Animal Reidentification using Restricted Set Classification

Ludmila I. Kuncheva

School of Computer Science and Electronic Engineering, Bangor University, UK
email: *l.i.kuncheva@bangor.ac.uk*

Abstract

Individual animal recognition and re-identification from still images or video are useful for research in animal behaviour, environment preservation, biology and more. We propose to use Restricted Set Classification (RSC) for classifying multiple animals simultaneously from the same image. Our literature review revealed that this problem has not been solved thus far. We applied RSC on a koi fish video using a convolutional neural network (CNN) as the individual classifier. Our results demonstrate that RSC is significantly better than applying just the CNN, as it eliminates duplicate labels in the same image and improves the overall classification accuracy.

Keywords: Individual animal recognition, Reidentification, Restricted Set Classification

1. Introduction

Consider an environmental project where the scientist is interested in wild animal behaviour, and is monitoring the movements of a group of animals on a daily basis. In order to study the behaviour and the dynamic within the group, each animal has to be identified with a unique tag or name. As recording and processing 24-hour video footage is impractical, time-lapse footage can be used instead. This modality will render tracking methods infeasible and will require other methods for re-identification of the animals in the group images. Individual animal recognition has made significant advances [1, 2, 3, 4]. Interestingly, the current approaches try to identify each animal individually, disregarding the possibility that several animals from the same group can be present in the image. Having several animals in the same image poses an instant constraint on the classification task. Suppose that four subimages were extracted from an image, each containing an individual animal. In classifying those four individuals, we will have the extra knowledge that they all have *different* identities. Taking this restriction into account is expected to improve on the individual classification accuracy.

In this study we propose a methods for re-identification of animals from images using Restricted Set Classification (RSC) [5, 6, 7]. RSC belongs to the general area of weak supervision and non-standard classification [8]. A *set* of objects are classified together so that each object receives a unique label but the relationship between the objects (the context) is also taken into account. Our experiments with a video of a fish pond demonstrate that the classification accuracy increased compared to that of the naive approach where each object is classified individually.

The rest of the paper is organised as follows. Section 2 gives an overview and categorisation of the computer vision methods for animal recognition and reidentification. RSC is explained in Section 3. The experiment is reported in Section 4 and our conclusions are given in Section 5.

2. Methods and approaches for animal recognition

Past methods for animal identification such as physical branding, tagging, tattooing or radio frequency identification (RFID) were accurate but often invasive or at least intrusive to the animal [9, 10, 11]. Computer vision has been gaining momentum as an inexpensive and non-intrusive alternative.

2.1. Tasks

A substantial amount of research has been devoted to automatic analysis of camera-trap images in order to find out whether there is an animal in the camera view and also to identify the species [3, 12, 13, 14].

On the other hand, individual animal identification and re-identification are of great interest to the animal behaviourist. Started as hand-drawn patterns and descriptions of re-captured animals (e.g., swan bill patterns [15]), visual biometrics now dominate the research landscape of animal re-identification. Burghardt and Campbell [16] point out in 2007 that while tools for human re-identification from images abound, animal re-identification does not enjoy the same level of attention. Nonetheless, there are many studies, especially recent ones, that propose adapted or new methodologies for animal re-identification. One of the matching directions between human and animal re-identification is face/head identification in the image and subsequent recognition, predominantly for primates [17, 18, 19] but also for other animals such as cats, tigers, pandas, foxes, cheetahs [20], lions [21], lemurs [22] and cows [23]. The main interest, however, lies in the identification of the unique coat/skin pattern such as spots, stripes, creases, etc. [24, 25, 26, 16, 27].

In *tracking* animals in video footage, the main focus is on the trajectories of the movements of the individual animals [28, 29, 30, 31]. The animals have to be re-identified in each frame of the video. This is typically done based on two sources of information: the predicted position of the animal and the appearance. The leading source is the former, especially in the case when the animals are very similar in appearance such as fruit flies or ants.

In this study, we are interested in identification of individual animals from separate images (e.g. from time-lapse video footage), which may not form a succession suitable for tracking. Thus, the appearance of the animal is the only source of information.

Notably, the task of *group recognition* has not been approached thus far in the literature. Such groups exist in many of the studies, as can be seen by the published images, for example, a group of primates [19], piglets [32], cows [33] or African penguins [34]. Here we argue that taking the group into account will improve on the accuracy of the individual recognition.

2.2. Machine Learning and Computer Vision methods

The overwhelming majority of the literature on animal identification is concerned with what Schneider et al. [3] name *feature engineering*. This is a collective term for methods from *Computer Vision* for extracting informative features from images and videos. *Machine Learning* has been widely applied in studying animal behaviour [35] but it can offer a lot more, specifically to animal re-identification. Typically, animal re-identification relies on a database of stored images and a comparison of a candidate image with the database to retrieve the closest match. This approach is the same as the nearest-neighbour classifier in machine learning. The use of state-of-the-art machine learning is handicapped by the relatively small number of images of a single individual in the database. This is the case in many applications, especially those relying on crowd sourcing for collecting images of individual animals. Even though the database may have a substantial size, containing data for thousands of animals, training accurate classifiers will be impossible due to the small count of images per animal. In our scenario, a group of animals is observed over a period of time, allowing for collecting an adequate number of samples for each individual. In this case, advanced classification methods could be applied.

