 cropped person images with a resolution of 64×128 pixels. The images are mainly standing people and have different orientations and backgrounds including crowds. The ETH pedestrian dataset is captured in crowd streets, which have 5 image sequences with different distances. It provides hard cases for pedestrian detection scenarios. The Caltech dataset is a widely used large pedestrian detection dataset, including about 250k labeled frames and 350k bounding boxes with a resolution of 640×480 pixels. It is captured in the streets of different cities, which reflect realistic scenarios for pedestrian detection. In experiment results on both three datasets, we used false positives per image (FPPI) as the evaluation metric, which is more suitable for evaluating the performance on full images. Moreover, we evaluated our approach with a precision-recall curve and average precision (AP).

The miss rate-false positives per image (FPPI) curves on the INRIA test set, the ETH Setup1 test set, and the Caltech test set are shown in Figs. 3–5. It can be seen that on the INRIA test set, the average miss rate of the method proposed in this paper is 0.7% lower than Faster R-CNN's from 10^{-2} to10^{0} FPPI, the average precision of the proposed method is more than Faster R-CNN's from 0.1 to 1 recall. Between the same FPPI as previous, the average miss rate of the method proposed in this paper is 1.5% lower than Faster R-CNN on the ETH Setup1 test set, and the average precision of the proposed method is more than Faster R-CNN's from 0.1 to 1 recall. Moreover, On the Caltech test set, the average miss rate of the method proposed in this paper is 2.4% lower than Faster R-CNN's, the average precision of the proposed method is more than Faster R-CNN's. It can be found that although LDCF is a shallow model-based method, it is still very closed to, even slightly better than, the deep learning model on INRIA.

