DETECTING KANGAROOS IN THE WILD: THE FIRST STEP TOWARDS AUTOMATED ANIMAL SURVEILLANCE

Teng Zhang[⋆], *Arnold Wiliem*[⋆], *Graham Hemson*[†] *and Brian C. Lovell*[⋆]

[⋆]The University of Queensland, Australia
[†]Queensland Parks and Wildlife Service, Australia

ABSTRACT

Recent studies in computer vision have provided new solutions to real-world problems. In this paper, we focus on using computer vision methods to assist in the study of kangaroos in the wild. In order to investigate the feasibility, we built a kangaroo image dataset from collected data from several national parks across the State of Queensland. To achieve reasonable detection accuracy, we explored a multi-pose approach and proposed a framework based on the state-of-the-art Deformable Part Model (DPM). Experiments show that the proposed framework outperformed the state-of-the-art methods on the proposed dataset. Also, the proposed vision tools are able to help our field biologists in studying kangaroo related problems such as population tracking for activity analysis.

***Index Terms*—** Object detection, animal, kangaroo, population tracking, DPM

1. INTRODUCTION

Monitoring animal populations and activities has significance for biology and ecology [1]. In uncontrolled field work environments like desert, forest or sea, it is desirable to develop computer vision tools to perform the surveillance tasks automatically instead of performing manual field observation. These automated tools could help biologists to perform much more efficient and cost effective field studies.

Performing vision tasks on wild animals can be very challenging due to the complex background, varying illumination, occlusion and the multiple shapes and poses of the animals. In addition, the task is usually an open set problem. To illustrate this, we present an example for the kangaroo detection task in Fig. 1. The figure shows both a successful detection and a false alarm. Although we confine ourselves to kangaroos, there are also some other animals such as emus, wild pigs, cats and dingos that are captured by the cameras. Thus, it is nearly impossible to include all possible animals that are not of interest during the training phase.

There have been several works that are related to animal images [2][3][4]. For instance, in [2], attribute based detection are studied for unseen animal detection and the authors also propose an animal dataset with attribute labels. Fine-grained classification for cats and dogs are explored in [3]. They use DPM to detect animal face and bag-of-words to describe the animal pattern. In [4], shape and texture features are extracted for cat's head detection. Unfortunately, most prior works were not primarily focussed on field work settings. The images are often collected from internet; makes these completely orthogonal to the focus of our work. In contrast to these works, we collected our kangaroo images from cameras mounted in the wild. Because of this, the images in our data is often in low resolution and occluded due to the varying poses.

(a) (b)

Fig. 1. Kangaroo detection with bounding-boxes. The left one is a correctly detection and the right one is a false alarm.

Perhaps the most closely related works are works proposed in [5, 6]. In [5], authors analyse the data collected in the wild of Kenya for animal identification. They use the standard SIFT features to perform the task. In [6], the authors work on animal recognition where the data are collected from the Mojave desert. They also use the simple SIFT-based method to assist the field biologist for an animal population study. However, the animals they studied are not extremely deformable like the kangaroo; thus, this task is significantly easier.

Contributions Our main contributions can be listed as follows: (1) we describe and propose a Kangaroo detection dataset which will be useful for practitioners in the field to develop their algorithms; (2) we propose an extension of DPM to address problems in Kangaroo detection; (3) we demonstrate that our proposed tools could be used to assist in the study of Kangaroos.

We continue this paper as follows. In Section 2, we propose the kangaroo dataset and show some samples. In Section 3, we describe the kangaroo detection problem and the proposed method in details. Experiments and analysis are given in Section 4. The conclusion and possible future directions are summarized in Section 5.

2. KANGAROO DATASET

Our data were collected in several national parks across Queensland State during 2013. In order to monitor wild animals for a long time, we used the RECONYX camera system as shown in Fig. 2. This camera works in both day and night conditions and has a long battery life. The camera is camouflaged to match tree branches to avoid aggression from the animals. We set up the cameras in front of the feeding locations to get a clear view of the animals. The cameras work on a low frame-rate to avoid high storage consumption. The collected videos were pre-processed and nearly 3,000 frames were extracted from each location. We present some examples in Fig. 3.

