Downloaded from https://onlinelibrary.wiley.com/doi/ by Queen'S University, Wiley Online Library on [25/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

9

Image Classification by Reinforcement Learning With Two-State Q-Learning

Abdul Mueed Hafiz

Department of E&C Engineering, Institute of Technology, University of Kashmir, Zakura Campus, Srinagar, J&K, India

Abstract

In this work, an efficient and simple image recognition classification system has been proposed. It consists of components from both reinforcement learning and deep learning. More specifically, Q-learning is used with an agent having two states, and two to three actions. This classifier is different from others, because the latter uses features of convolutional nets and also uses past histories in addition to Q-states. The other techniques found in literature have issues due to the large number of states used, as the dimensions of their feature maps are quite large. Since the novel technique proposed as only two Q-states, it has the advantage of being simple and also having significantly lesser parameters to optimize. Also, it has a straightforward reward process. Another advantage of the proposed classifier is usage of unique action set for image processing not found in literature. Accuracy of the proposed classifier has been compared with various contemporary algorithms on important datasets from ImageNet, Caltech-101, and *Cats and Dogs*. The classifier given in this work performs better than other classifiers on the various datasets used experimentally.

Keywords: Image classification, ImageNet, Q-learning, reinforcement learning, ResNet50, InceptionV3, deep learning

9.1 Introduction

Reinforcement learning (RL) [1–4] has garnered much attention [4–13]. In computer vision, good initial work [5, 7, 9, 11, 12, 14–20] has been undertaken. In [14], the authors aim to reduce the large computational costs of using large images, by proposing a RL agent which uses iterative selection of the image resolution used in their detector for every image. They have trained their agent with double rewards by choosing lesser resolution for low-level detection of larger objects appearing in the image, and higher resolution for higher-level detection for smaller objects also appearing in the same image. In [20], the authors propose an object detection technique based on reinforcement Q-learning [21, 22]. They use a policy search based on analytic gradient computation with continuous

Email: mueedhafiz@uok.edu.in

Mukhdeep Singh Manshahia, Valeriy Kharchenko, Elias Munapo, J. Joshua Thomas and Pandian Vasant (eds.) Handbook of Intelligent Computing and Optimization for Sustainable Development, (171–182) © 2022 Scrivener Publishing LLC

reward. They report almost two orders of magnitude speed-up over other popular techniques found in literature. In [23], an adaptive deep Q-learning technique has been used for improving and shortening computational time for digit recognition. They refer to their novel network as Q-learning deep belief network (Q-ADBN). This network uses feature maps from a deep auto-encoder [24]. These feature maps are used as active states of the Q-learning technique. After conducting experiments on MNIST dataset [25], the authors of the above work claim that their technique is superior to other techniques in terms of accuracy and running time. The authors of [26] have proposed an image recognition technique which zooms and translates repeatedly in order to refine the Bounding Boxes (Bboxes) of the object categories. The authors of the same have used the VGG-16 CNN [27], concatenated the features alongside the history, and then fed the output to a Q-network. The authors of [28] have proposed an algorithm wherein the agent is learning for deformation of Bbox with the use of various basic transformations, having the goal of obtaining specific object locations. The actions used are horizontal moves, vertical moves, scale changes, and aspect ratio changes.

Taking a hint from the work in area of Maximum Power Point Tracking (MPPT) which is used in Photovoltaic Arrays [29], a simple and efficient technique has been proposed for image classification which gives high accuracy. It is based on deep learning as well as RL. The technique involves using feature maps obtained from a pre-trained CNN like ResNet50 [30], InceptionV3 [31], or AlexNet [32]. Next, RL is used for optimal action proposal generation (rotation by a specific angle or translation) on the image. After application of the final action to the original test image, and obtaining feature map from the CNN [30, 31, 33–37], classification is done using a second classifying structure, like a Support Vector Machine (SVM) [38–40] or a Neural Network (NN) which has been trained on the CNN feature maps of training images.