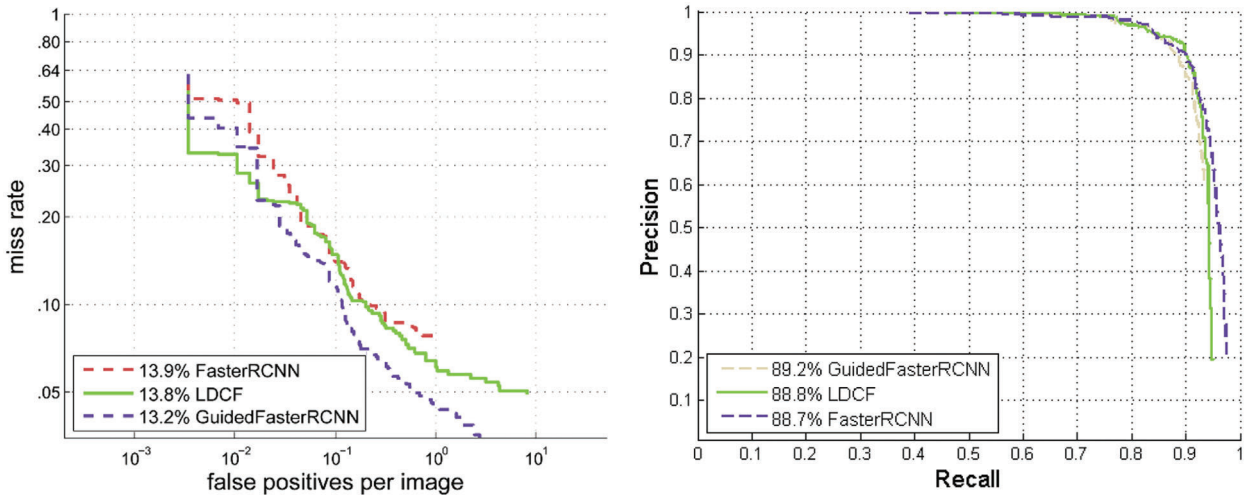

Figure 3: Experiment results on the INRIA dataset

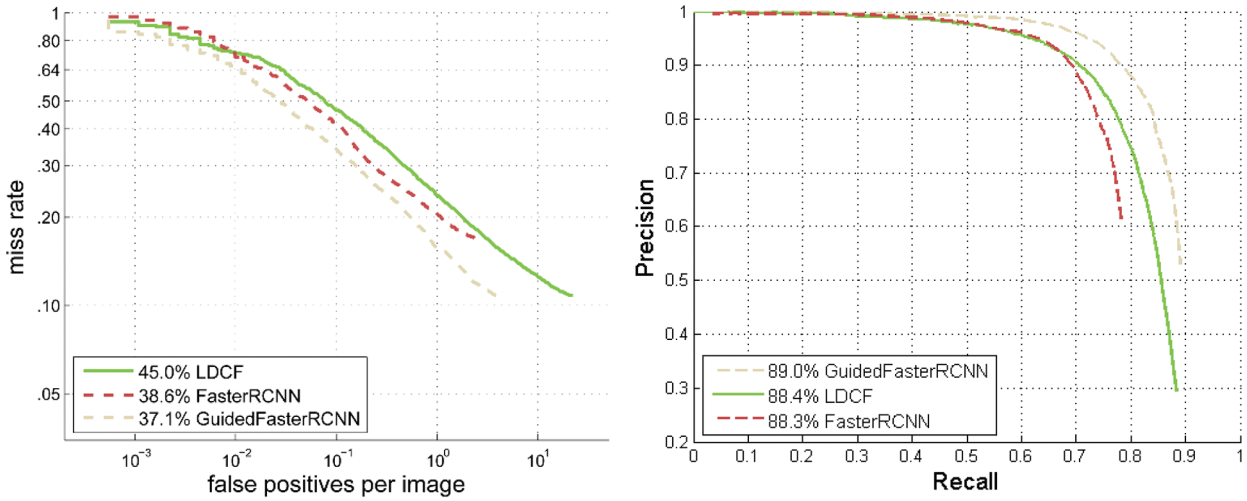

Figure 4: Experiment results on ETH Setup1 dataset

The average precision (AP) for the different methods on the INRIA test set, ETH Setup1 test set, and Caltech test set are shown in Table 2. The AP of guided Faster R-CNN increased by 0.5% compared to Faster R-CNN's on the INRIA test set. The AP of guided Faster R-CNN increased by 0.7% compared to Faster R-CNN on the ETH Setup1 test set. The AP of Guided Faster R-CNN increased by 0.3% compared to Faster R-CNN's on the Caltech test set.

Examples of the detection of partial images on the INRIA test set, the ETH dataset Setup1 test set and the Caltech test set are shown in Fig. 6. The figure shows the comparison between standard dropout and guided dropout. The first line in Figs. 6a–6c represents the detection effect of standard dropout, and the second line represents the detection effect of guided dropout. Among them, Fig. 6a shows that the method of this paper improves recall (true positives) and Fig. 6b shows that the method of this paper reduces false positives. Tiny pedestrians are detected more precisely and some human-like architecture can be avoided compared with standard dropout. Fig. 6c shows that the method of this paper makes bounding boxes more accurate.

