Automated Detection of Hummingbirds in Images: A Deep Learning Approach

Sergio A. Serrano[1], Ricardo Benítez-Jimenez[1], Laura Nuñez-Rosas[2], Ma del Coro Arizmendi[3], Harold Greeney[4], Veronica Reyes-Meza[2], Eduardo Morales[1], and Hugo Jair Escalante[1](✉)

[1] Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Puebla, Mexico
{sserrano,ricardo.benitez,emorales,hugojair}@inaoep.mx

[2] Centro Tlaxcala de Biología de la Conducta, Universidad Autónoma de Tlaxcala, Tlaxcala, Mexico
lnunezr18@gmail.com, veronica.reyesm@uatx.mx

[3] Universidad Nacional Autónoma de México, Mexico City, Mexico
coro@unam.mx

[4] School of Natural Resources and the Environment, University of Arizona, Tucson, USA
greeney@email.arizona.edu

Abstract. The analysis of natural images has been the topic of research in uncountable articles in computer vision and pattern recognition (e.g., natural images has been used as benchmarks for object recognition and image retrieval). However, despite the research progress in such field, there is a gap in the analysis of certain type of natural images, for instance, those in the context of animal behavior. In fact, biologists perform the analysis of natural images manually without the aid of techniques that were supposedly developed for this purpose. In this context, this paper presents a study on automated methods for the analysis of natural images of hummingbirds with the goal to assist biologists in the study of animal behavior. The automated analysis of hummingbird behavior is challenging mainly because of (1) the speed at which these birds move and interact; (2) the unpredictability of their trajectories; and (3) its camouflage skills. We report a comparative study of two deep learning approaches for the detection of hummingbirds in their nest. Two variants of transfer learning from convolutional neural networks (CNNs) are evaluated in real imagery for hummingbird behavior analysis. Transfer learning is adopted because not enough images are available for training a CNN from scratch, besides, transfer learning is less time consuming. Experimental results are encouraging, as acceptable classification performance is achieved with CNN-based features. Interestingly, a pretrained CNN without fine tunning and a standard classifier performed better in the considered data set.

Keywords: Image classification · Convolutional neural network
Transfer learning · Animal behavior analysis · Hummingbird detection

J. F. Martínez-Trinidad et al. (Eds.): MCPR 2018, LNCS 10880, pp. 155–166, 2018.
https://doi.org/10.1007/978-3-319-92198-3_16

1 Introduction

The analysis of natural images, and more specifically, of images depicting animals, has served as motivation and justification for many landmark papers in computer vision and pattern recognition, see e.g. [1–6], contributing to the development and establishment of fields such as image categorization, image retrieval and even object recognition. For instance, reference benchmarks depicting animals include: ImageNet [7][1], Caltech-101[2], VOC[3], Mammal animals [8], SAIAPRTC12 [9], among others. However, it is remarkable that related fields needing this sort of methods have not been benefited that much from this progress. This is the case of animal behavior analysis, in which biologist must be carefully trained and later manually analyze large amounts of images and videos in order to draw conclusions about the behavioral patterns of living organisms.

Among birds, the nesting behavior is complicated to analyze. Specially hummingbirds are difficult to analyze during nesting period because of their high speed movements, cryptic colors, and the trouble of accessing to the places where they build their nests. The aim of this paper is to develop tools that facilitate the analysis of hummingbirds nesting behavior. Specifically, the study focuses on methods for detecting the presence of hummingbirds in nests recorded in videos. Knowing the time spent in nests is important for studying maternal care and investment, and making accurate descriptions about breeding strategies and the relationship between mother and offspring. Additionally, this is the first time that image analysis methods are applied for the analysis of hummingbirds behavior.

The problem of detecting objects in images has been studied since the beginning of computer vision. Thanks to the achievements in this field, and those in related fields like machine learning, nowadays there are available methods that show outstanding performance in a number of tasks focusing on image and video analysis (e.g., face verification [10]). In recent years, these methods are converging to a single modeling methodology: deep learning [11]. Convolutional neural networks have rapidly established as reference methods in the analysis of spatio-temporal data. However, the success of this model depends on a number of aspects, most importantly the amount of data available for training the models: large amounts of labeled data are required for learning the huge number of parameters (commonly on the order of hundreds of millions).