Many RL techniques [18, 20, 26, 28] used for image classification use actions like zoom and translation according to visual detection in humans. Thus, they miss the important action of rotating the field of view used in human visual image comprehension. In this paper rotation of image by specific angle(s) has been used which is novel in itself. Also Q-Learning has been used in RL. Q-states which have been used in the other approaches of RL-based object detection, use features with high dimensions combined with state history. This technique usually leads to large state-space, in turn leading to optimization problems. Addressing these problems was the motivation behind this work. The proposed technique uses only two states, and two or three actions. As a consequence of this strategy, the Q-table has two rows and two/three columns. To the best of available knowledge, this is the first technique using two Q-states, as well as using image rotation as an action. As a result, the overall task of using RL in computer vision becomes simple and also becomes efficient. Better results are obtained in comparison to other techniques involving CNNs like ResNet50 [30, 37, 41] and InceptionV3 [31].

The major highlights of the paper are as follows:

- A novel image recognition technique is proposed which is based on RL. This is a first to the best of knowledge.
- Rotation of image has been used for image recognition, which is akin to tilting of vision. This is also a first to the best of knowledge.
- The proposed technique outperforms others on all the datasets used in the current work.

Downloaded from https://onlinelibrary.wiley.com/doi/ by Queen'S University, Wiley Online Library on [25/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The rest of the paper is structured as follows. Section 9.2 discusses the proposed approach. Section 9.3 gives a brief description of the various datasets used in the study. Section 9.4 discusses the experimentation. Conclusion is presented in Section 9.5.

9.2 Proposed Approach

In this paper, a hybrid approach of deep learning [7, 14, 42–46] and RL is proposed. The CNNs used are ResNet50 [30, 37, 41], InceptionV3 [31], and AlexNet [32]. First, feature-map of the CNN is obtained for every training sample by feeding it to the CNN. Let the set of all of these feature maps be referred to as $\boldsymbol{F_{Train}}$. A secondary classifier like a SVM, or a NN is trained on $\boldsymbol{F_{Train}}$. For classifying a test sample, a filtering criterion for "hard to classify" samples must be used. If the test sample is tagged as hard, it is classified by RL. If not, then CNN is used for classification. This paper is not about filtering criteria, and hence, no such criterion has been used except that all test samples misclassified by the CNN are tagged as "hard to classify". Every "hard" test sample is first fed to the CNN. Next, the feature map of the CNN, *viz.*, $\boldsymbol{F_{Sample}}$ for the test image is obtained. RL based classification is done as follows. A random action is selected from a bank of actions whose number does not exceed 3 in all the experiments. Next, the action permutation is applied to the test image. Then, the new feature map $(\boldsymbol{F'_{Sample}})$ of the permuted image is obtained from the CNN. The action permutations used in this work are mainly image rotation with specific angles and sometimes diagonal translation. For RL, Q-Learning with random policy is used. Two states ($\boldsymbol{n} = 2$) and "two or three" actions ($\boldsymbol{a} = 2$) *or* ($\boldsymbol{a} = 3$) are used. The current state is decided after observing a metric, *viz.*, "standard deviation" of the prediction scores (of the second classifier) before and after applying the permutation. Let $\boldsymbol{M}$ be metric for original image and $\boldsymbol{M_1}$ be metric after applying current action. The new state is decided based on the criteria whether $\boldsymbol{M_1}$ is lesser than, equal to, or larger than $\boldsymbol{M}$, respectively. The Q-table having two rows ($\boldsymbol{n}$=2) and $\boldsymbol{a}$ columns ($\boldsymbol{a}$ = **2 or 3**) is initialized to zero. Reward $\boldsymbol{r}$ is based on the comparison as shown below:

$$r = \left\{ \begin{array}{rl} +1, & if\ \boldsymbol{M_1} > \boldsymbol{M} \\ 0, & if\ \boldsymbol{M_1} = \boldsymbol{M} \\ -1, & if\ \boldsymbol{M_1} < \boldsymbol{M} \end{array} \right\} \tag{9.1}$$

Number of iterations for updating the Q-table is $N = a \times m$, where $a = 2$ or 3 and m is a constant (usually 20). After each iteration, the Q-value entry for the current "state-action pair" with state s and action a, i.e., $Q(s,a)$ present in the Q-table, is updated as per the Q-Learning Update Rule:

$$Q(s,a) = Q(s,a) + a\left[r + \gamma \max_{\forall b \in A} Q(s',b) - Q(s,a)\right] \tag{9.2}$$

where s' is new state, the learning rate $\alpha = 0.4$, and the discount rate $\gamma = 0.3$. Flowchart for the proposed RL algorithm is given in Figure 9.1. After completing N iterations of Q-Learning,

Downloaded from https://onlinelibrary.wiley.com/doi/ by Queen'S University, Wiley Online Library on [25/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

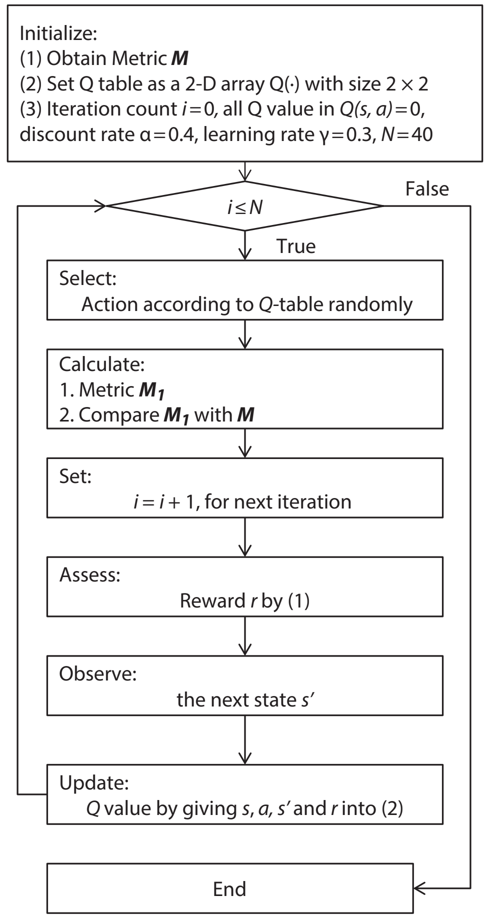

Figure 9.1 Flowchart of the proposed RL algorithm.

optimal action is chosen as the action having highest value in Q-table. Finally, the optimal action is applied to the original sample/test image. Next, the image is fed to the CNN giving its feature map. This feature map is fed to the second classifying structure for classification. Figure 9.2 shows the proposed approach in modular fashion.

The NNs used on top of ResNet50 and InceptionV3 networks are shown in Figure 9.3.

9.3 Datasets Used

Benchmarking has been done on three popular databases, *viz*., ImageNet [47, 48], *Cats and Dogs* Dataset [49], and Caltech-101 Dataset [50].

9.3.1 ImageNet

ImageNet [47] is an extensive database of images for image recognition. It has about 14 million hand-annotated images. Bboxes are also given for almost one million images.

Downloaded from https://onlinelibrary.wiley.com/doi/ by Queen'S University, Wiley Online Library on [25/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