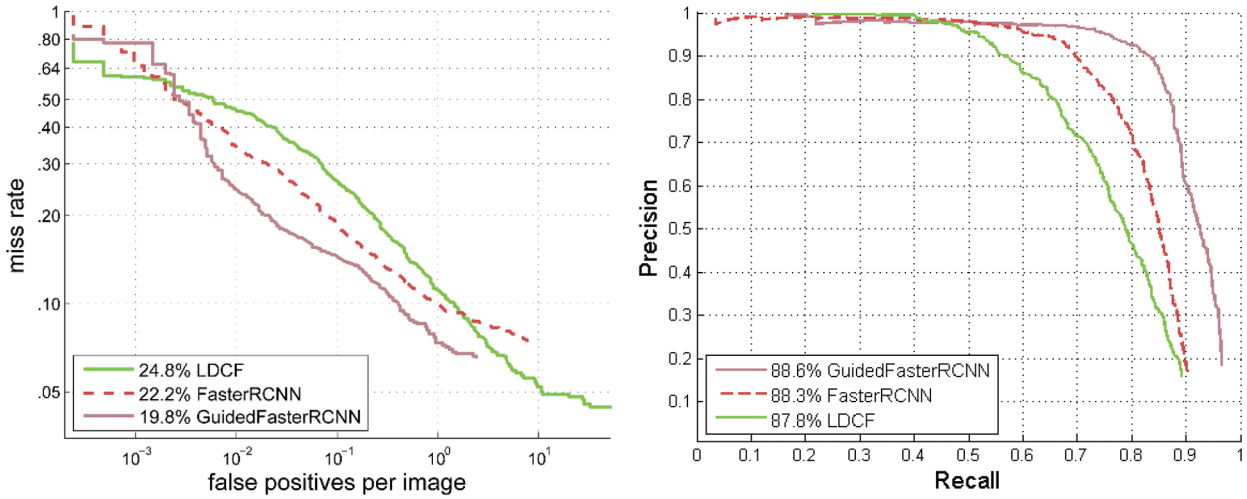

Figure 5: Experiment results on the Caltech dataset

Table 2: Average precision on INRIA/ETH/Caltech dataset

Methods	AP (%)		
	INRIA	ETH	Caltech
Guided faster R-CNN	89.2	89.0	88.6
LDCF	88.8	88.4	87.8
Faster R-CNN	88.7	88.3	88.3

In terms of miss rate, Faster R-CNN is better than LDCF on ETH and Caltech, but LDCF is better than Faster R-CNN on INRIA. In terms of average precision, LDCF is slightly better than Faster R-CNN on INRIA and ETH, but Faster R-CNN is better than LDCF on Caltech. It indicates that, on the whole, the gap between the shallow model and the deep model in the miss rate is greater than the average precision. When the human body is large, the miss rate of the shallow learning model is lower than that of the deep model. When the human body is small, the average precision of the deep model is higher than that of the shallow model. Guided Faster R-CNN achieved the best results, which indicated that it combined the advantages of the deep model and shallow model.

From the above experimental results, the guided dropout proposed in this paper has achieved the best detection results on three datasets. The proposed method still has good generalization when the background is complex and changeable. By analyzing of miss rate-FPPI curve, precision-recall curve and AP, it can be seen that guided dropout is better than standard dropout, thanks to the guide of shallow learning in the networks.