For the problem approached in this paper, labeled data is scarce and difficult to obtain. In this scenario, transfer learning is a strategy that aims at alleviating the scarcity of data. Transfer learning aims to tailor models learned for related tasks to solve the problem at hand. In this regard, several variants have been proposed. In this paper, two transfer learning strategies are adopted for learning representations directly from raw pixels. We perform a comparative study between both methods using real imagery collected by biologists. Experimental

[1] http://www.image-net.org/.
[2] http://www.vision.caltech.edu/Image_Datasets/Caltech101/.
[3] http://host.robots.ox.ac.uk/pascal/VOC/.

results reveal that a straightforward pretraining formulation, results in a better performance when compared to another popular and more promising strategy.

The contributions of this paper can be summarized as follows:

- A comparative study between two transfer learning methodologies for image classification of natural images.
- The application of the considered methodologies for the detection of hummingbirds in videos with the goal of supporting animal behavior research, where one of the evaluated methods obtained acceptable performance in a real data set.
- Experimental results evidencing the usefulness of deep learning methods for approaching real problems in animal behavior analysis.

The remainder of this paper is organized as follows. In the next section we present background information on the application domain and on convolutional neural networks. In Sect. 3 we describe in detail the methodology followed in the development of our research. In Sect. 4 we present the performed experiments and the results we obtained. Finally, in Sect. 5 the conclusions and future work are presented.

2 Background and Related Work

In this section, we provide some background information about the analysis of hummingbird behavior and about convolutional neural networks and transfer learning.

2.1 Hummingbird Behavior Analysis

The hummingbirds (Aves: Trochilidae) are endemic to the American continent, there are 330 different species. Their distribution range is wide from sea level to 4500 m above sea level. These small birds, just weight from 2 to 22 grams, are responsible of pollinating more than 1300 different plants [12], they are the only birds that can fly sideways and backwards, flapping up to 60 wingbeats per second, this is the reason why they have the highest in-flight metabolism of any bird species. They eat principally nectar but, during breeding season, they also eat arthropods and small insects [13,14]. Males are polygynous, therefore, after mating they usually search for other females to mate [15]. The females build small nests in hidden places and care the nestlings until they fledge [16].

Although reproduction is a very important period for hummingbird survival, little is known about it [13,17–22]. Generating information about breeding sites preferences, reproductive success and maternal investment for incubation and fledged is important for describing the natural history of these animals and promote their conservation. However, studying this period is quite complicated because it is difficult to find the nest, to get visual access and avoid to be detected by the bird, additionally it implies very long observation periods. Such difficulties could be overcome using breakthrough technology for visual analysis. This paper presents a study in such direction.

2.2 Convolutional Neural Networks

Convolutional Neural Networks (CNN) are a special type of artificial neural network that are characterized, among other things, by applying convolutional operations. Unlike standard neural networks, CNNs retrieve input data in the form of a n-dimensional tensor (typically 2-dimensional) to which a set of convolutional operators is applied, also called kernels. Each kernel operates throughout the whole tensor and generates as a result a smaller tensor called feature map. For instance in Fig. 1, six different kernels are applied to the input matrix which generate six feature maps which are sub-sampled to create even smaller feature maps. Each set of kernels that share the same input data constitute a convolutional layer, and each step where tensors are sub-sampled is called a pooling layer.

In addition to the convolutional and pooling layers, usually one might incorporate to a CNN non-linear activation functions in order to transform the feature maps, *e.g.* the ReLU function [23]. Similarly, it is a common practice to attach at the end of a CNN a fully connected neural network to represent the network's output in the form of a vector of real values, such as *softmax* [24]. The output is a vector constituted by n elements, each of them represents the probability that the input belongs to the i^{th} class.

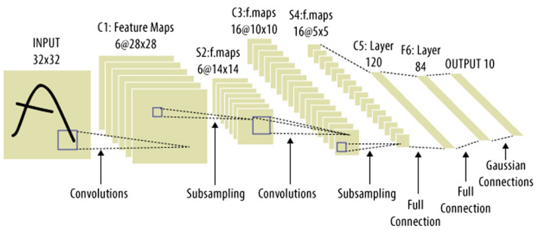

Fig. 1. General architecture of a CNN constituted by convolution, sampling and fully connected layers.

