On detection of novel categories and subcategories of images using incongruence

Dalia Coppi* dalia.coppi@unimore.it
Teofilo de Campos †§ t.decampos@surrey.ac.uk
Fei Yan † f.yan@surrey.ac.uk
Josef Kittler † j.kittler@surrey.ac.uk
Rita Cucchiara † rita.cucchiara@unimore.it

*DII - University of Modena and Reggio Emilia via Vignolese 905 Modena, Italy

† CVSSP University of Surrey Guildford, GU2 7XH UK

§ DCS and SITraN University of Sheffield Reg. Court, 211 Portobello Sheffield, S1 4DP, UK

ABSTRACT

Novelty detection is a crucial task in the development of autonomous vision systems. It aims at detecting if samples do not conform with the learnt models. In this paper, we consider the problem of detecting novelty in object recognition problems in which the set of object classes are grouped to form a semantic hierarchy. We follow the idea that, within a semantic hierarchy, novel samples can be defined as samples whose categorization at a specific level contrasts with the categorization at a more general level. This measure indicates if a sample is novel and, in that case, if it is likely to belong to a novel broad category or to a novel sub-category. We present an evaluation of this approach on two hierarchical subsets of the Caltech256 objects dataset and on the SUN scenes dataset, with different classification schemes. We obtain an improvement over Weinshall et al. and show that it is possible to bypass their normalisation heuristic. We demonstrate that this approach achieves good novelty detection rates as far as the conceptual taxonomy is congruent with the visual hierarchy, but tends to fail if this assumption is not satisfied.

Keywords

Novelty detection, Hierarchical classification, SVMs, One Class SVMs.

Categories and Subject Descriptors

I.5.2 [**Pattern Recognition**]: Design Methodology; H.3.1 [**Information Systems Applications**]: Content Analysis and Indexing; I.4.8 [**Object recognition**]:

General Terms

Algorithms

ICMR '14 Apr 01-04 2014, Glasgow, United Kingdom

1. INTRODUCTION

Content-based image retrieval systems must not only be able to compare pairs of images efficiently, but also identify cues from the query images, such as their semantic attributes or categories [Douze et al., 2011, Wang et al., 2012]. There has been a significant advance in image categorisation methods: they currently report impressive results on large and realistic data sets [Vedaldi and Zisserman, 2012, Duchenne et al., 2011, Chatfield et al., 2011, Everingham et al., 2010]. However, they rely on a pre-defined set of classes that are learnt during training in order to identify attributes and categories. In real life big data problems, one cannot expect that labels are available for all possible classes during training. The ability to automatically detect and model novel classes is crucial for scalability of these methods. