 CMC, vol.55, no.3, pp.435-446, 2018

Rare Bird Sparse Recognition via Part-Based Gist Feature Fusion and Regularized Intraclass Dictionary Learning

Jixin Liu[1, *], Ning Sun[1, 2], Xiaofei Li[1], Guang Han[1], Haigen Yang[1] and Quansen Sun[3]

Abstract: Rare bird has long been considered an important in the field of airport security, biological conservation, environmental monitoring, and so on. With the development and popularization of IOT-based video surveillance, all day and weather unattended bird monitoring becomes possible. However, the current mainstream bird recognition methods are mostly based on deep learning. These will be appropriate for big data applications, but the training sample size for rare bird is usually very short. Therefore, this paper presents a new sparse recognition model via improved part detection and our previous dictionary learning. There are two achievements in our work: (1) after the part localization with selective search, the gist feature of all bird image parts will be fused as data description; (2) the fused gist feature needs to be learned through our proposed intraclass dictionary learning with regularized K-singular value decomposition. According to above two innovations, the rare bird sparse recognition will be implemented by solving one l1-norm optimization. In the experiment with Caltech-UCSD Birds-200-2011 dataset, results show the proposed method can have better recognition performance than other SR methods for rare bird task with small sample size.

Keywords: Rare bird, sparse recognition, part detection, gist feature fusion, regularized intraclass dictionary learning.

1 Introduction

In the research field of bird monitoring and preservation, rare bird is undoubtedly one of the most valuable topic. However, it is also the most difficult to implement regulations. The reason is that, unlike human face or action, bird behavior is complicated and uncontrollable. In other words, the traditional manual observation can not be suitable for bird object, let alone use for rare bird. With the gradual popularization of the IOT (Internet of Things)-based video surveillance, all day and weather unattended bird monitoring becomes possible. Due to the above, new requirements for rare bird intelligent identification have been put forward.

[1] Engineering Research Center of Wideband Wireless Communication Technology, Ministry of Education, Nanjing University of Posts and Telecommunications, Nanjing 210003, PR China.

[2] Herbert and Florence Irving Medical Center, Columbia University, New York, NY 10032, United States of America.

[3] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, PR China.

[*] Corresponding Author: Jixin Liu. Email: liujixin@njupt.edu.cn.

As one application of pattern recognition, bird recognition has always been focused by researchers in the field of airport security, biological conservation, environmental monitoring, and so on. At present, for bird recognition, most achievements look at the aspect of audio data [Evangelista, Priolli, Silla Jr. et al. (2014); Ventura, Oliveira, Ganchev et al. (2015); Boulmaiz, Messadeg, Doghmane et al. (2016); Raghuram, Chavan, Belur et al. (2016); Chakraborty, Mukker, Rajan et al. (2017)]. But the study of bird image recognition might be relatively few [Li, Zhang and Yan (2014); Marini, Turatti, Britto et al. (2015); Karmaker, Schiffner, Strydom et al. (2017)]. In practical application, the audio recognition is not a ideal choice for bird monitoring. Because the real environment is easy to be influenced by noise interference. Therefore, the image data under video surveillance will be more suitable for bird recognition. For this purpose, it become necessary and urgent to carry out research in bird recognition for image or video data.

Image bird recognition is a kind of typical fine-grained recognition. For this kind of problem, CNN (convolutional neural network) [Han, Quan, Zhang et al. (2018)] is the most popular solution. Zhang et al. [Zhang, Donahue, Girshick et al. (2014)] propose a model for fine-grained categorization that overcomes these limitations by leveraging deep convolutional features computed on bottom-up region proposals. Lin et al. [Lin, Roychowdhury and Maji (2015)] present bilinear CNNs, an architecture that efficiently represents an image as a pooled outer product of two CNN features, that is effective at fine-grained recognition tasks. Wei et al. [Wei, Xie and Wu (2016)] propose a novel end-to-end Mask-CNN model without the fully connected layers for fine-grained recognition. Although these studies have yielded some results, the limitation of CNN is undeniable. That is due to the fact that deep learning with CNN will be more appropriate for big data applications. But the rare bird recognition task is usually very difficult to have a large enough training samples for CNN modeling. Hence, for rare bird recognition with small size, we need to select new ways to ensure high robustness under natural scene.