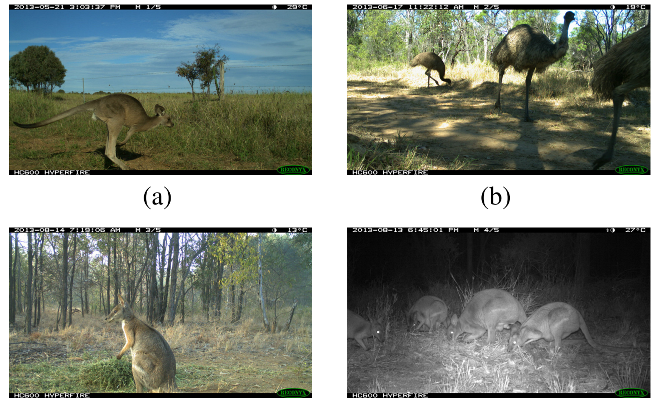
Fig. 2. RECONYX Camera.

We discarded the frames that contain only background and blurry images. After these frames were discarded, we extracted 1,900 cropped images to build the kangaroo dataset. The dataset contained 250 positive samples and 450 negative samples for training. The test set comprised 600 positive samples and 600 negative samples. We opted to use the True Positive Rate (TPR), False Positive Rate (FPR) and Average Precision (AP) as the performance metrics. These are calculated from the True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) as follows. TPR $= \frac{\text{TP}}{\text{TP+FN}}$, FPR $= \frac{\text{FP}}{\text{FP+TN}}$, and AP $= \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}}$. All positive training samples were labelled with a bounding box in the XML format. The dataset will be available for download at `http://www.itee.uq.edu.au/sas/datasets`.

3. PROPOSED VISION TOOLS

Unlike most object detection tasks, kangaroo detection possesses markedly more difficult challenges as kangaroos could have multiple poses and extremely deformed body (e.g. a Kangaroo could significantly twist their body). Significant variations to illuminant and background also exist due to the uncontrolled natural environment. Fig. 4 provides some examples of kangaroos in different poses. Significant appearance differences in various poses present difficulties in the detection task.

Despite the challenges, from our empirical observation, one of the state-of-the-art methods called, Deformable Part



(a) (b) (c) (d)

Fig. 3. Sample frames from collected frames. (a): Kangaroo jumping, (b): Emus walking, (c): Kangaroo standing, (d): Kangaroos eating in the night.

(a) (b) (c)

Fig. 4. Different poses of kangaroo. (a): Kangaroo showing its back to the fixed camera, (b): Kangaroo showing its frontal body, (c):Kangaroo showing its left part of the body

Model based object detection method, can achieve quite reasonable detection accuracy. This could be due to the fact that the core idea of DPM is to capture the variation in an object's appearance. The model is defined by a coarse root filter, several higher resolution part filters and a spatial model for the location of each part relative to the root [7]. The spatial model captures the varying appearance of objects in the same class. More precisely, Let $\boldsymbol{I} \in \mathbb{R}^{k\times l}$ be an image region extracted in sliding window manner; $\boldsymbol{H} \in \mathcal{R}^{k\times l}$ be the HOG [8] feature map; $\boldsymbol{p}$ denotes a location (x, y) in $\boldsymbol{H}$ and $\boldsymbol{w}_0 \in \mathbb{R}^{w\times h}$ be a $w \times h$ filter. Let $\phi(\boldsymbol{p}, \boldsymbol{H}, w, h)$ denote the vector obtained by concatenating the feature vectors in the $w \times h$ sub-window of $\boldsymbol{H}$ with top-left corner at $\boldsymbol{p}$ in row-major order. Then the filter score of $\boldsymbol{w}_0$ at location $\boldsymbol{p}$ is $\boldsymbol{w}_0^\top \phi(\boldsymbol{p}, \boldsymbol{H}, w, h)$. Similarly, the i-th part is defined by filter $\boldsymbol{w}_i$ and $\boldsymbol{d}_i \in \mathbb{R}^4$ is a four dimensional vector specifying coefficients of a quadratic function defining a deformation cost for each possible placement of the part. For clarity, we will shorten the expression $\phi(\boldsymbol{p}, \boldsymbol{H}, w, h)$ to $\phi(\boldsymbol{p}, \boldsymbol{H})$. An object hypothesis can be represented by $\{\boldsymbol{p}_0, \boldsymbol{p}_1, \cdots, \boldsymbol{p}_n\}$ where $\boldsymbol{p}_0$ is the root location and $\boldsymbol{p}_i$ is the location of the i-th part. The sliding image window containing the hypothesis is $\boldsymbol{I}$. Finally, the score of a hypothesis in $\boldsymbol{H}$ is given