As previously mentioned, the performance of CNNs depends on the availability of a large enough data set from which CNNs' parameters can be adjusted. However, in many scenarios, including ours, labeled data is scarce and difficult to obtain. Hence, additional tricks or procedures must be performed to make CNNs work. In this context, a methodology that has been increasingly applied along with convolutional models is *transfer learning*, which was already being developed before CNN became a trend, for example, see [25–28]. In broad terms, transfer learning aims at approaching a target task A, by using as basis a model learned for task B that is often adjusted/modified to solve A. In the context of CNNs, transfer learning is a very popular solution for effectively using CNNs in tasks for which not enough labeled data is available.

3 Transfer Learning in CNNs for Detection of Hummingbirds in Images

In our study, labeled data is scarce: there are not enough labeled images depicting hummingbirds in nests, hence training a CNN from scratch is not an option. This particularity of the problem and the proved success of transfer learning in the context of CNNs inspired us for relying on transfer learning mechanisms for detecting hummingbirds in images with CNNs. In the remainder of this section we present fundamentals of CNNs, as well as on the two *transfer learning* methods compared in this paper: feature extraction with a pretrained CNN and fine tuning. Afterwards, we described the methodology followed in this research for the specific task we considered.

3.1 Transfer Learning Strategies

As previously mentioned, according to [27], transfer learning attempts to improve the learning of an objective function f_A that operates over a domain D_A, by using knowledge of another domain D_B, where $D_A \neq D_B$. What and how to transfer knowledge between different domains are the main research questions in its field, however, in the case of CNNs there are two transfer learning methods that have reported good results on image recognition tasks [29], features extraction along with a classifier and fine-tuning. Both strategies are considered in our study and described below.

Features Extraction from a Pretrained CNN

Usually, a CNN is constituted by several convolutional layers and their respective pooling layers, and at its end a fully connected neural network (see Fig. 1). This transfer learning approach uses the representation that a CNN, previously trained with millions of examples, generates for the instances retrieved in the input layer. The main idea behind this approach is to interpret the representations generated by the CNN as feature vectors, and use them to train a classifier, *e.g.* in [30] they show how effective this method can be. In Fig. 2 one may observe how the blue square encloses the CNN's layer from which the generated representation is taken and used for the classifier's training.

Fine-Tuning

We have previously mentioned the outstanding ability CNNs have in terms of large scale image classification. However, there are some tasks that do not require the recognition of a large amount of classes as in the ILSVRC challenge [6], and instead, few classes and a few number of training images are available. For this kind of smaller problems, the fine-tuning method is a good alternative. Fine-tuning is a form of transfer learning consisting of using a sub-set of parameters from a CNN that has already been trained over a general dataset, and compute the rest of the parameters by means of *back propagation*, training on a more specialized dataset, thus, the network will adjust itself to perform efficiently over the specialized task for which it was trained. The main advantage of using fine-tuning is that, unlike training a CNN from scratch, the training time decreases

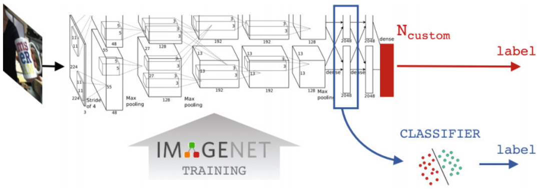

Fig. 2. The two transfer learning methods applied with CNNs: features extraction (blue) and fine-tuning (red). Figure from [31]. (Color figure online)

significantly and enables the usage of CNNs on problems with small datasets. In Fig. 2, the red square represents the section of layers that is re-trained using fine-tuning.

3.2 Detecting Hummingbirds with Transfered Learning from CNNs

The main goal of this paper is to determine the effectiveness of feature extraction and fine-tuning approaches along with an CNN, when applied to the task of detecting the presence of a hummingbird in an image. Our main motivation is to provide support tools for biologists that must manually analyze large amounts of videos. Implicitly, our aim is two-fold, (i) to prove the effectiveness of modern image classification algorithms in a real-world challenging domain and (ii) to verify whether either of the two transfer learning methods is significantly more effective than the other one on this specific problem. In the following we describe the considered data set and the adopted evaluation framework.

Data Collection and Division: The data set used in the development of our research was captured as follows. The videos were recorded in several places in Ecuador and Arizona, USA. The species recorded were *Aglaiocercus coelestis, Doryfera johannae, Heliangelus strophianus, Selaspherus platycercus* and *Topaza pyra*. First, the nest was located and then, videos were manually recorded with a camera for 45 min on average, making several recordings per day in the same nest.