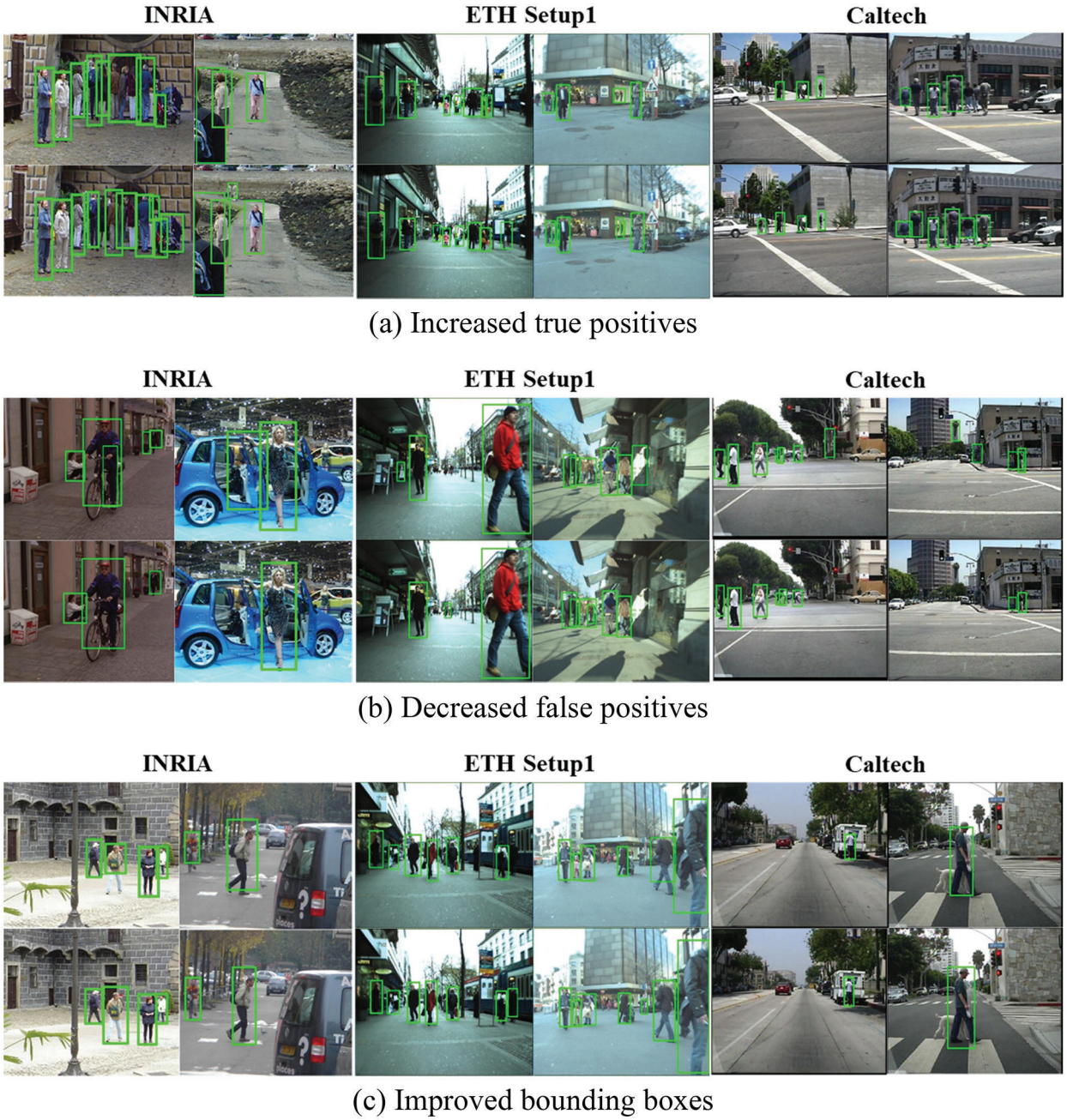

Figure 6: Detection examples on three datasets. Detected persons are enclosed in rectangles. Both in (a), (b) and (c): First row: standard dropout; Second row: guided dropout

5 Conclusion

This paper proposes an effective dropout method, which takes advantage of shallow learning. The work in this paper explores some new ideas for the study of deep learning. The core value focuses on using the guided and reasonable probability to dropout neurons based on an unsupervised clustering algorithm and leads to better results than the standard dropout approach. In this work, the proposed method has been conducted using three challenging datasets and achieved the best experimental results. On this basis, we

conclude that the proper combination of deep learning and shallow learning may achieve better results. The other conclusion that can be drawn is that the gap between the shallow model and the deep model in the miss rate is greater than the average precision. In future work, we will study task-aware guided dropout and simplify the training process. In addition, we try to apply the concept of guided dropout to improve the performance of deep networks in other fields such as telemedicine [31] and blockchain [32].

Funding Statement: This work is supported by the National Natural Science Funds of China (Project No. U19B2036).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] Y. Xue, Y. H. Tang, X. Xu, J. Y. Liang and F. Neri, "Multi-objective feature selection with missing data in classification," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 6, no. 2, pp. 355–364, 2022.

[2] Y. Xue, Y. K. Wang, J. Y. Liang and A. Slowik, "A Self-adaptive mutation neural architecture search algorithm based on blocks," *IEEE Computational Intelligence Magazine*, vol. 16, no. 3, pp. 67–78, 2021.

[3] C. Song, X. Cheng, Y. X. Gu, B. J. Chen and Z. J. Fu, "A review of object detectors in deep learning," *Journal on Artificial Intelligence*, vol. 2, no. 2, pp. 59–77, 2020.

[4] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[5] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh *et al.,* "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[6] M. Lu, S. Niu and Z. Gao, "An efficient detection approach of content aware image resizing," *Computers, Materials & Continua*, vol. 64, no. 2, pp. 887–907, 2020.

[7] G. H. Yu, H. H. Fan, H. Y. Zhou, T. Wu and H. J. Zhu, "Vehicle target detection method based on improved SSD model," *Journal on Artificial Intelligence*, vol. 2, no. 3, pp. 125–135, 2020.

[8] H. Qian, X. Zhou and M. Zheng, "Abnormal behavior detection and recognition method based on improved resnet model," *Computers, Materials & Continua*, vol. 65, no. 3, pp. 2153–2167, 2020.

[9] X. Liu and X. Chen, "A survey of GAN-generated fake faces detection method based on deep learning," *Journal of Information Hiding and Privacy Protection*, vol. 2, no. 2, pp. 29–36, 2020.

[10] M. Fabbri, G. Brasó, G. Maugeri, O. Cetintas, R. Gasparini *et al.,* "MOTSynth: How can synthetic data help pedestrian detection and tracking?," in *Proc. of the IEEE/CVF Int. Conf. on Computer Vision (ICCV)*, Montreal, Canada, pp. 10849–10859, 2021.

[11] S. Q. Ren, K. H. He, R. Girshick and J. Sun, "Faster R-CNN: Towards realtime object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.