Recently, deep neural networks (DNN) have established themselves as the preferred tool for various tasks in animal re-identification [36, 37, 13, 14]. While most of the applications are about detecting bounding boxes, face matching and similarity evaluation [36], with a sufficient number of images per individual in the data set, DNN can be used as a high-accuracy classifier. We use a DNN classifier in the experiment in this study. We show that taking advantage of the group context improves the classification accuracy of the DNN classifier.

3. Restricted Set Classification

3.1. Definitions

Definition 1. The *restricted set classification problem* is defined as follows [5, 7]. Let $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ be a set of instances such that at most k_i instances come from class $\omega_i \in \Omega$, where $\Omega = \{\omega_1, \ldots, \omega_c\}$ is the set of class labels (animal identities). The task is to find labels for all elements of X so that the restriction holds.

Note that $k_1 + \ldots + k_c = k \geq m$.

Definition 2. A *base classifier* D is a classifier that assigns a class label to an instance $\mathbf{x} \in \mathbb{R}^n$

$$D : \mathbb{R}^n \to \Omega. \tag{1}$$

We also require that D provides estimates of the posterior probabilities $P(\omega_1|\mathbf{x}), \ldots, P(\omega_c|\mathbf{x})$.

For the animal re-identification problem, we assume that there is a group of c animals that we wish to monitor. The c animals are the classes of interest. Assuming that there are no newcomers to the group, in any given image, there may be at most c different animals. This problem is a version of the restricted set classification problem, which we termed "who-is-missing" [6]. In this case, $k_1 = k_2 = \ldots = k_c = 1$, and $m \leq c$. Classifier D will output the probabilities for the c classes for a given animal sub-image $\mathbf{x}$.

Definition 3. A *super-label* for set X is any collection of m labels from Ω so that any instance $\mathbf{x} \in X$ receives a single label. A super-label will be called *consistent* if it satisfies the requirement that at most k_i labels are equal to ω_i, $i = 1, \ldots, c$.

Denote by $\mathcal{S}$ the set of all possible super-labels of X. For cardinality $|X| = m$, $\mathcal{S}$ has $\frac{c!}{(c-m)!}$ elements. Let $P = [p_{ij}]$ be a matrix of size $m \times c$ that contains the posterior probability estimates obtained from the base classifier D applied to X. Entry p_{ij} is the estimate of $P(\omega_j|\mathbf{x}_i)$. Let $\mathcal{P}$ be the set of all matrices P.

Definition 4. A *set classifier* D_{set} assigns a super-label to any set X using the output of classifier D, that is

$$D_{\text{set}}(X, D) : \mathcal{P} \rightarrow \mathcal{S}. \tag{2}$$

3.2. Evaluation of accuracy of a set classifier

We consider two type of estimates of the accuracy of D_{set} for a given set X:

- A_T, total accuracy: $A_T = 1$ if *all* labels are correctly assigned to the instances in X, and $A_T = 0$, otherwise;

- A_P, partial accuracy: A_P is the *proportion* correctly labelled instances across the whole set of instances (identical to classification accuracy).

3.3. Three set classifiers

We consider here the following set classifiers:

(1) Independent set classifier (Baseline) D^i_{set}. This classifier takes the labels suggested by D without any modification and collates them to make the super-label of X. Note that this approach does not guard against having multiple labels of the same animal for different objects in X (sub-images). Thus, the super-label is not guaranteed to be consistent.

In D^i_{set}, all instances are labelled independently. Assuming that D's accuracy is p, the accuracy measures of D^i_{set} are

$$A_T(D^i_{\text{set}}) = p^{\mathbb{E}[m]}, \tag{3}$$

where $\mathbb{E}[m]$ is the expected value of the cardinality of X, and

$$A_P(D^i_{\text{set}}) = p. \tag{4}$$

(2) Greedy set classifier D^g_{set}. The input to this set classifier are the posterior probabilities $P(\omega_i|\mathbf{x})$ produced by classifier D for $i = 1 \ldots, c$ for the given $\mathbf{x} \in \mathbb{R}^n$. The Greedy Set Classifier labels X according to the following algorithm:

1. Initialise a set $V = \emptyset$ to store the assigned object-class pairs.
2. Identify the largest posterior probability $P(\omega^*_j|\mathbf{x}^*_j)$ among the objects and classes not assigned so far.
3. Remove ω^*_j from the list of available classes, and $\mathbf{x}^*_j$ from the list of available objects, and add the pair to set V.

4. If there are no objects left, stop and return V. Else, continue from step 2.

The Greedy set classifier guarantees consistent super-labels. It can be formally proved [7] that for two-class problems, and 2 instances in each image,

$$A_P(D^g_{\text{set}}) > A_P(D^i_{\text{set}}) \ . \tag{5}$$

(3) Hungarian set classifier D^h_{set}. Here we propose to use this set classifier for the animal re-identification problem. It is based on the Hungarian assignment algorithm further developed by Kuhn and Munkres, also known as Kuhn-Munkres algorithm [38]. Proposed originally for $c \times c$ matrices, the Hungarian algorithm has been extended for rectangular matrices [39]. Below we demonstrate the mathematical rationale behind the Hungarian set classifier.