The original data set is made up of 18 videos, this set was separated into 5 subsets, each formed by videos obtained from the same scene. One of these subsets was discarded due to its poor resolution and a lack of certainty when we manually attempted to label it. Figure 3 shows positive (hummingbird in nest) and negative (hummingbird not in nest) frames extracted form the four subsets retained; whereas Fig. 4 shows frames from the removed subset (for this subset it was not possible to determine the presence of the hummingbird for manual annotators).

From each subset of videos we extracted a set of frames using a sample rate that allowed us to gather at least 1000 positive and 1000 negative examples (that later were manually labeled), from which a frame centered on the hummingbird's nest was extracted and these frames were resized to 299 $\times$ 299 pixels. Finally, in order to label the adjusted frames, we applied to each of them the following

Fig. 3. The images in the upper row belong to the positive class and the ones in the lower row belong to the negative class. The columns, from left to right, correspond to the subsets **A**, **B**, **C** and **D**.

Fig. 4. Frames from the subset of videos that were omitted due to their poor resolution.

criterion: there were only two possible classes to be assigned, positive and negative classes. For a frame to be labeled as positive, a hummingbird should be in its nest. Otherwise, in order to label a frame as negative, no hummingbird nor any partial view of it should be captured within the image. As result, each set of frames was downsized due to those frames which did not satisfy any of the conditions previously described. To evaluate the classification performance of the models, the F1-Score was chosen.

Experimental Design: Being the main objective of this research to compare the performance of the features extraction method along with a classifier, and the fine-tuning method in the classification of images containing a hummingbird, we used *TensorFlow's* [32] implementation of the *Inception V3* CNN [33]. For the feature extraction approach, we opted for a SVM classifier and performed preliminary tests with several kernels, at the end we selected the linear kernel which was the one that reported the best results. On the other hand, with the fine-tuning approach, preliminary tests were carried out with different configurations in learning rate values and number of epochs to determine the appropriate parameters, the learning rate that reported better results was 0.01 and this was established as constant for the rest of the tests, while the optimal number of epochs ranged from 200 to 3000.

4 Experiments and Results

To compare both transfer learning methods, 14 tests have been defined, each one consisting of a combination of the four subsets of data, which we called **A**, **B**, **C** and **D**. In this way, the training and testing sets are not only disjoint, but also they come from different scenes making it more challenging. Each test is configured by the subsets used for training and those designated for evaluating the trained model, *i.e.* the test subsets. The number of training samples in every test was defined as 300, where of them 150 were positive and 150 of them were negative instances, randomly selected. Regarding the test samples, 1000 positive and 1000 negative samples were randomly selected for each test subset. Positive and negative examples of each of the subsets of images can be seen in Fig. 3. One of the main reasons of why we decided to gather only 1000 examples from each class is that in consecutive frames, even after sampling, the images have little noticeable differences, at least for the human eye.

The number of training samples for our models was determined experimentally after several preliminary tests, where the number of training samples was varied in each test. We observed that, with both approaches, for a number of examples greater than 400 the precision was high but with a low recall, and for less than 200 they had high recall but a very poor precision.

Table 1 shows the F1-Score obtained by the two evaluated methods. To compare the performance of the classifiers, the statistical Wilcoxon signed-rank test has been selected [34]. In order to apply this statistical test, we first defined the null hypothesis, h_0 = *In the task of classifying hummingbird images, the performance of the features extraction method with SVM and the Fine-tuning method are not significantly distinct.* Then, the absolute values of the F1-Score difference in each test were assigned a range, starting with the lowest value with range 1, up to the largest of the differences with range 14. These values are used by the Wilcoxon test to calculate a sum for each classifier. In this case, the sums of the ranks of each classifier were $R_{SVM} = 88$ y $R_{FineTuning} = 16$. Checking Wilcoxon's table of critical values for a confidence of $\alpha = 0.05$ and $N = 14$ tests, the difference between the classifiers is significant if the lesser of the sums is less than or equal to 21. This last condition is met by $R_{FineTuning} \leqslant 21$, therefore, we reject the null hypothesis h_0 and we can affirm that the features extraction approach with SVM is significantly better than fine-tuning to classify hummingbird images when there are few sets of images. In addition, the average of the F1-Score of SVM (≈0.6837) is superior that fine-tuning (≈0.6384) with a difference of ≈0.0453.