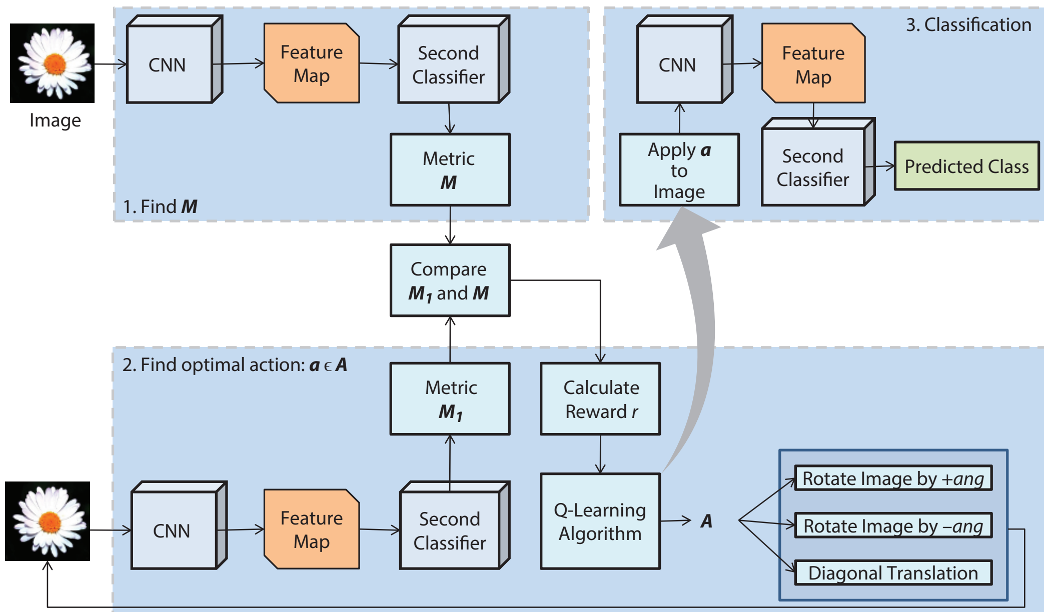

Figure 9.2 Proposed technique.

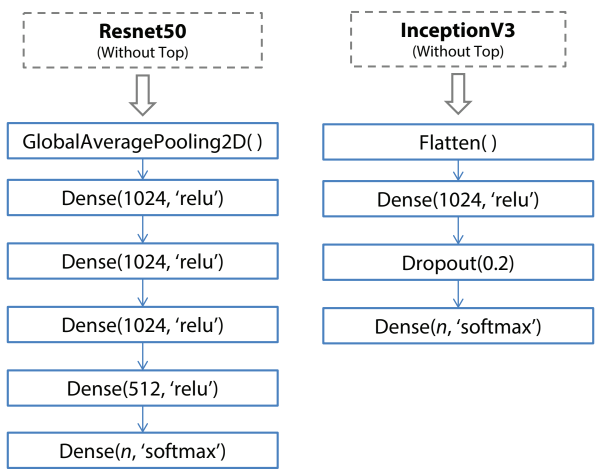

Figure 9.3 NNs used after the CNNs, *viz.*, ResNet50 (left) and InceptionV3 (right), which are implemented in TensorFlow without top; n is number of classification categories.

The database has about 20,000 classes with each class typically consisting of hundreds of images. The annotation-database consists of external image URL entries. The latter are available from the website maintained by the ImageNet Project.

9.3.2 Cats and Dogs Dataset

Web services have often been protected with challenges that are relatively easier for people to solve as compared to computers, e.g., CAPTCHA or Human Interactive Proof (HIP).

Downloaded from https://onlinelibrary.wiley.com/doi/ by Queen'S University, Wiley Online Library on [25/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 9.1 Distribution of data experimentally.

Dataset	Classes used	Training images	Validation images	Testing images
ImageNet [47]	4 (Bikes, Ships, Tractors, Wagons)	1,531	788	745
Cats and Dogs [49]	2 (Cats, Dogs)	2,000	500	500
Caltech-101 [51]	50	750	-	1,250

The latter are available for various actions like email-spam blocking and prevention of hacking. Some HIPs ask people for identification of pictures of dogs and cats. Although it is not easy for machines, it is known that humans can do the same easily. Petfinder.com has provided Microsoft more than three million digital pictures of dogs and cats, which have been already identified by humans. Kaggle offers a subset of this dataset [49] for research.

9.3.3 Caltech-101

Caltech-101 [50] is a database of images created by a group of researchers at the California Institute of Technology, used for image classification and recognition, and also for object detection. The database consists of 9,146 digital pictures with 101 image classes. It also has a set of annotations which describe the outlines in each image.

The distribution of data among the experimental setups is shown in Table 9.1.