We shall assume that the objects in X are drawn independently from their respective classes, that is, $\mathbf{x}_i$ is drawn from the distribution of class ω_i, independently of the remaining $m-1$ objects. It can be argued that the appearance of a given animal in the image does not depend on the appearances of the other animals. For example, one chimpanzee's face could be in full frontal view in the image while another's could be in semi-profile. However, animals interact in certain ways, and there may be patterns of interactions that will correlate the animals' appearance. For example, all animals can be on high alert and looking in the direction of the approaching danger. Also, they may all be looking at a food source. Correlated appearances may be used to improve the accuracy of the set classifier. For such correlated appearances to be evaluated and used, we need a large amount of data. While this is an interesting research line, for the purposes of this study, we will assume independent appearances.

With this assumption in place, the likelihood of a super-label $S = \langle s_1, \ldots, s_m \rangle$, $s_i \in \Omega$ is

$$L(S|X) \propto \prod_{i=1}^{m} P(s_i|\mathbf{x}_i) \ . \tag{6}$$

The optimal super-label S^* will be the one maximising L (equivalently $\log(L)$), that is

$$S^* = \arg\max_{S\in\mathcal{S}} \sum_{i=1}^{m} \log(P(s_i|\mathbf{x}_i)) \ , \tag{7}$$

It can be shown that the Greedy set classifier D^g_{set} will not guarantee the optimal solution. We can cast the problem defined by Equation (7) as a linear programming problem (LP). Let T be a reward matrix with entries $t_{i,j} = P(\omega_i|\mathbf{x}_j)$. Introducing the unknowns $r_{(i,j)} \in \{0,1\}$, $i = 1,\ldots,m$, $j = 1,\ldots,c$, the LP is

$$\max \sum_{i=1}^{m}\sum_{j=1}^{c} r_{(i,j)} \ \log(t_{i,j}),$$

subject to

$$\sum_{i=1}^{m} r_{(i,j)} \leq 1, \quad j = 1,\ldots,c,$$
$$\sum_{j=1}^{c} r_{(i,j)} = 1, \quad i = 1,\ldots,m.$$

The Hungarian assignment algorithm provides the solution to this LP problem, guaranteeing that the obtained super-label is S^*.[1]

[1] MATLAB code for the Restricted Set Classification with the three set classifiers is available at `https://github.com/LucyKuncheva/Restricted-Set-Classification` .

4. Experiment

4.1. Data

Koi is an informal group name of the coloured variants of the Amur carp (*Cyprinus rubrofuscus*) that are kept for decorative purposes. We sourced a video from Pixabay to use as an example with multiple animals in the same frame. The video consists of 536 frames with 9 fish in total. We named the fish randomly (regardless of their true gender): Catherine, Dwayne, Florence, Humphrey, JP, Jack, Ruby, Selwyn and Siobhan. Each frame of the video was manually segmented by defining a bounding box around the visible part of the fish. Each sub-image was stored with the respective name tag. By segmenting the video manually, we bypass the main bottleneck of animal re-identification. We did this on purpose, because the claim in this study concerns the last stage of the classification.

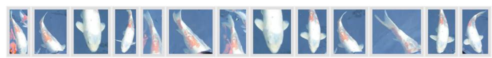
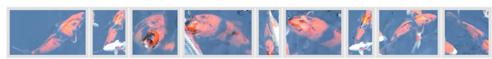
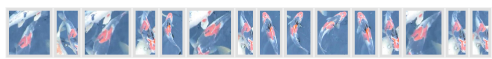
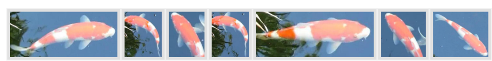
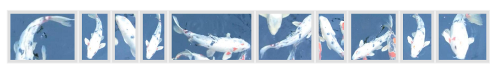
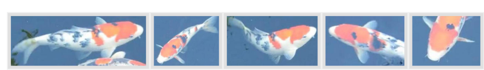
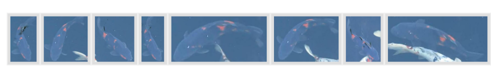
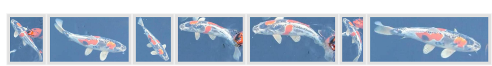
Overall, 1640 sub-images were cropped from the 536 frames, each one containing one fish individual. Table 1 shows the distribution of the classes as well as examples from each class.

4.2. The Independent classifier D

The Independent classifier D used in this experiment was a deep neural network (Convolutional Neural Network, CNN) from the Deep Learning Toolbox of MATLAB, version 12.1 (R2019a). We chose this model because of the overwhelming evidence in the literature reviewed in Section 2 in favour of deep learning models. Since we have nine classes, and the data set is not very large compared to standard deep learning set-ups, we kept the CNN as simple as possible using its default structure and training choices:

- Structure:

Table 1: Distribution and examples of the classes (nine individual fish).

Name	Number	Examples
Catherine	103 (6.28%)	
Dwayne	228 (13.90%)	
Florence	145 (8.84%)	
Humphrey	152 (9.27%)	
JP	233 (14.21%)	
Jack	161 (9.82%)	
Ruby	265 (16.16%)	
Selwyn	94 (5.73%)	
Siobhan	259 (15.79%)	
Total	1640 (100.00%)	

- an input layer with colour images sized 56-by-56 pixels
- a convolution layer with 10 filters of size 5-by-5
- a RELU layer
- a max pooling layer with pool size [2, 2] and stride [2, 2]
- a fully connected layer for 9 classes
- a softmax layer returning the posterior probabilities for the classes.