According to the above requirements, this paper present a new SR (sparse recognition) method for rare bird recognition. Fig. 1 shows the processing flow of this method. There are two innovation points in our work: Firstly, the local (such head as torso) and global (the whole object) image patches, based on part detection, will be fused as feature description under GIST [Oliva and Torralba (2001)] space. Secondly, by introducing regularized K-singular value decomposition, our previous work [Liu and Sun (2016)] will be improved as a new classifier in the solving performance of l_1 optimization. This paper will be organized as follows: Section 2 gives a brief introduction for SR method. In Section 3 the proposed SR for rare bird fine-grained recognition is detailed. Experimental results are analyzed in Sections 4 and Section 5 concludes the paper with a discussion.

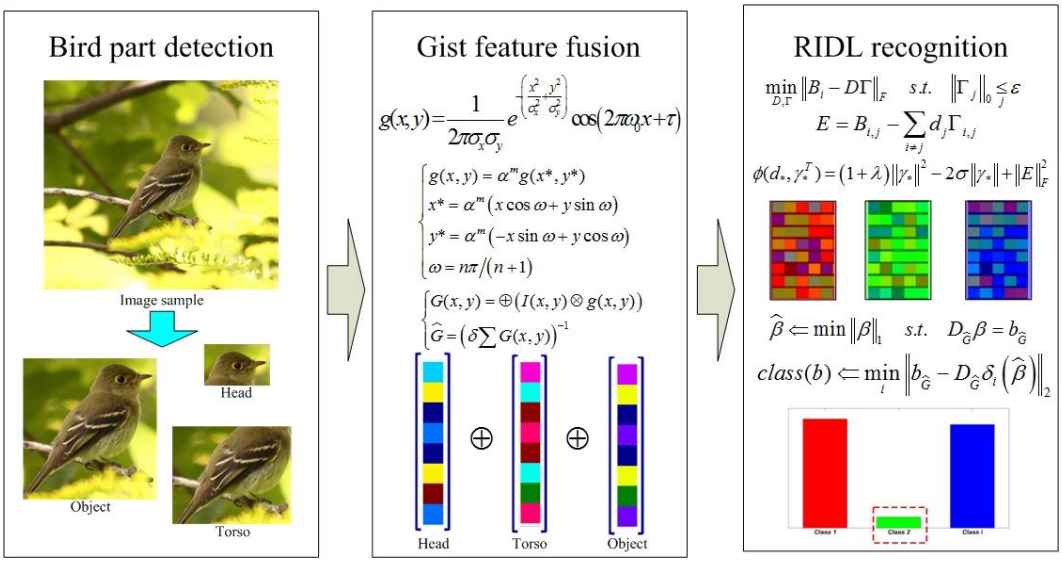


Figure 1: System flow of the proposed SR method

2 Sparse recognition and the related works

In the study of SR, there are two mainstream approaches at present. One classical method is SRC (sparse representation-based classification). It is derived from the theory of compressed sensing which is presented by Candes et al. [Candes and Tao (2006)] and Donoho [Donoho (2006)]. In this method, any test sample b can be sparsely measured through the global recognition matrix from the training sample set $B=\left[B_1,\ldots,B_i,\ldots,B_k\right]$. And the process will be implemented as

$$\gamma \Leftarrow \min\|\gamma\|_1 \quad s.t. \quad B\gamma = b \tag{1}$$

From this the recognition task can be accomplished by the following judgment

$$class(b) \Leftarrow \min_i \left\| b - B\delta_i\left(\gamma\right)\right\|_2 \tag{2}$$

Unfortunately, the performance of SRC will rely on some preprocessing (such as alignment [Ma, Luong, Philips et al. (2012)] or registration [Mohammadi, Fatemizadeh and Mahoor (2014)]).