$$\Phi_{dpm}(\boldsymbol{I}, \{\boldsymbol{p}_i\}_{i=0}^{i=n}) = \boldsymbol{w}_0^\top \phi(\boldsymbol{p}_0, \boldsymbol{H}) + \sum_{i=1}^{n} \boldsymbol{w}_i^\top \phi(\boldsymbol{p}_i, \boldsymbol{H}) - \boldsymbol{d}_i^\top \lambda_d(\boldsymbol{p}_i, \boldsymbol{p}_0) + b \quad (1)$$

where $\lambda_d(\cdot)$ is the deformation feature defined by the derivations of the pixel distance between $\boldsymbol{p_i} = (x_i, y_i)$ and $\boldsymbol{p_0} = (x_0, y_0)$. Specifically,

$$\lambda_d(\boldsymbol{p}_i, \boldsymbol{p}_0) = (dx_i, dy_i, dx_i^2, dy_i^2) \quad (2)$$

where $(dx_i, dy_i) = (x_i, y_i) - (2(x_0, y_0) + \boldsymbol{v}_i)$ gives the displacement of the i-th part relative to its anchor position $\boldsymbol{v}_i$. The vector $\boldsymbol{v}_i$ is a two-dimensional vector specifying an anchor position for part i relative to the root position. The bias term b is introduced to make the scores of multiple models comparable. When it comes to detection, the location of parts are inferred by maximizing the part appearance score $\boldsymbol{w}_i^\top \phi(\cdot)$ minus the deformation cost in Eqn. 3. Interested readers are referred to [7] for a full treatment of DPM.

$$\boldsymbol{p}_i = arg \max_{\boldsymbol{p}_i} \boldsymbol{w}_i^\top \phi(\boldsymbol{p}_i, \boldsymbol{H}) - \boldsymbol{d}_i^\top \lambda_d(\boldsymbol{p}_i, \boldsymbol{p}_0) \quad (3)$$

where $\boldsymbol{p}_i$ traverses possible locations of the part.

DPM is a classifier of good specificity. This is due to the fact that DPM first detects the body parts in an image before making an inference in an image region. Unfortunately, kangaroo body part visibility varies depending on its pose. In other words, the DPM may fail to detect on some poses where important body parts are occluded. From our empirical observations, we found that although DPM fails to detect kangaroos due to occlusion of important body parts, its hypothesis score defined in Eqn. 1 is still significantly higher than the true negatives (i.e. the background). We shall call these as weak negatives. The set of images belong to weak negatives is denoted $\mathcal{S}_{wn}$.

In our work, we focus on the weak negatives to improve DPM detection performance. To that end, we propose to use pose specific SVM models in addition to the original DPM model. This formulation allows the system to markedly improve the performance while maintaining low computational complexity.

Let $\mathcal{G}_j = \{\boldsymbol{x_i}\}_{i=1}^{i=m_j}$ be the positive training samples for the j-th pose, where m_j denotes the number of positive training samples and $\boldsymbol{x}_i$ denotes the HOG histogram representation of the positive exemplar. The pose specific SVM can be trained using the general max margin training via.

$$min \frac{1}{2} \|\boldsymbol{\gamma}_j\|^2 \;\; s.t.,\; y_i(\boldsymbol{x_i} + \beta_j) \geq 1, i = 1, .., m^j \quad (4)$$

where $\boldsymbol{\gamma}_j$ and β_j are the parameters for pose model j.