- Training parameters: We used the default stochastic gradient descent with momentum (SGDM) optimiser with 30 maximum number of epochs and initial learning rate 0.0001. The data was shuffled after each epoch.

- Data Augmentation:

  In view of the relatively small data size, we opted for augmentation. Each image in the data set was processed twice with random augmentation, thereby tripling the training data size. MATLAB function `imageDataAugmenter` was applied with the following augmentation choices:

  - random rotation at an angle between 0 and 360 degrees. This transformation was deemed reasonable because the fish were swimming in any direction in the video.
  - random scaling on the x-axis and a separate random scaling on the y-axis at a ratio between 0.8 and 1.
  - random shear on the x-axis and a separate random shear on the y-axis at an angle between -20 degrees and 20 degrees.
  - random translation on the x-axis and a separate random translation on the y-axis at ± 4 pixels.

The experiments were carried out on a HP Pavilion Laptop 15-cs1xxx with graphics card NVIDIA GeForceGTX 1050 with Max-Q Design and operating system Windows 10 Home 64-bit.

We considered the classification accuracy sufficient for the purposes of this study.

4.3. An example

The expected improvement on the classification accuracy by using RSC is illustrated by the following example. Figure 1 shows the original image containing five fish and Catherine's head. Catherine was not segmented in this frame because she would not be identifiable from such a small part.

Figure 1: Original image with five recognisable fish.

Figure 2 shows the labels assigned by the Independent set classifier. This classifier labelled two fish as Jack and mistook Humphrey for Florence.

The two proper set classifiers guarantee that the restriction is observed (no repeated labels). The Greedy set classifier (Figure 3) resolves the conflict

Figure 2: Labelling by the Independent set classifier. Correct labels are marked with a yellow bounding box, repeated labels with red, and wrong (non-repeated) labels, with blue.

by relabelling the bottom "Jack" to "Catherine", which is also largely white in colour. It, however fails to recover the correct label of "Florence".

Finally, the Hungarian set classifier reassigns the labels to their correct values as shown in Figure 4.

4.4. Cross-validation experiment

To compare the independent set classifier with the two proposed variants, we ran a 2-fold, 3-fold, 5-fold and 10-fold cross-validation by splitting the set of *frames* into folds. We then retrieved the objects in the training frames to collate a training set of *objects* for D. The objects in the testing frames were pooled to create the testing data set. Care should be taken when preparing training and testing data from a video. The individual fish images cropped from consecutive frames will be very similar. Thus, if two consecutive frames are randomly assigned to the training and the testing part, respectively, the

Figure 3: Labelling by the Greedy set classifier. Correct labels are marked with a yellow bounding box and wrong labels, with blue.

classifier may achieve a deceptively high accuracy. Therefore, we carried out the cross-validation by splitting the video into time intervals. For example, in the two-fold cross-validation experiment, the first 268 frames were taken as the first fold and the remaining 268 frames, as the second fold. All cross-validation and data shuffle experiments were carried out in this manner.

The classification accuracies A_P and A_T, averaged across the folds, are shown in Tables 2 and 3, respectively. The last column in Table 2 shows the number of frames with repeated labels in the testing set. These numbers show how many chances there have been for the RSC to improve on the Independent set classifier. As D becomes more accurate with the growing size of the training data, the number of frames with repeated labels declines.

Tables 2 and 3 show that the best option is the Hungarian set classifier for both A_P and A_T. Naturally, the total accuracy is lower than the partial accuracy, as it requires that *all* objects in the frame are correctly labelled.

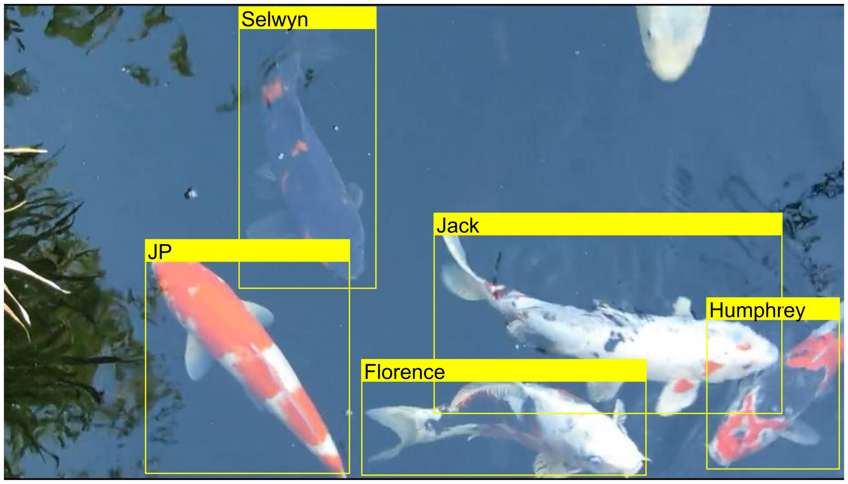


Figure 4: Labelling by the Hungarian set classifier. Correct labels are marked with a yellow bounding box.

A_T is more affected by the training sample size than A_P. In the 2-fold cross-validation the training sample contained 820 objects on average, while in the 10-fold cv, this size was 1476.