Considering the limitation of SRC, another SR idea is presented. That is so-called DSR (dictionary-based sparse recognition) [Patel, Wu, Biswas et al. (2012); Zhang, Sun, Porikli et al. (2017)]. The key of DSR is based on one dictionary learning process as

$$\min_{D,\Gamma}\left\|B_i - D\Gamma\right\|_F \quad s.t. \quad \left\|\Gamma_j\right\|_0 \underset{j}{\leq} \varepsilon \tag{3}$$

the optimal solution Γ should meet the sparse level ε. For this problem, the most common algorithm is K-SVD (K-singular value decomposition) [Aharon, Elad and Bruckstein (2006)]. Simply speaking, the solving process can be realized by an alternate iteration between D and Γ. The first part, under a fixed initial dictionary D_0, is to acquire

the initial sparse representation with some optimization algorithms (such as orthogonal matching pursuit [Tropp (2004)] (OMP))

$$\Gamma_1 \Leftarrow \min_{\Gamma} \left\| B_i - D_0 \Gamma \right\|_F \quad s.t. \quad \left\| \Gamma_j \right\|_0 \leq_j \varepsilon \tag{4}$$

In the second part, under a fixed sparse representation, the dictionary will be replaced as

$$\left\| B_i - D\Gamma \right\|_F^2 = \left\| B_i - \sum_{j=1}^{K} d_j \gamma_S^j \right\|_F^2 = \left\| \left(B_i - \sum_{j \neq p} d_j \gamma_S^j \right) - d_p \gamma_S^p \right\|_F^2 = \left\| B_p^* - d_p \gamma_S^p \right\|_F^2 \tag{5}$$

$$B_p^* = U \Delta V' \quad \Rightarrow \quad d_p = U_1, \quad \gamma_S^p = \Delta_1^1 V'_1 \tag{6}$$

So the recognition task in DSR will be changed from Eq. (2) as

$$class(b) \Leftarrow \min_i \left\| b - D_{B_i} \left(D_{B_i}^T D_{B_i} \right)^{-1} D_{B_i}^T b \right\|_2 \tag{7}$$

Patel et al. [Patel, Wu, Biswas et al. (2012)] indicates that, DSR could be more robust than SRC without any preprocessing. But, it is easy to be local optimum because of a lack of global measurement.

In order to integrate the superiority of SRC and DSR, we propose the concept of intraclass dictionary learning (IDL) [Liu and Sun (2016)]. In this method, the global recognition matrix like SRC framework will be replaced with the IDL (not DSR) result from each class training sample set. Thus the SR under IDL can be improved from Eq. (1)

$$\beta \Leftarrow \min \left\| \beta \right\|_1 \quad s.t. \quad D_B \beta = \hat{b} \tag{8}$$

Here the global matrix D_B is generated by the IDL with K-SVD algorithm. Then the judgment also will become as

$$class(b) \Leftarrow \min_i \left\| \hat{b} - D_B \delta_i \left(\beta \right) \right\|_2 \tag{9}$$

By the experiment under some data sets, such as LFW [Huang, Ramesh, Berg et al. (2007)], Caltech101 [Li, Fergus and Perona (2007)] and ISR [Quattoni and Torralba (2001)], the proposed IDL shows the preferable recognition performance for image object in natural scene. Hence this paper will try to use it for rare bird fine-grained sparse recognition.

3 The proposed SR method for rare bird recognition

3.1 Challenges in rare bird image data

The major diversity of bird image is in the size, color and texture of bird parts. Take the popular Caltech-UCSD Birds-200-2011 [Wah, Branson, Welinder et al. (2011)] (CUB200-2011) as one example. In this database, each class has at least three orientations for bird head. There is no doubt that other parts will be more complicated. So the SRC for human face [Wright, Yang, Ganesh et al. (2009)] will be inadvisable. Because this model usually depends on the preprocessing. Besides, the small sample size for rare bird can easily affect the sparsity precondition in SRC framework. Thus it can be

seen, SR via dictionary learning should be taken seriously.

In the selection between DSR and IDL, we think that the latter is better. The reason is that, our previous work [Liu and Sun (2016)] shows that IDL has better robustness for object recognition under natural scene. When SR model can be determined, the new problem is how to realize feature description for dictionary learning.

From the current research achievement for bird recognition [Lin, Roychowdhury and Maji (2015); Wei, Xie and Wu (2016)], it is not hard to see that the part detection is one mainstream critical processing. So this paper, inspired by these studies, needs to select some part localization methods to generate suitable data representation. For this, we have some representative methods [Han, Quan, Zhang et al. (2018)] to leverage. Bourdev and Malik [Bourdev (2009)] propose a two-layer classification/regression model for detecting people and localizing body components; Felzenszwalb et al. [Felzenszwalb, Girshick, Mcallester et al. (2010)] described an object detection system based on mixtures of multiscale deformable part models; Uijlings et al. [Uijlings, Sande, Gevers et al. (2013)] introduce selective search which combines the strength of both an exhaustive search and segmentation; Long et al. [Long, Shelhamer and Darrell (2017)] show that a fully convolutional network trained end-to-end, pixels-to-pixels on semantic segmentation exceeds the state-of-the-art without further machinery.

Considering the lack of training samples for rare bird, Uijlings' selective search [Uijlings, Sande, Gevers et al. (2013)] will be very attractive. For CUB200-2011, this paper use selective search as part localization to extract the head, torso and object for each image sample. Fig. 2 shows the basic process.

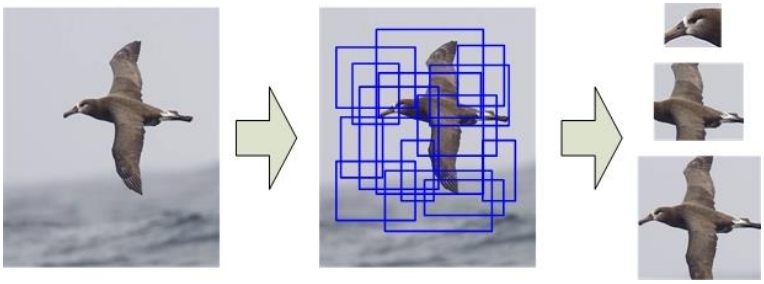


Figure 2: Selective search for bird image

3.2 Gist feature fusion based on part detection

When the main parts have been acquired, feature description becomes critical step. From the view of bird recognition [Zhang, Donahue, Girshick et al. (2014); Wei, Xie and Wu (2016)], HOG (histogram of oriented gradients) [Dalal and Triggs (2005)] is one common filter for feature representation. Although it might a good choice for deformable parts model (DPM) [Felzenszwalb, Girshick, Mcallester et al. (2010); Azizpour and Laptev (2012)], our experiments show that HOG can not make it work to its advantage under SR system. In contrast, gist descriptor seems more appropriate for this paper.

About the gist feature, the original goal of Oliva et al. [Oliva and Torralba (2001)] is to build a computational model of the recognition of real world scenes that bypasses the

segmentation and the processing of individual objects or regions. The core of gist is Gabor filter. Assume one image is $I(x, y)$, its 2D Gabor function can be as

$$g(x,y)=\frac{1}{2\pi\sigma_x\sigma_y}e^{-\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)}\cos\left(2\pi\omega_0 x+\tau\right) \tag{10}$$

On this basis, self-similarity Gabor can be structured as

$$\begin{cases} g(x,y)=\alpha^m g(x^*,y^*) \\ x^*=\alpha^m\left(x\cos\omega+y\sin\omega\right) \\ y^*=\alpha^m\left(-x\sin\omega+y\cos\omega\right) \\ \omega=n\pi/\left(n+1\right) \end{cases} \tag{11}$$

From this, the gist feature can be extracted as

$$\begin{cases} G(x,y)=\oplus\left(I(x,y)\otimes g(x,y)\right) \\ G=\left(\delta\sum G(x,y)\right)^{-1} \end{cases} \tag{12}$$

When the size of image grid unit is 4×4 under four scales and eight orientations, the gist feature dimensionality will be 512 (=4×4×4×8). In this paper, our fusion strategy is to cascading all parts' gist features as one data representation. Theoretically, the gist descriptor belongs to a kind of global feature. But the feature fusion in our work is derived from various local patches. This makes our gist feature having both local and global superiority in image description. The subsequent experiment will prove this point.