For each pose j, we use the pose information manually labelled from the training set to create the positive training samples $\mathcal{G}_j$. We then train the SVM model by using $\mathcal{G}_j$ as positive exemplars as well as all the negative exemplars. In our work we define eight different poses: front, rear, left, right, front-left, front-right, rear-left and rear-right. As for DPM, we use all positive exemplars from all poses together with the negative exemplars.

During the detection process, we combine the score from DPM and the multi-pose SVM as follows.

$$\Psi(\boldsymbol{I}) = \begin{cases} \Phi_{pose}(\boldsymbol{I}), & \text{if } \boldsymbol{I} \in \mathcal{S}_{wn} \\ \Phi_{dpm}(\boldsymbol{I}), & otherwise \end{cases} \quad (5)$$

where $\Psi(\cdot)$ is our final detector score; $\mathcal{S}_{wn}$ is the weak negative set; $\Phi_{dpm}(\cdot)$ is the DPM detection score defined in Eqn. 1 and $\Phi_{pose}(\cdot)$ is the multi-pose SVM detection score defined via.

$$\Phi_{pose}(\boldsymbol{I}) = \max \left(\{\varphi_{pose}^j(\boldsymbol{I})\}_{j=1}^{j=q} \right) \quad (6)$$

where $\varphi_{pose}^j(\boldsymbol{I})$ is the j-th pose SVM score (i.e. $\varphi_{pose}^j(\boldsymbol{I}) = \boldsymbol{\gamma}_j^\top \boldsymbol{x} + \beta_j$, where $\boldsymbol{x}$ is the HOG representation of $\boldsymbol{I}$).

In order to determine whether an image region $\boldsymbol{I}$ belongs to the weak negative set $\mathcal{S}_{wn}$, we could use the following.

$$\boldsymbol{I} \in \begin{cases} \mathcal{S}_{sp}, & \text{if } \Phi_{dpm}(\chi) > \tau_1 \\ \mathcal{S}_{wn}, & \text{if } \tau_1 \geq \Phi_{dpm}(\chi) > \tau_2 \\ \mathcal{S}_{sn}, & \text{if } \Phi_{dpm}(\chi) \leq \tau_2 \end{cases} \quad (7)$$

where τ_1 and τ_2 are the predefined thresholds from cross validation. $\mathcal{S}_{wn}$, $\mathcal{S}_{sp}$ and $\mathcal{S}_{sn}$ are the weak negative, strong positive and strong negative sets, respectively.

4. EXPERIMENTAL RESULTS

In the first evaluation, our proposed approach contrasted to two approaches: a baseline method HOG+SVM [8] and the state-of-the-art DPM [7] on the proposed Kangaroo dataset. All the hyper-parameters were determined empirically from the cross validation set.

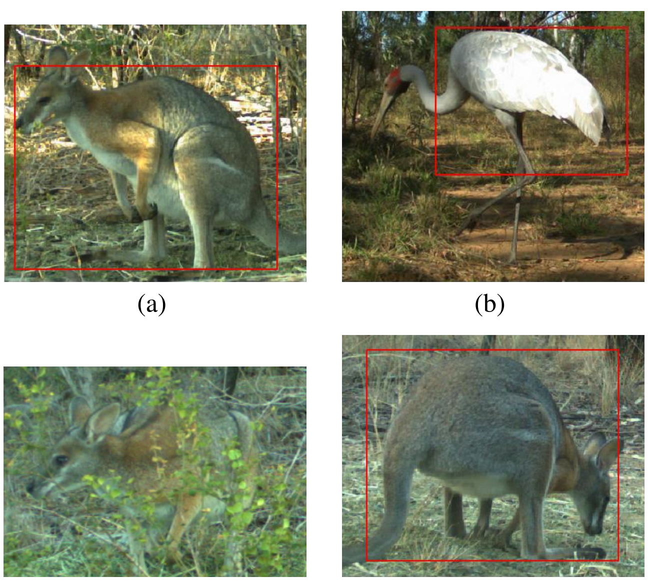
Table 1 shows the evaluation results. As we can see, the proposed approach achieves marked improvement over standard DPM. It is also worth noting that DPM has the lowest False Positive Rate (FPR); corroborating the previous analysis that says DPM has high specificity. As our proposal is an extension of DPM, it also has low FPR on par with DPM. The challenge posed by the dataset is also reflected in the performance of HOG+SVM method which could be considered as a good baseline method for pedestrian detection. (Other state-of-the-art pedestrian detection methods such as [9] will be investigated in the future.)