Statistical validation of these results is only feasible for the 10-fold cross-validation experiment. Table 4 shows the relationship between the set classifiers. The probabilities were calculated using a Bayesian correlated t-test proposed by Benavoli et al. [40]. The paper argues that the p-value based statistical analyses are inaccurate and misleading. The proposed alternative directly answers the question of 'what is the probability that classifier A is better than classifier B?' For two classifiers and one data set, the authors propose using a cross-validation and a Bayesian correlated t-test as a replacement of the conventional t-test or even the corrected t-test due to Nadeau

Table 2: Partial accuracy A_P for the cross-validation experiment and the three set classifiers. The last column shows the number of frames with repeated labels.

	Hungarian	Greedy	Independent	Repeated labels
2-fold	0.4257	0.4063	0.4248	206
3-fold	0.5204	0.4856	0.4613	250
5-fold	0.4819	0.4743	0.4701	177
10-fold	0.6243	0.6138	0.6039	165

Table 3: Total accuracy A_T for the cross-validation experiment and the three set classifiers

	Hungarian	Greedy	Independent
2-fold	0.1856	0.1819	0.1726
3-fold	0.2559	0.2236	0.1593
5-fold	0.1611	0.1739	0.1412
10-fold	0.2502	0.2481	0.2219

and Bengio [41].[2] Again, the Hungarian set classifier is substantially better than the Greedy set classifier and the Independent classifier in view of both partial accuracy and total accuracy.

To illustrate the difference in the performances of H and I, we compare their partial accuracy A_P. Based on the number of folds of the cross-validation and the number of repeats, a distribution of the paired differences between the classification accuracies was calculated and plotted in Figure 5. It is an extended Student distribution with degrees of freedom $N-1$, where

[2] Python library `baycomp` contains the functions for this analysis `https://github.com/janezd/baycomp`. Here we used a MATLAB version available at `https://github.com/LucyKuncheva/Bayesian-Analysis-for-Comparing-Classifiers`.

Table 4: Probabilities from the Bayesian correlated t-test for A_P. The value in cell (i, j) is the probability that Method i (the row) is better than Method j (the column).

	Partial accuracy A_P			Total accuracy A_T		
	H	G	I	H	G	I
Hungarian	0.0000	0.9221	0.8581	0.0000	0.9884	0.9988
Greedy	0.0779	0.0000	0.7282	0.0116	0.0000	0.9989
Independent	0.1419	0.2718	0.0000	0.0012	0.0011	0.0000

Note: The column headings are abbreviated as H (Hungarian set classifier), G (Greedy set classifier) and I (Independent classifier)

N is the number of differences, the mean is equal to the sample mean $\hat{x}$, the variance is

$$v = \left(\frac{1}{N} + \frac{\rho}{(1-\rho)}\right)\hat{\sigma}^2$$

where $\rho = 1/K$ for a K-fold cross-validation and $\hat{\sigma}^2$ is the sample variance. The larger shaded area to the right of 0 indicates that the differences are mostly positive, and the Hungarian set classifier is better than the Independent set classifier.

Based on these results, we recommend the Hungarian set classifier for the problem of individual animal recognition from images containing groups of animals.

4.5. Data-shuffle experiment

This part of the experiment examines the effect of the training set size on the improvement offered by the set classifiers. We carried out 100 runs of training and testing with a given proportion split. As explained earlier, we kept the testing set as a time-contiguous part of the video with a random

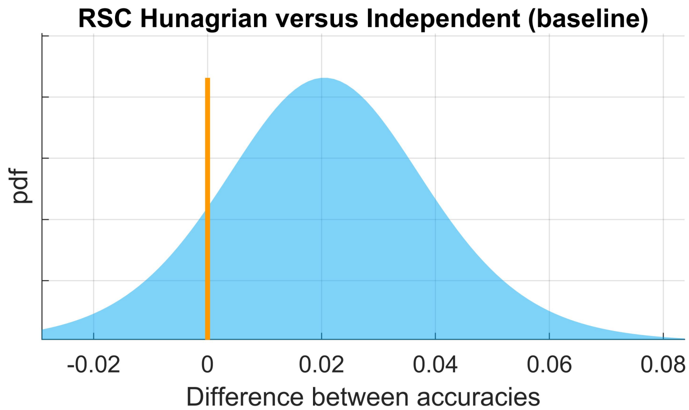


Figure 5: Distribution of the difference $A_P(D^h_{\text{set}}) - A_P(D^i_{\text{set}})$ [40]. Most values lie to the right of zero (orange line). The shaded area to the right of zero gives the probability 0.8581 that the Hungarian set classifier is better than the Independent classifier.

starting point. We ran experiments with split proportions $\{0.5, 0.7, 0.9\}$. The accuracies A_P and A_T for the three methods are shown in Tables 5 and 6, respectively.

Table 5: Partial accuracy A_P for the data shuffle experiment and the three set classifiers for splitting proportions P. The last column (R) shows the number of frames with repeated labels.

P	Hungarian	Greedy	Independent	R
0.5	0.3823	0.3624	0.3350	131.7100
0.7	0.4216	0.4079	0.3696	74.0500
0.9	0.5862	0.5783	0.5675	15.1300

The results ascertain again the advantage of the RSC classifiers over the independent labelling of the objects. Notably, there is a dramatic fall in the classification accuracy for smaller training data sizes. Further statistical analyses using the Bayesian correlated t-test also unequivocally select the

Table 6: Total accuracy A_T for the data shuffle experiment and the three set classifiers for splitting proportions P.

P	Hungarian	Greedy	Independent
0.5	0.1155	0.1132	0.1001
0.7	0.0659	0.0641	0.0494
0.9	0.1427	0.1414	0.1071

Hungarian set classifier as the best of the three alternatives. The dominance of the Hungarian set classifier is even more prominent compared to the cross-validation experiment. For all split proportions, for both A_P and A_T, the probability that the Hungarian set classifier is better the Greedy and the Independent set classifiers was evaluated at 1. Similarly, The Greedy set classifier dominated the Independent set classifier in all experiments with probability evaluated at 1.

The *rate* of improvement was not affected much by the training sample size, unlike the accuracy itself. While accuracy A_P drops by over 20% (in absolute units) when 50% of the data is used for training compared to 90% of the data, the improvement offered by the RSC methods is within 5% for both training proportions. Still, the less accurate individual classifier for the 50% split leaves more room for improvement. To illustrate this, we show two box-plots in Figure 6. The three methods were plotted next to one another for split proportions 0.5 and 0.9. In both plots, the box for the Hungarian set classifier is higher than the other two boxes. The dotted line marks the median of the Independent set classifier taken as baseline.

While the improvement achieved by the Hungarian and the Greedy set classifiers is visible, the *rate* of improvement is similar between the two proportions. The reason for this is that if the independent classifier D is not very good (the case with the smaller training data), then there are too many mistakes. For example, a frame containing fish that is labelled wrongly but there are no repeated labels, will receive the same labels from all set classifiers. Therefore, the potential of correcting the repeated labels through any set classifier is limited by the accuracy of D.

Fine-grained classifiers based on more complex CNN architectures may be accurate, provided there is enough data for training. In this case, the effect of RSC may be too small to warrant its use.

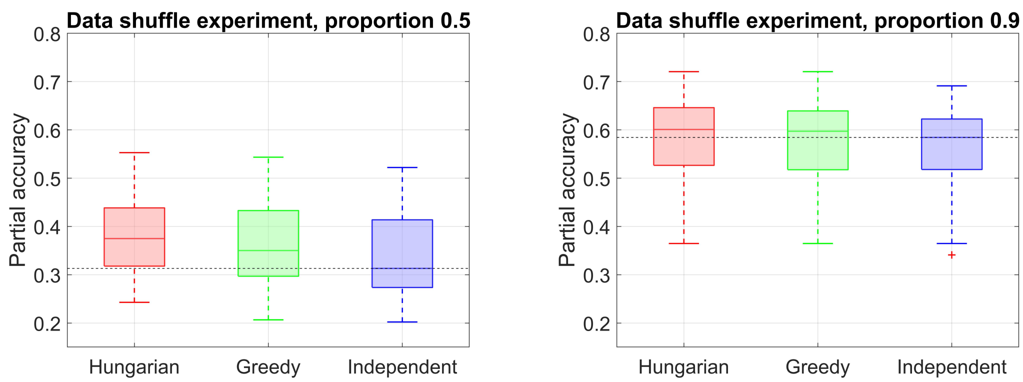


Figure 6: Box-plots of the classification accuracies A_P for split proportions 0.5 and 0.9. The dashed horizontal line indicates the median of the baseline method (the Independent set classifier).

Table 7 shows the computational times of the data shuffle experiments. The table also shows the number of *frames* in the training data as well as the average number of images in the training data (after augmentation) and the average number of frames with repeated labels. The times are calculated as the average of the 100 runs and are shown in seconds. The CNN training takes the bulk of the time. Testing is a small fraction of the training time, and the add-ons through the RSC classifier are also relatively small. The Hungarian set classifier is slightly more expensive than the Greedy set classifier and can be used as a viable extension of the individual CNN in the RSC setting.

Table 7: Computational times in seconds for the data shuffle experiment (100 runs with each proportion split). P is the proportion of the video for training; N is the number of individual training sub-images per run; R is the number frames with repeated labels per run.

			Times (s)			
P (frames)	N	R	Train	Test	Hungarian	Greedy
0.5 (268)	2151.0	131.7	15.2	0.120	0.027	0.021
0.7 (375)	3309.0	74.0	20.8	0.088	0.016	0.013
0.9 (482)	4146.0	15.1	28.8	0.055	0.005	0.005

5. Conclusion

In this study we advocate using a Restricted Set Classification instead of independent classification for individual animal recognition. We demonstrate that RSC can correct mistakes when there are more than one animal in the same image assigned to the same individual label. The best RSC version was the Hungarian set classifier which assigns the most probable labels while observing the restriction of no repeated labels in the same frame.

Note that no further labelling is needed for RSC to work. The only information that it uses is that the set of sub-images that are labelled together come from the same frame.

We observed that the potential of correcting repeated labels through a set classifier is limited by the accuracy of the base classifier. It would be interesting to prove the limits of the improvement and determine how calibration of posterior probabilities affects it.

This study bypasses the considerable problem of image segmentation by assuming that the correct bounding boxes and labels are available. The result from any classification will be preconditioned by the accuracy of the segmentation. If the segmentation is accurate, then the classifier (CNN) could be good on its own. This means that the individual classification accuracy will be high, and the benefits from RSC may not be that great. The proposed approach will be most useful when the individual accuracy leaves room for improvement. It will be interesting to study the effect of automatic segmentation on the improvement potential of RSC.

While the results in favour of the RSC classifier are compelling, we have used only one data set (the koi fish video). We could not find another suitable annotated and labelled data set with multiple individual animals to expand our experiment. We are currently preparing a second data set from a video containing a group of pigeons.