9.4 Experimentation

Experiments were conducted on a machine having an *Intel® Xeon®* processor (with 2 Cores), 12.6 GB RAM, and 12 GB GPU. For benchmarking of the performance of the proposed technique, its performance was compared with that of the pre-trained CNN used alone after being fine-tuned on the datasets. TensorFlow [52] has been used for implementing the CNNs (pre-trained, having ImageNet weights) and algorithms. For training the CNNs using transfer learning, 10 *training epochs* were used, with optimizer: *Adam*, Loss: *Categorical Crossentropy*, and Learning Rate: *0.001*. Benchmarking has been done on three popular databases, *viz.*, ImageNet [47, 48], *Cats and Dogs* Dataset [49], and Caltech-101 Dataset [50].

Tables 9.2 to 9.4 show the benchmarking for the performance of the proposed approach on the datasets used. It should be noted that for conventional CNN usage no rotation is done during evaluation. It should be noted that for **Caltech-101**, a two action set comprising of angular rotation by 12.5° or by −12.5° gave best results. For **ImageNet** and ***Cats and Dogs* Dataset**, a three action set comprising of angular rotation by 90°, or by 180°, or *downward and rightward* diagonal translation by 15 pixels gave best results. The action sets gave best results, as compared to others including no-rotation action. It should be noted that rotation permutation during evaluation leads to better results than the conventional CNNs which do not use rotation. This can be due to alignment between test image and previous training

Downloaded from https://onlinelibrary.wiley.com/doi/ by Queen'S University, Wiley Online Library on [25/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 9.2 Classification accuracy of various approaches on ImageNet (**Second Classifier Used**: NN; **Metric Used**: Std. Deviation of Softmax scores, **Feature Map Size**: 1 × 1,024).

Approach	**Secondary NN used on top of layer**	**Image size**	**Accuracy**
ResNet50	#174: @(conv5_block3_out)	150 × 150 × 3	.8242
Proposed Approach using ResNet50	#174: @(conv5_block3_out)	150 × 150 × 3	**.8309**
Inception V3	#228: @(mixed7)	150 × 150 × 3	.8564
Proposed Approach using InceptionV3	#228: @(mixed7)	150 × 150 × 3	**.8644**

Table 9.3 Classification accuracy of various approaches on *Cats and Dogs* dataset (**Second Classifier Used**: NN; **Metric Used**: Std. Deviation of Softmax scores, **Feature Map Size:** 1 × 1,024).

Approach	**Secondary NN used on top of layer**	**Image size**	**Accuracy**
ResNet50	#174: @(conv5_block3_out)	224 × 224 × 3	.9780
Proposed Approach using ResNet50	#174: @(conv5_block3_out)	224 × 224 × 3	**.9860**
Inception V3	#228: @(mixed7)	150 × 150 × 3	.9440
Proposed Approach using InceptionV3	#228: @(mixed7)	150 × 150 × 3	**.9520**

Table 9.4 Classification accuracy of various approaches on Caltech-101 dataset (**Second Classifier Used**: Binary-SVM Ensemble; **Metric Used**: Std. Deviation of SVM prediction scores, **Feature Map Size:** 1 × 4,096).

Approach	**Secondary SVM used on top of layer**	**Image size**	**Accuracy**
AlexNet	-	227 × 227 × 3	.841
CNN-SVM Hybrid Approach [53]	#20: @(fc7)	227 × 227 × 3	.882
Proposed Approach using AlexNet	#20: @(fc7)	227 × 227 × 3	**.898**

image (due to rotation during evaluation) in the CNN. One advantage of this technique is that training need not be done extensively, thus saving resources like memory and time considerably for systems having both offline and online training. This statement is subject to speculation and will be more revealed and explained in future work.

Downloaded from https://onlinelibrary.wiley.com/doi/ by Queen'S University, Wiley Online Library on [25/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

As is observed from Table 9.2, larger image sizes lead to higher accuracies. However, the training times also increase.