3.3 Regularized IDL for rare bird sparse recognition

As the comparison in Section 2, IDL could be an appropriate choice for rare bird recognition. If the gist feature with all parts has been generated, the SR classier can be set as

$$\beta \Leftarrow \min\|\beta\|_1 \quad s.t. \quad D_G\beta=b_G \tag{13}$$

Here D_G is from the IDL processing. And the common algorithm for IDL undoubtedly is K-SVD.

From Eq. (4) to Eq. (6), we can see the basic process for K-SVD. However, in each iteration, it implies that the update of dictionary and sparse representation would be not at the same time. So it might be likely to produce singular point. For solving this problem, Wei et al. [Wei, Xu and Wang (2012)] try to change the objective function as

$$f_\lambda(D,\Gamma)=\|B_i-D\Gamma\|_F^2+\lambda\|\Gamma\|_F^2 \tag{14}$$

Although this improvement could prevent the singular point, it is a pity that its performance will decline dramatically when the size of training sample is not enough.

The latest solution is presented by Dumitrescu et al. [Dumitrescu and Irofti (2017)]. In their so-called regularized K-SVD (RK-SVD), the signal error during sparse

representation update will be changed as

$$E = B_{i,j} - \sum_{i \neq j} d_j \Gamma_{i,j} \tag{15}$$

Then the optimal measurement will be inferred as

$$\phi(d_*, \gamma_*^T) = (1+\lambda)\|\gamma_*\|^2 - 2\sigma\|\gamma_*\| + \|E\|_F^2 \tag{16}$$

Based on this RK-SVD, our IDL could be improved as Tab. 1. And we name it RIDL (regularized intraclass dictionary learning). Finally, the SR result for rare bird will be judged from the following criterion

$$class(b) \Leftarrow \min_i \left\| b_G - D_G \delta_i(\beta) \right\|_2 \tag{17}$$

Table 1: Algorithm of the proposed RIDL

The pseudo-code for IDL based on RK-SVD.

1. Input

Initial dictionary *D*, *i*-th class training sample set under gist feature space G_{B_i} and iteration times *K*.

2. IDL with RK-SVD

for *i*=1 to *c*

Start dictionary learning:

for *k*=1 to *K*

Sparse representation: *D* is fixed and Γ is solved by OMP;

Dictionary update:

for *j*=1 to *n*

Use Γ to find d_j and acquire (σ, u, v) by SVD;

Set $d_j = u$ and update the sparse representation as $\Gamma_{i,j} = \sigma v / (1+u)$;

end for

end for

i-th class intraclass dictionary: $D_{G_{B_i}} = D$.

end for

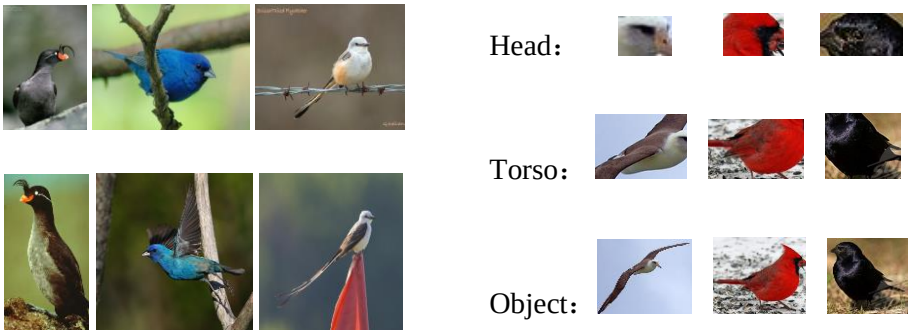
3. Output

Generate the global recognition matrix $D_G = \left\{ D_{G_{B_i}} \right\}_i$.