[12] L. Ba and B. Frey, "Adaptive dropout for training deep neural networks," in *Proc. Int. Conf. on Neural Information Processing Systems*, Red Hook, NY, USA, pp. 3084–3092, 2013.

[13] J. Duyck, M. H. Lee and E. Lei, "Modified dropout for training a neural network," in *Advanced Introduction to Machine Learning Course, Tech. Rep, School of Computer Science, Carnegie-Mellon University*, Pittsburgh, PA, USA, 2014. [Online]. Available: https://www.cs.cmu.edu/~epxing/Class/10715/project-reports/DuyckLeeLei.pdf.

[14] L. Wan, M. Zeiler, S. Zhang, Y. Lecun and R. Fergus, "Regularization of neural networks using dropconnect," in *Proc. Int. Conf. on Machine Learning*, Atlanta, GA, USA, pp. 1058–1066, 2013.

[15] E. Barrow, M. Eastwood and C. Jayne, "Selective dropout for deep neural networks," in *Proc. Int. Conf. on Neural Information Processing*, Kyoto, Japan, pp. 519–528, 2016.

[16] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, San Diego, CA, USA, pp. 886–893, 2005.

[17] W. Nam, P. Dollár and J. H. Han, "Local decorrelation for improved pedestrian detection," in *Proc. Advances in Neural Information Processing Systems*, Cambridge, MA, USA, pp. 424–432, 2014.

[18] Y. Liu, L. Zou, J. Li, J. Yan, W. Shi *et al.,* "Segmentation by weighted aggregation and perceptual hash for pedestrian detection," *Journal of Visual Communication and Image Representation*, vol. 36, pp. 80–89, 2016.

[19] P. Dollar, S. Belongie and P. Perona, "Fast feature pyramids for object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 8, pp. 1532–1545, 2016.

[20] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 779–788, 2016.

[21] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed *et al.,* "SSD: Single shot multibox detector," in *Proc. European Conf. on Computer Vision*, Amsterdam, Netherlands, pp. 21–37, 2016.

[22] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. on Computer Vision*, Venice, Italy, pp. 2961–2969, 2017.

[23] R. Meng, S. G. Rice, J. Wang and X. M. Sun, "A fusion steganographic algorithm based on faster R-CNN," *Computers, Materials & Continua*, vol. 55, no. 1, pp. 1–16, 2018.

[24] S. Li, X. Cao and Y. Nan, "Multi-level feature-based ensemble model for target-related stance detection," *Computers, Materials & Continua*, vol. 65, no. 2, pp. 1373–1384, 2020.

[25] C. Anitescu, E. Atroshchenko, N. Alajlan and T. Rabczuk, "Artificial neural network methods for the solution of second order boundary value problems," *Computers, Materials & Continua*, vol. 59, no. 1, pp. 345–359, 2019.

[26] D. Arthur and S. Vassilvitskii, "K-Means++: The advantages of careful seeding," in *Proc. the Eighteenth Annual ACM-SIAM Symp. on Discrete Algorithms*, New Orleans, Louisiana, USA, pp. 1027–1035, 2006.

[27] M. Halkidi and M. Vazirgiannis, "Clustering validity assessment: Finding the optimal partitioning of a data set," in *Proc. IEEE Int. Conf. on Data Mining*, San Jose, CA, USA, pp. 187–194, 2001.

[28] M. D. Matthew and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. European Conf. on Computer Vision*, Zurich, Switzerland, pp. 818–833, 2014.

[29] A. Ess, B. Leibe, K. Schindler and L. Gool, "A mobile vision system for robust multi-person tracking," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Anchorage, AK, USA, pp. 1–8, 2008.

[30] C. Wojek, P. Dollár, B. Schiele and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 743–761, 2012.

[31] X. R. Zhang, W. Z. Zhang, W. Sun and A. G. Song, "A new soft tissue deformation model based on Runge-Kutta: Application in lung," *Computers in Biology and Medicine*, vol. 148, pp. 105811, 2022.

[32] Y. J. Ren, Y. Leng, J. Qi, P. K. Sharma, J. Wang *et al.,* "Multiple cloud storage mechanism based on blockchain in smart homes," *Future Generation Computer Systems*, vol. 115, no. 2, pp. 304–313, 2021.