As is observable from Tables 9.2 to 9.4, the proposed approach outperforms other approaches on all the datasets used. This technique can pave the way for a novel image recognition approach using tilt of vision in classifiers, etc. Also, the use of RL in computer vision is a first, to the best of available knowledge. This technique can pave the way for a whole new generation of computer vision classifiers based on RL. It should also be noted that in this study dimensional reduction [54] is not used. Also, that the proposed approach is instance-based. Thus, the processing time is more as compared to other techniques in this area. This is one of the limitations which are intended to be addressed in future work. In future, more work will be done on using larger datasets. Also, work would be done on making the proposed approach faster by using techniques like dimensional reduction, or using smaller feature maps. Also, work would be done on using the proposed approach on other interesting computer vision tasks like instance segmentation [55].

9.5 Conclusion

In this paper, a straightforward and efficient learning system is investigated which combines deep learning with reinforcement learning. The proposed technique is simpler than other contemporary techniques found elsewhere. This is for the reason that others use high number of states while as the proposed approach uses only two states. Thus, optimization is easy and the reward function is straightforward. Other approaches use visualization tasks like zoom and translation. A novel technique, i.e., rotation, has been used which is similar to tilt of visual field. Three databases have been used in the experimentation here. These are ImageNet, *Cats and Dogs* Dataset, and Caltech-101 Dataset. Benchmarking of the proposed classifier has been done. The proposed approach outperforms other approaches including ResNet50 and InceptionV3, on all the three datasets used.

References

1. Arulkumaran, K., Deisenroth, M.P., Brundage, M., Bharath, A.A., A Brief Survey of Deep Reinforcement Learning, *arXiv:170805866v2*, 28 Sep 2017.
2. Sutton, R.S. and Barto, A.G., *Reinforcement Learning: An Introduction*, The MIT Press, Cambridge, MA, United States, 2017.
3. Bernstein, A. and Burnaev, E., Reinforcement learning in computer vision, in: *Proc. SPIE 10696, Tenth International Conference on Machine Vision (ICMV 2017)*, 106961S, 13 April 2018, https://doi.org/10.1117/12.2309945
4. Botvinick, M., Ritter, S., Wang, J.X., Kurth-Nelson, Z., Blundell, C., Hassabis, D., Reinforcement learning, fast and slow. *Trends Cognit. Sci.*, 23, 5, 408–422, 2019.
5. Liu, Z., Wang, J., Gong, S., Lu, H., Tao, D., Deep reinforcement active learning for human-in-the-loop person re-identification, in: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 6122–6131, 2019.
6. Wirth, C., Akrour, R., Neumann, G., Fürnkranz, J., A Survey of Preference-Based Reinforcement Learning Methods. *J. Mach. Learn. Res.*, 18, 1–46, December, 2017.

Downloaded from https://onlinelibrary.wiley.com/doi/ by Queen'S University, Wiley Online Library on [25/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