An important future line of research is accommodating objects from classes that were not represented in the training data. In other words, the individual classifier D should be able to realise its own competence by outputting a probability vector that does not necessarily sum up to 1. The 'leftover' probability will allow for assigning label 'don't know' to accommodate unseen classes. The set classifier should be modified accordingly. Further on, constrained clustering can be used to label the objects in the 'don't know' category into different classes (individuals). The constrained version of the clustering will ensure that the RSC restrictions are in place, that is, there cannot be more than one individual with the same class label in a single image.

Acknowledgements

The koi fish video was sourced from Pixabay under the Pixabay license. https://pixabay.com/videos/koi-carp-fish-ornamental-fish-swim-5652/

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] H. S. Kühl, T. Burghardt, Animal biometrics: quantifying and detecting phenotypic appearance, Trends in Ecology & Evolution 28 (2013) 432–441. doi:10.1016/j.tree.2013.02.013.

[2] J. Parham, C. Stewart, J. Crall, D. Rubenstein, J. Holmberg, T. Berger-Wolf, An animal detection pipeline for identification, in: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2018. doi:10.1109/wacv.2018.00123.

[3] S. Schneider, G. W. Taylor, S. Kremer, Deep learning object detection methods for ecological camera trap data, in: 2018 15th Conference on Computer and Robot Vision (CRV), IEEE, 2018. doi:10.1109/crv.2018.00052.

[4] S. Kumar, S. K. Singh, Visual animal biometrics: survey, IET Biometrics 6 (2017) 139–156. doi:10.1049/iet-bmt.2016.0017.

[5] L. I. Kuncheva, Full-class set classification using the Hungarian algorithm, International Journal of Machine Learning and Cybernetics 1 (2010) 53–61. doi:DOI10.1007/s13042-010-0002-z.

[6] L. I. Kuncheva, A. S. Jackson, Who is missing? A new pattern recognition puzzle, in: Intenrational Conference on Statistical, Structural and Syntactic Pattern Recognition (S+SSPR), volume LNCS 8621, Springer, Joensuu, Funland, 2014, pp. 243–252.

[7] L. I. Kuncheva, J. J. Rodríguez, A. S. Jackson, Restricted set classification: Who is there?, Pattern Recognition 63 (2017) 158–170.

[8] J. Hernández-González, I. Inza, J. A. Lozano, Weak supervision and other non-standard classification problems: A taxonomy, Pattern Recognition Letters 69 (2016) 49 – 55. URL: http://www.sciencedirect.com/science/article/pii/S0167865515003505. doi:https://doi.org/10.1016/j.patrec.2015.10.008.

[9] S. Schneider, G. W. Taylor, S. Linquist, S. C. Kremer, Past, present and future approaches using computer vision for animal re-identification from camera trap data, Methods in Ecology and Evolution 10 (2019) 461–470. doi:10.1111/2041-210x.13133.

[10] A. I. Awad, From classical methods to animal biometrics: A review on cattle identification and tracking, Computers and Electronics in Agriculture 123 (2016) 423–435. doi:10.1016/j.compag.2016.03.014.

[11] W. J. Eradus, M. B. Jansen, Animal identification and monitoring, Computers and Electronics in Agriculture 24 (1999) 91–98. doi:10.1016/S0168-1699(99)00039-3.

[12] M. S. Norouzzadeh, A. Nguyen, M. Kosmala, A. Swanson, M. S. Palmer, C. Packer, J. Clune, Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning, Proceedings of

the National Academy of Sciences 115 (2018) E5716–E5725. doi:10.1073/pnas.1719367115.

[13] G. Chen, T. X. Han, Z. He, R. Kays, T. Forrester, Deep convolutional neural network based species recognition for wild animal monitoring, in: 2014 IEEE International Conference on Image Processing (ICIP), IEEE, 2014. doi:10.1109/icip.2014.7025172.

[14] A. G. Villa, A. Salazar, F. Vargas, Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks, Ecological Informatics 41 (2017) 24–32. doi:10.1016/j.ecoinf.2017.07.004.

[15] M. E. Evans, Recognizing individual Bewick's Swans by bill pattern, Wildfowl 28 (1977) 153–158.

[16] T. Burghardt, N. Campbell, Individual animal identification using visual biometrics on deformable coat-patterns., in: Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007), Icvs, 2007, p. 10 pp.

[17] D. Deb, S. Wiper, S. Gong, Y. Shi, C. Tymoszek, A. Fletcher, A. K. Jain, Face recognition: Primates in the wild, in: 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS), IEEE, 2018. doi:10.1109/btas.2018.8698538.

[18] S. Schneider, G. W. Taylor, S. Linquist, S. C. Kremer, Similarity learning networks for animal individual re-identification – beyond the capabilities of a human observer (2019). arXiv:1902.09324v4.

[19] D. Schofield, A. Nagrani, A. Zisserman, M. Hayashi, T. Matsuzawa, D. Biro, S. Carvalho, Chimpanzee face recognition from videos in the wild using deep learning, Science Advances 5 (2019) eaaw0736. doi:10.1126/sciadv.aaw0736.

[20] W. Zhang, J. Sun, X. Tang, From tiger to panda: Animal head detection, IEEE Transactions on Image Processing 20 (2011) 1696–1708. doi:10.1109/TIP.2010.2099126.