The best result obtained by the fine-tuning method is experiment 11, which was trained with sets **B**, **C**, **D** and tested in set **A**. However, the best result was obtained by the SVM approach in the experiment 14, where the training was carried out with the sets **A**, **B**, **C** and tested in **D**, in Fig. 5 examples of frames for this configuration are shown.

Table 1. Performance of both classifiers

ID	Training subsets	Testing subsets	# of test examples	F1-Score		F1-Score abs. diff.
				Feat. ext. + SVM	Fine tuning	
1	A	B, C, D	6,000	**0.6549**	0.6358	0.0191
2	B	A, C, D	6,000	**0.6913**	0.6744	0.0169
3	C	A, B, D	6,000	**0.7339**	0.6494	0.0845
4	D	A, B, C	6,000	0.6667	0.6667	0.0
5	A, B	C, D	4,000	**0.6667**	0.6665	0.0002
6	A, C	B, D	4,000	**0.6453**	0.5687	0.0766
7	A, D	B, C	4,000	**0.6026**	0.5235	0.0791
8	B, C	A, D	4,000	0.6715	**0.6921**	0.0206
9	B, D	A, C	4,000	0.6897	**0.7246**	0.0349
10	C, D	A, B	4,000	**0.6841**	0.6690	0.0151
11	B, C, D	A	2,000	**0.7502**	0.7281	0.0221
12	A, C, D	B	2,000	**0.6560**	0.3653	0.2907
13	A, B, D	C	2,000	**0.6682**	0.6675	0.0007
14	A, B, C	D	2,000	**0.7903**	0.7062	0.0841
Avg.				**0.6837**	0.6384	0.0453

Fig. 5. Sample positive instances from the experiment 14.

5 Conclusions and Future Work

We presented a methodology for detecting hummingbirds in images. The goal of the study is to provide biologists with support tools that can help them to analyze animal behavior and make new discoveries. The problem was approached as one of classification and a real data set was considered for experimentation. Since the number of distinctive available images is scare for the considered domain, we relied on transfer learning techniques. We presented a comparative analysis on two image classification methods based on transfer learning in CNNs: features extraction along with a SVM and *fine tuning*. Given the nature of our data, which was a scenario with few and high-dimensional data, we observed a better performance from the SVM approach and noticed that the *fine tuning* approach requires a more variated set of training examples in order to increase its precision when classifying new unseen instances. We think F1-score values obtained from the SVM approach are acceptable considering the low variance within the training example sets. Moreover, we performed other tests where the training and test example were extracted from the same sub-set, the f1-score values obtained from the *fine tuning* approach vary over the range of 0.8654 to 1, while the SVM approach obtained values ranging from 0.7263 up to 0.9823, which indicate the great precision CNNs can achieve when they are trained under scenarios that are not as restricted as the ones designed in our analysis. It is worth mentioning that we started training a standard CNN (*Alexnet* [35]) from scratch to have a reference performance. However, we confirmed this procedure was too computationally expensive when compared to transfer learning (1 epoch for *Alexnet* took 25 min, while 1000 epochs for the transfer learning configuration lasted 86 s). This is in addition to the expected low performance of the network. As future work, we plan to train CNNs from scratch with data augmentation mechanisms, also, we will explore the use of methods that provide localization, in addition to recognition, of objects in images.

References

1. Duygulu, P., Barnard, K., de Freitas, J.F.G., Forsyth, D.A.: Object recognition as machine translation: learning a lexicon for a fixed image vocabulary. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2353, pp. 97–112. Springer, Heidelberg (2002). https://doi.org/10.1007/3-540-47979-1_7
2. Barnard, K., Duygulu, P., Forsyth, D., de Freitas, N., Blei, D.M., Jordan, M.I.: Matching words and pictures. J. Mach. Learn. Res. **3**(Feb), 1107–1135 (2003)
3. 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: Proceedings of CVPRW, p. 178 (2004)
4. Griffin, G., Holub, G., Perona, P.: The caltech-256. Technical report. California Institute of Technology, Pasadena, California (2007)
5. Everingham, M., Zisserman, A., Williams, C.K.I., Van Gool, L.: The PASCAL Visual Object Classes Challenge 2006 (VOC2006) Results. http://www.pascal-network.org/challenges/VOC/voc2006/results.pdf
6. Russakovsky, O., Deng, J., Hao, S., 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. Vis. **115**(3), 211–252 (2015)
7. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: CVPR 2009 (2009)
8. Fink, M., Ullman, S.: From aardvark to zorro: a benchmark for mammal image classification. Int. J. Comput. Vis. **77**(1), 143–156 (2008)
9. Escalante, H.J., Hernández, C.A., Gonzalez, J.A., López-López, A., Montes, M., Morales, E.F., Sucar, L.E., Villaseñor, L., Grubinger, M.: The segmented and annotated IAPR TC-12 benchmark. Comput. Vis. Image Underst. **114**(4), 419–428 (2010)
10. Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: a unified embedding for face recognition and clustering. In: CVPR (2015)

11. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**, 436–444 (2015)
12. del Coro Arizmendi, M., Rodríguez-Flores, C.I.: How many plant species do hummingbirds visit? Ornitol. Neotrop. **23**, 71–75 (2012)
13. Elliott, A., del Hoyo, J., Sargatal, J.: Handbook of the Birds of the World, Volume 5, Barn-Owls to Hummingbirds, pp. 388–435. Lynx Edicions, Barcelona (1999)
14. Colwell, R.K.: Rensch's rule crosses the line: convergent allometry of sexual size dimorphism in hummingbirds and flower mites. Am. Nat. **156**(5), 495–510 (2000)
15. Bleiweiss, R.: Phylogeny, body mass, and genetic cponsequences of lek-mating behavior in hummingbirds (1998)
16. Johnsgard, P.A.: The Hummingbirds of North America. Smithsonian Institution, Washington (2016)
17. Vleck, C.M.: Hummingbird incubation: female attentiveness and egg temperature. Oecologia **51**(2), 199–205 (1981)
18. Baltosser, W.H.: Nesting success and productivity of hummingbirds in Southwestern New Mexico and Southeastern Arizona. Wilson Bull. **98**(3), 353–367 (1986)
19. Brown, B.T.: Nesting chronology, density and habitat use of black-chinned hummingbirds along the Colorado River Arizona. J. Field Ornithol. **63**(4), 393–400 (1992)
20. Greeney, H.F., Hough, E.R., Hamilton, C.E., Wethington, S.M.: Nestling growth and plumage development of the black-chinned hummingbird (Archilochus alexandri) in Southeastern Arizona. Huitzil. Revista Mexicana de Ornitología **9**(2), 35–42 (2008)
21. Greeney, H.F., Wethington, S.M.: Proximity to active accipiter nests reduces nest predation of black-chinned hummingbirds. Wilson J. Ornithol. **121**(4), 809–812 (2009)
22. Smith, D.M., Finch, D.M., Hawksworth, D.L.: Black-chinned hummingbird nest-site selection and nest survival in response to fuel reduction in a Southwestern Riparian forest. Condor **111**(4), 641–652 (2009)
23. Nair, V., Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning, ICML 2010, pp. 807–814 (2010)
24. Memisevic, R., Zach, C., Pollefeys, M., Hinton, G.E.: Gated softmax classification. In: Advances in Neural Information Processing Systems, pp. 1603–1611 (2010)
25. Dai, W., Yang, Q., Xue, G.-R., Yu, Y.: Boosting for transfer learning. In: Proceedings of the 24th International Conference on Machine Learning, pp. 193–200. ACM (2007)
26. Raina, R., Battle, A., Lee, H., Packer, B., Ng, A.Y.: Self-taught learning: transfer learning from unlabeled data. In: Proceedings of the 24th International Conference on Machine Learning, pp. 759–766. ACM (2007)
27. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**(10), 1345–1359 (2010)
28. Taylor, M.E., Stone, P.: Transfer learning for reinforcement learning domains: a survey. J. Mach. Learn. Res. **10**(Jul), 1633–1685 (2009)
29. Oquab, M., Bottou, L., Laptev, I., Sivic, J.: Learning and transferring mid-level image representations using convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1717–1724 (2014)
30. Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T.: DeCAF: a deep convolutional activation feature for generic visual recognition. In: International Conference on Machine Learning, pp. 647–655 (2014)

31. Pasquale, G., Ciliberto, C., Rosasco, L., Natale, L.: Object identification from few examples by improving the invariance of a deep convolutional neural network. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, pp. 4904–4911. IEEE (2016)
32. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., et al.: TensorFlow: large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467 (2016)
33. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of CVPR, pp. 2818–2826 (2016)
34. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. JMLR **7**(Jan), 1–30 (2006)
35. 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)