To further show the efficacy of our method, we present some qualitative results in Fig. 5. Since the proposed approach is based on DPM's strength in detecting body parts, we can see the example of a correct detection in (a) and a bird falsely detected as a kangaroo in (b). In (c), the proposed approach fails together with DPM due to the serious occlusion of most body parts. However, when only a few body parts are hard to detect like in (d), the proposed approach can work where DPM fails. This is due to the fact that the multi-pose SVM models incorporate the pose information when they are initially trained.

Methods	TPR	FPR	AP
HOG+SVM [8]	52.0%	18.0%	67.0%
DPM [7]	66.0%	**0.2%**	82.9%
Proposed	**68.8%**	0.3%	**84.25%**

Table 1. Detection Performance and Comparison with the state-of-the-art methods

In the second evaluation, we used our proposed approach in a Kangaroo population study. More precisely, we applied our detector on the raw frames collected in the national parks. The number of detections was then divided into 5 time scopes over 24 hours (i.e. 00:00-3:00, 3:00-7:00, 7:00-18:00, 18:00-22:00, 22:00-24:00). The time information from each frame was extracted from the time stamp provided from the camera. We collected 4,000 frames from four different locations (each location 1,000 frames). For each location, the 1,000 frames were further divided according to the 5 time scopes. The accumulated kangaroo population over the 24 hour time space is presented in Fig. 6.



(a) (b)

(c) (d)

Fig. 5. Kangaroo detection results made by proposed approach. (a): correct detection. (b): an example of false detection. (c): missed detection due to occlusion. (d): correct detection made by the proposed approach where DPM fails.

From this result, we found that kangaroos tend to be more active in the early evening in these locations. In addition, kangaroo is much more active in the night than day time. This finding corroborates previous biological studies suggesting that most kangaroos are nocturnal animals [10]. It is interesting to note that even using non-optimised research code, we only required six hours to compute the population over the 24-hour period at four locations; significantly more efficient and effective than manual methods. This demonstrates the efficacy of our method for future kangaroo studies.

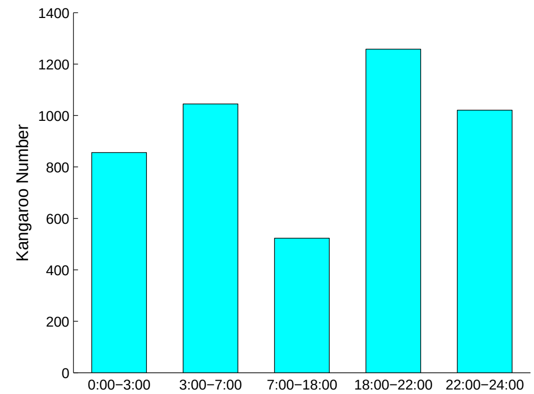


Fig. 6. Kangaroo's Population Distribution in 24-hours.