[21] T. Burghardt, J. Ćalić, Analysing animal behaviour in wildlife videos using face detection and tracking, IEE Proceedings: Vision, Image and Signal Processing 153 (2006) 305–312. doi:10.1049/ip-vis:20050052.

[22] D. Crouse, R. L. Jacobs, Z. Richardson, S. Klum, A. Jain, A. L. Baden, S. R. Tecot, LemurFaceID: a face recognition system to facilitate individual identification of lemurs, BMC Zoology 2 (2017). doi:10.1186/s40850-016-0011-9.

[23] L. Bergamini, A. Porrello, A. C. Dondona, E. D. Negro, M. Mattioli, N. D'Alterio, S. Calderara, Multi-views embedding for cattle re-identification, 2019. arXiv:1902.04886v1.

[24] M. Lahiri, C. Tantipathananandh, R. Warungu, D. I. Rubenstein, T. Y. Berger-Wolf, Biometric animal databases from field photographs: Identification of individual zebra in the wild, in: Proceedings of the 1st ACM International Conference on Multimedia Retrieval, ICMR'11, 2011. doi:10.1145/1991996.1992002.

[25] P. Buehler, B. Carroll, A. Bhatia, V. Gupta, D. E. Lee, An automated program to find animals and crop photographs for individual recogni-

tion, Ecological Informatics 50 (2019) 191–196. doi:10.1016/j.ecoinf.2019.02.003.

[26] T. Burghardt, B. Thomas, P. Barham, J. Calic, Automated Visual Recognition of Individual African Penguins, Computer (2004) 1–10. URL: http://epubs.surrey.ac.uk/531569/.

[27] C. W. Speed, M. G. Meekan, C. J. Bradshaw, Spot the match - Wildlife photo-identification using information theory, Frontiers in Zoology 4 (2007) 1–11. doi:10.1186/1742-9994-4-2.

[28] Z. Khan, T. Balch, F. Dellaert, MCMC-based particle filtering for tracking a variable number of interacting targets, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2005) 1805–1819.

[29] A. Pérez-Escudero, J. Vicente-Page, R. C. Hinz, S. Arganda, G. G. de Polavieja, idTracker: tracking individuals in a group by automatic identification of unmarked animals, Nature Methods 11 (2014) 743–748. doi:10.1038/nmeth.2994.

[30] F. Romero-Ferrero, M. G. Bergomi, R. C. Hinz, F. J. H. Heras, G. G. de Polavieja, idtracker.ai: tracking all individuals in small or large collectives of unmarked animals, Nature Methods 16 (2019) 179–182. doi:10.1038/s41592-018-0295-5.

[31] R. Ardekani, A. Biyani, J. E. Dalton, J. B. Saltz, M. N. Arbeitman, J. Tower, S. Nuzhdin, S. Tavaré, Three-dimensional tracking and behaviour monitoring of multiple fruit flies, Journal of The Royal Society Interface 10 (2013) 20120547. doi:10.1098/rsif.2012.0547.

[32] M. Kashiha, C. Bahr, S. Ott, C. P. Moons, T. A. Niewold, F. Ödberg, D. Berckmans, Automatic identification of marked pigs in a pen using image pattern recognition, Computers and Electronics in Agriculture 93 (2013) 111–120. doi:10.1016/j.compag.2013.01.013.

[33] W. Andrew, C. Greatwood, T. Burghardt, Visual localisation and individual identification of holstein friesian cattle via deep learning, in: Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017, volume 2018-January, 2017, pp. 2850–2859. doi:10.1109/ICCVW.2017.336.

[34] R. B. Sherley, T. Burghardt, P. J. Barham, N. Campbell, I. C. Cuthill, Spotting the difference: Towards fully-automated population monitoring of African penguins Spheniscus demersus, Endangered Species Research 11 (2010) 101–111. doi:10.3354/esr00267.

[35] J. J. Valletta, C. Torney, M. Kings, A. Thornton, J. Madden, Applications of machine learning in animal behaviour studies, Animal Behaviour 124 (2017) 203–220. URL: http://dx.doi.org/10.1016/j.anbehav.2016.12.005. doi:10.1016/j.anbehav.2016.12.005.

[36] S. Schneider, Deep learning methods and applications for animal re-identification, in: 2020 IEEE Winter Conference on Applications of Computer Vision Workshops, WACVW, 2020. URL: https://openaccess.thecvf.com/WACV2020_workshops/WACV2020_w2.

[37] M. Willi, R. T. Pitman, A. W. Cardoso, C. Locke, A. Swanson, A. Boyer, M. Veldthuis, L. Fortson, Identifying animal species in camera trap images using deep learning and citizen science, Methods in Ecology and Evolution 10 (2018) 80–91. doi:10.1111/2041-210x.13099.

[38] H. W. Kuhn, The Hungarian Method for the assignment problem, Naval Research Logistic Quarterly 2 (1955) 83–97.

[39] F. Bourgeois, J.-C. Lassalle, An extension of the Munkres algorithm for the assignment problem to rectangular matrices, Communications ACM 14 (1971) 802–804.

[40] A. Benavoli, G. Corani, J. Demšar, M. Zaffalon, Time for a change: a tutorial for comparing multiple classifiers through bayesian analysis, Journal of Machine Learning Research 18 (2017) 1–36.

[41] C. Nadeau, Y. Bengio, Inference for the generalization error, Machine Learning 52 (2003) 239–281.