7. Gärtner, E., Pirinen, A., Sminchisescu, C., Deep Reinforcement Learning for Active Human Pose Estimation, in: *AAAI*, pp. 10835–10844, 2020.
8. Wiering, M.A., van Hasselt, H., Pietersma, A., Schomaker, L., Reinforcement learning algorithms for solving classification problems, in: *2011 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)*, pp. 91–96, 2011, doi: 10.1109/ADPRL.2011.5967372.
9. Furuta, R., Inoue, N., Yamasaki, T., Fully convolutional network with multi-step reinforcement learning for image processing, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 3598–3605, 2019.
10. Lagoudakis, M.G. and Parr, R., Reinforcement Learning as Classification: Leveraging Modern Classifiers, in: *ICML*, 2003.
11. Toromanoff, M., Wirbel, E., Moutarde, F., Deep Reinforcement Learning for autonomous driving, in: *Workshop on "CARLA Autonomous Driving challenge", IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019.
12. Jiang, M., Hai, T., Pan, Z., Wang, H., Jia, Y., Deng, C., Multi-agent deep reinforcement learning for multi-object tracker. *IEEE Access*, 7, 32400–32407, 2019.
13. Hafiz, A.M. and Bhat, G.M., Deep Q-Network Based Multi-agent Reinforcement Learning with Binary Action Agents, *arXiv preprint arXiv: 200804109*, 2020.
14. Uzkent, B., Yeh, C., Ermon, S., Efficient object detection in large images using deep reinforcement learning, in: *The IEEE Winter Conference on Applications of Computer Vision*, pp. 1824–1833, 2020.
15. Zhang, D., Han, J., Zhao, L., Zhao, T., From Discriminant to Complete: Reinforcement Searching-Agent Learning for Weakly Supervised Object Detection, in: *IEEE Transactions on Neural Networks and Learning Systems,* vol. 31, no. 12, pp. 5549–5560, Dec. 2020, doi: 10.1109/TNNLS.2020.2969483.
16. König, J., Malberg, S., Martens, M., Niehaus, S., Krohn-Grimberghe, A., Ramaswamy, A., Multi-stage Reinforcement Learning for Object Detection, in: *Science and Information Conference*, Springer, pp. 178–191, 2019.
17. Pais, G.D., Dias, T.J., Nascimento, J.C., Miraldo, P., OmniDRL: Robust pedestrian detection using deep reinforcement learning on omnidirectional cameras, in: *2019 International Conference on Robotics and Automation (ICRA)*, IEEE, pp. 4782–4789, 2019.
18. Pirinen, A. and Sminchisescu, C., Deep Reinforcement Learning of Region Proposal Networks for Object Detection, in: *CVPR*, 2018.
19. Hierarchical Object Detection with Deep Reinforcement Learning, in: *NIPS*, 2016.
20. Mathe, S. and Pirinen, A., Reinforcement Learning for Visual Object Detection, in: *CVPR*, June 2016.
21. Watkins, C.J.C.H., *Learning from Delayed Rewards*. Ph.D. Thesis, University of Cambridge, Cambridge, United Kingdom, 1989.
22. Watkins, C.J.C.H. and Dayan, P., Q-learning. *Mach. Learn.*, 8, 3–4, 279–292, 1992.
23. Qiao, J., Wanga, G., Li, W., Chen, M., An adaptive deep Q-learning strategy for handwritten digit recognition. *Neural Networks*, 107, 61–71, 2018.
24. Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., Alsaadi, F.E., A survey of deep neural network architectures and their applications. *Neurocomputing*, 234, 11–26, 2017.
25. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., Gradient-based learning applied to document recognition. *Proc. IEEE*, 86, 11, 2278–2324, 1998.
26. König J., Malberg S., Martens M., Niehaus S., Krohn-Grimberghe A., Ramaswamy A., Multi-stage Reinforcement Learning for Object Detection, in: *Advances in Computer Vision. CVC 2019. Advances in Intelligent Systems and Computing*, K. Arai and S. Kapoor (eds), vol. 943, Springer, Cham, 2020. https://doi.org/10.1007/978-3-030-17795-9_13

Downloaded from https://onlinelibrary.wiley.com/doi/ by Queen'S University, Wiley Online Library on [25/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