4 Experiment and analysis

In this section, the rare bird training sample set comes from the CUB200-2011. This

dataset has 200 bird classes with about 60 images in each class. In China, rare birds under the key state protection list are 58 species, and 16 species in it are endangered. Unfortunately, there is no complete correspondence category in CUB200-2011 for these birds. For this reason, we can only use some similar family or genus in CUB200-2011 instead. Our experiment will select 11 classes (such as Parakeet Auklet, Belted Kingfisher, White Pelican, and so on) from CUB200-2011 with 30 image samples in each class randomly. About each sample, three parts (head, torso and the whole object) will be segmented by selective search. And the feature fusion strategy is cascade mode. Fig. 3 shows some samples in CUB200-2011 and some results with part detection.



(a) (b)

Figure 3: (a) Some samples and (b) results with part detection.

4.1 Experiment 1

For comparison of gist and other feature descriptors, RGB color histogram and HOG will be studied. Considering the possible way of feature fusion, five compound modes should be set including head, torso, object, head+torso, and head+torso+object. Recognition system will run 300 times with 25 training samples in each class, and recognition rate would be counted as evaluation index. Fig. 4 is the result for this experiment.

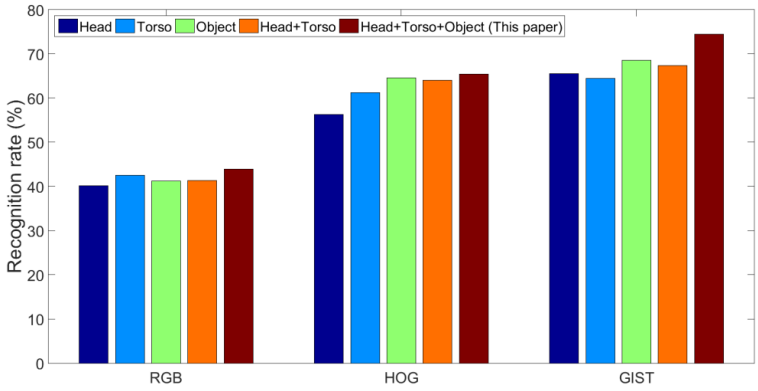


Figure 4: Five part compound modes under different feature space

From the Fig. 4, it can be seen that: (1) In the three representative feature descriptors, the gist fusion has better recognition rate than other two method; (2) Through the

comparison in the five fusion patterns, the head+torso+object shows the best application effect. These results means that the proposed gist feature fusion based on part detection could mix the global description from gist feature and the local segmentation from selective search.

4.2 Experiment 2

Because one innovation in this paper is to use the RK-SVD to improve our previous IDL as a new SR classier. For comparing the application effect of the proposed RIDL, three typical SR methods (SRC, DSR and IDL) will be tested. And another aim in this experiment is to study how the small sample size problem of rare bird influences the SR modes. So the recognition rate for these four SR modes will be contrasted under five training sample sizes (5, 10, 15, 20 and 25). Fig. 5 shows the result of above experiment.

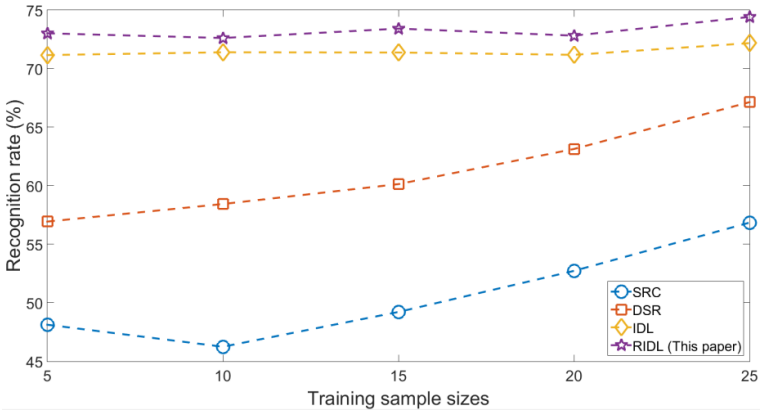


Figure 5: SR methods with five training sample sizes

Fig. 5 shows two aspects of this experiment: (1) From the view of SR methods, the recognition rate of SRC is far less than other dictionary learning approaches; (2) With the change of training sample size, our IDL and RIDL will be more robust and stable than other classic methods. So it is not hard to see that the proposed RIDL could be more appropriate for rare bird recognition with small sample size.