5. CONCLUSION

In this paper, we address problems in current kangaroo research such as population tracking and activity monitoring. However, traditional manual observation can be less-effective and more expensive than automated systems especially for kangaroo surveillance in the wild. As an important step for conducting experiments using computer vision tools, a suitable field dataset of kangaroos does not exist. To that end, we created a novel kangaroo dataset from field-work data. The dataset could be used by practitioners in developing automatic systems. As the initial step in the automation, we proposed a kangaroo detector which could be used to detect kangaroos over the video. The detector is based on the state-of-the-art Deformable Part Model approach. Since DPM may fail to detect kangaroos with occluded parts caused by different poses, we proposed to combine pose-specific SVM classifiers with DPM. The evaluation done in the proposed benchmark dataset shows the efficacy of our proposed detector. In this work, we also demonstrated that the proposed kangaroo detector could be used to conduct a study on kangaroo activity over time. In the future, we plan to enlarge the dataset with other Australian native animals such as dingos and emus. In addition other approaches such as attribute feature [11], context based systems [12, 13] and active learning [14] will be investigated.

6. ACKNOWLEDGEMENT

The authors like to thank the Queensland Parks and Wildlife Service for their animal surveillance data.

7. REFERENCES

[1] T. Clutton-Brock and B.C. Sheldon, "Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology," *Trends in Ecology and Evolution*, vol. 25, pp. 562–573, Sept. 2010.

[2] C.H. Lampert, H. Nickisch, and S. Harmeling, "Learning to detect unseen object classes by between-class attribute transfer," in *Computer Vision and Pattern Recognition, International Conference on (CVPR)*. IEEE, 2009, pp. 951–958.

[3] O.M. Parkhi, A. Vedaldi, A. Zisserman, and C.V. Jawahar, "Cats and dogs," in *Computer Vision and Pattern Recognition, International Conference on (CVPR)*. IEEE, 2012, pp. 3498–3505.

[4] W.W. Zhang, J. Sun, and X.O. Tang, "Cat head detectionhow to effectively exploit shape and texture features," in *Computer Vision, European Conference on (ECCV)*, 2008, pp. 802–816.

[5] J.P. Crall, C.V. Stewart, T.Y. Berger-Wolf, and D.I. etc. Rubenstein, "Hotspotter patterned species instance recognition," in *Application of Computer Vision, IEEE Winter Conference on (WACV)*. IEEE, 2013, pp. 230–237.

[6] M.J. Wilber, W.J. Scheirer, P. Leitner, and B. etc. Heflin, "Animal recognition in the mojave desert: Vision tools for field biologists," in *Application of Computer Vision, IEEE Winter Conference on (WACV)*. IEEE, 2013, pp. 206–213.

[7] P.F. Felzenszwalb, R.B. Girshick, D.McAllester, and D. Ramanan, "Object detection with discriminatively trained part based models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on (PAMI)*, vol. 32, pp. 1627–1645, Sept. 2010.

[8] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, International Conference on (CVPR)*. IEEE, 2005, vol. 1, pp. 886–893.

[9] P. Luo, Y.L. Tian, X.G. Wang, and X.O. Tang, "Switchable deep network for pedestrian detection," in *Computer Vision and Pattern Recognition, International Conference on (CVPR)*. IEEE, 2014, pp. 899–906.

[10] M. Daly, P.R. Behrends, and M.I. Wilson, "Activity patterns in small mammals," *Ecological Studies*, vol. 141, pp. 145–158, 2000.

[11] L.C. Liu, A. Wiliem, S.K. Chen, and B.C. Lovell, "Automatic image attribute selection for zero-shot learning of object categories," in *Pattern Recognition, Internation Conference on (ICPR)*. IEEE, 2014, pp. 2619–2624.

[12] Arnold Wiliem, Vamsi Madasu, Wageeh Boles, and Prasad Yarlagadda, "A context-based approach for detecting suspicious behaviours," in *Digital Image Computing: Techniques and Applications*. IEEE, 2009, pp. 146–153.

[13] Arnold Wiliem, Vamsi Madasu, Wageeh Boles, and Prasad Yarlagadda, "A suspicious behaviour detection using a context space model for smart surveillance systems," *Computer Vision and Image Understanding*, vol. 116, no. 2, pp. 194–209, 2012.

[14] Y. Yang, Z.G. Ma, F.P. Nie, X.J. Chang, and A.G. Hauptmann, "Multi-class active learning by uncertainty sampling with diversity maximization," *International Journal of Computer Vision (IJCV)*, pp. 1–15, Nov. 2014.