27. Simonyan, K. and Zisserman, A., Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:14091556*, 2014.
28. Caicedo, J.C. and Lazebnik, S., Active Object Localization with Deep Reinforcement Learning, in: *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 2488–2496, 2015.
29. Hsu, R.C., Liu, C.-T., Chen, W.-Y., Hsieh, H.-I., Wang, H.-L., A Reinforcement Learning-Based Maximum Power Point Tracking Method for Photovoltaic Array. *Int. J. Photoenergy*, 12, 2015, 2015.
30. He, K., Zhang, X., Ren, S., Sun, J., Deep residual learning for image recognition, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
31. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., Rethinking the inception architecture for computer vision, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.
32. Krizhevsky, A., Sutskever, I., Hinton, G.E., Imagenet classification with deep convolutional neural networks, in: *Advances in neural information processing systems*, pp. 1097–1105, 2012.
33. LeCun, Y., Bengio, Y., Hinton, G., Deep learning. *nature*, 521, 7553, 436, 2015.
34. Schmidhuber, J., Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117, 2015.
35. Goodfellow, I., Bengio, Y., Courville, A., *Deep learning*, MIT press, Cambridge, MA, United States, 2016.
36. Shin, H.-C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M., Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imaging*, 35, 5, 1285–1298, 2016.
37. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K., Aggregated Residual Transformations for Deep Neural Networks, in: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 21-26 July 2017, pp. 5987–5995, 2017.
38. Cortes, C. and Vapnik, V., Support vector machine. *Mach. Learn.*, 20, 3, 273–297, 1995.
39. Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J., Scholkopf, B., Support vector machines. *IEEE Intell. Syst. Appl.*, 13, 4, 18–28, 1998.
40. Chang, C.C., Hsu, C.W., Lin, C.J., *Practical Guide to Support Vector Classification,* Department of Computer Science, National Taiwan University, Taipei 106, Taiwan, 2009.
41. He, K., Zhang, X., Ren, S., Sun, J., Identity mappings in deep residual networks, in: *European conference on computer vision*, Springer, pp. 630–645, 2016.
42. Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., Pietikäinen, M., Deep Learning for Generic Object Detection: A Survey. *Int. J. Comput. Vision*, 128, 2, 261–318, 2020.
43. Hafiz, A.M. and Bhat, G.M., A Survey of Deep Learning Techniques for Medical Diagnosis, in: *Singapore, Information and Communication Technology for Sustainable Development*, pp. 161–170, Springer, Singapore, 2020.
44. Hafiz, A.M. and Bhat, G.M., Digit Image Recognition Using an Ensemble of One-Versus-All Deep Network Classifiers, in: *Information and Communication Technology for Competitive Strategies (ICTCS 2020)*. Lecture Notes in Networks and Systems, vol. 190, Springer, Singapore, 2021. https://doi.org/10.1007/978-981-16-0882-7_38
45. Hafiz A.M. and Bhat G.M., Fast Training of Deep Networks with One-Class CNNs, in: *Modern Approaches in Machine Learning and Cognitive Science: A Walkthrough. Studies in Computational Intelligence*, V.K. Gunjan and J.M. Zurada (eds), vol. 956, Springer, Cham, 2021. https://doi.org/10.1007/978-3-030-68291-0_33
46. Hafiz, A.M. and Bhat, G.M., Deep Network Ensemble Learning applied to Image Classification using CNN Trees, *arXiv preprint arXiv:200800829*, 2020.

Downloaded from https://onlinelibrary.wiley.com/doi/ by Queen'S University, Wiley Online Library on [25/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

47. Deng, J., Dong, W., Socher, R., Li, L., Kai, L., Li, F.-F., ImageNet: A large-scale hierarchical image database, in: *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 20-25 June 2009, pp. 248–255, 2009.
48. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vision*, 115, 3, 211–252, 2015.
49. Parkhi, O.M., Vedaldi, A., Zisserman, A., Jawahar, C.V., Cats and dogs, in: *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 16-21 June 2012, pp. 3498–3505, 2012.
50. Fei-Fei, L., Fergus, R., Perona, P., Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories, in: *2004 conference on computer vision and pattern recognition workshop*, IEEE, pp. 178–178, 2004.
51. Li, F.-F., Andreetto, M., Ranzato, M.A., Caltech 101, 2004.
52. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Tensorflow: A system for large-scale machine learning, in: *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, pp. 265–283, 2016.
53. Tang, Y., Deep learning using linear support vector machines, *arXiv preprint arXiv:13060239*, 2013.
54. Maaten, L. J. P. v. d. and Hinton, G.E., Visualizing High-Dimensional Data Using t-SNE. *J. Mach. Learn. Res.*, 9, 2579–2605, 2008.
55. Hafiz, A.M. and Bhat, G.M., A survey on instance segmentation: state-of-the-art. *Int. J. Multimed. Inf. Retr.*, 9, 3, 171–189, 2020.

Downloaded from https://onlinelibrary.wiley.com/doi/ by Queen'S University, Wiley Online Library on [25/11/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License