5 Conclusion

For rare bird recognition, this paper proposes a new SR method based on gist feature fusion and regularized IDL. In our SR system, there are two key steps. One is that three parts (head, torso and object) of each bird image sample will be extracted through selective search before the feature fusion is implemented. Another is the proposed RIDL which can be considered as the improvement of our previous IDL via RK-SVD. The experimental results, under CUB200-2011, show the feasibility of our work for rare bird intelligence monitoring.

Acknowledgement: This work was supported by the China National Natural Science Funds (Grant No. 61401220 and No. 61471206) and the Scientific Research Foundation of Nanjing University of Posts and Telecommunications (Grant No. NY218066).

References

Aharon, M.; Elad, M.; Bruckstein, A. (2006): The K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311-4322.

Azizpour, H.; Laptev, I. (2012): Object detection using strongly-supervised deformable part models. *European Conference on Computer Vision*, vol. 7572, pp. 836-849.

Boulmaiz, A.; Messadeg, D.; Doghmane, N.; Taleb-Ahmed, A. (2016): Robust acoustic bird recognition for habitat monitoring with wireless sensor networks. *International Journal of Speech Technology*, vol. 19, no. 3, pp. 1-15.

Bourdev, L.; Malik, J. (2009): Poselets: Body part detectors trained using 3D human pose annotations. *IEEE International Conference on Computer Vision*, vol. 30, pp. 1365-1372.

Candes, E. J.; Romberg, J.; Tao, T. (2006): Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489-509.

Candes, E. J.; Tao, T. (2006): Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406-5425.

Chakraborty, D.; Mukker, P.; Rajan, P.; Dileep, A. D. (2017): Bird call identification using dynamic kernel based support vector machines and deep neural networks. *IEEE International Conference on Machine Learning and Applications*, pp. 280-285.

Dalal, N.; Triggs, B. (2005): Histograms of oriented gradients for human detection. *IEEE Computer Vision and Pattern Recognition 2005*, vol. 1, pp. 886-893.

Donoho, D. L. (2006): Compressed sensing. *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289-1306.

Dumitrescu, B.; Irofti, P. (2017): Regularized K-SVD. *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 309-313.

Evangelista, T. L. F.; Priolli, T. M.; Silla Jr, C. N.; Angelico, B. A.; Kaestner, C. A. A. (2014): Automatic segmentation of audio signals for bird species identification. *IEEE International Symposium on Multimedia*, vol. 21, pp. 223-228.

Felzenszwalb, P. F.; Girshick, R. B.; Mcallester, D.; Ramanan, D. (2010): Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 32, no. 9, pp. 1627-1645.

Han, J.; Quan, R.; Zhang, D.; Nie, F. (2018): Robust object co-segmentation using background prior. *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 1639-1651.

Han, J.; Zhang, D.; Cheng, G.; Liu, N.; Xu, D. (2018): Advanced deep-learning techniques for salient and category-specific object detection: A Survey. *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 84-100.

Huang, G.; Ramesh, M.; Berg, T.; Learned-Miller, E. (2007): Labeled faces in the wild. *Technical Report TR 07-49*. University of Massachusetts, Amherst, USA.

Karmaker, D.; Schiffner, I.; Strydom, R.; Srinivasan, M. V. (2017): WHoG: A weighted HoG-based scheme for the detection of birds and identification of their poses in natural environments. *International Conference on Control, Automation, Robotics and Vision*, pp. 1-7.

Li, F. F.; Fergus, R.; Perona, P. (2007): Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer Vision and Image Understanding*, vol. 106, no. 1, pp. 178.

Li, J.; Zhang, L.; Yan, B. (2014): Research and application of bird species identification algorithm based on image features. *International Symposium on Computer, Consumer and Control*, pp. 139-142.

Lin, T.; Roychowdhury, A.; Maji, S. (2015): Bilinear CNN models for fine-grained visual recognition. *IEEE International Conference on Computer Vision*, pp. 1449-1457.

Liu, J.; Sun, Q. (2016): Sparse recognition via intra-class dictionary learning using visual saliency information. *Neurocomputing*, vol. 196, pp. 70-81.

Long, J.; Shelhamer, E.; Darrell, T. (2017): Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 39, no. 4, pp. 640-651.

Ma, X.; Luong, H. Q.; Philips, W.; Song, H.; Cui, H. (2012): Sparse representation and position prior based face hallucination upon classified over-complete dictionaries. *Signal Processing*, vol. 92, no. 9, pp. 2066-2074.

Marini, A.; Turatti, A. J.; Britto, A. S.; Koerich, A. L. (2015): Visual and acoustic identification of bird species. *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2309-2313.

Mohammadi, M. R.; Fatemizadeh, E.; Mahoor, M. H. (2014): PCA-based dictionary building for accurate facial expression recognition via sparse representation. *Journal of Visual Communication & Image Representation*, vol. 25, no. 5, pp. 1082-1092.

Oliva, A.; Torralba, A. (2001): Modeling the shape of the scene: a holistic representation of the spatial envelope. *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145-175.

Patel, V. M.; Wu, T.; Biswas, S.; Phillips, P. J.; Chellappa, R. (2012): Dictionary-based face recognition under variable lighting and pose. *IEEE Transactions on Information Forensics & Security*, vol. 7, no. 3, pp. 954-965.

Quattoni, A.; Torralba, A. (2001): Recognizing indoor scenes. *IEEE Computer Vision and Pattern Recognition 2009*, pp. 413-420.

Raghuram, M. A.; Chavan, N. R.; Belur, R.; Koolagudi, S. G. (2016): Bird classification based on their sound patterns. *International Journal of Speech Technology*, vol. 19, no. 4, pp. 791-804.

Tropp, J. (2004): Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, vol. 50, pp. 2231-2242.

Uijlings, J. R. R.; Sande, K. E. A. V. D.; Gevers, T.; Smeulders, A. W. M. (2013):

Selective search for object recognition. *International Journal of Computer Vision*, vol. 104, no. 2, pp. 154-171.

Ventura, T. M.; Oliveira, A. G. D.; Ganchev, T. D.; Figueiredo, J. M. D.; Jahn, O. et al. (2015): Audio parameterization with robust frame selection for improved bird identification. *Expert Systems with Applications an International Journal*, vol. 42, no. 22, pp. 8463-8471.

Wah, C.; Branson, S.; Welinder, P.; Perona, P.; Belongie, S. (2011): The Caltech-UCSD birds-200-2011 dataset. *Technical Report CNS-TR-2011-001*.

Wei, D.; Xu, T.; Wang, W. (2012): Simultaneous codeword optimization (SimCO) for dictionary update and learning. *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6340-6353.

Wei, X.; Xie, C. W.; Wu, J. (2016): Mask-CNN: Localizing parts and selecting descriptors for fine-grained image recognition. *29th Conference on Neural Information Processing Systems*, pp. 1-9.

Wright, J.; Yang, A. Y.; Ganesh, A.; Sastry, S. S.; Ma, Y. (2009): Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 31, no. 2, pp. 210-227.

Zhang, N.; Donahue, J.; Girshick, R.; Darrell, T. (2014): Part-based R-CNNs for fine-grained category detection. *European Conference on Computer Vision 2014*, vol. 8689, pp. 834-849.

Zhang, G.; Sun, H.; Porikli, F.; Liu, Y.; Sun, Q. (2017): Optimal couple projections for domain adaptive sparse representation-based classification. *IEEE Transactions on Image Processing*, vol. 26, no. 12, pp. 5922-5935.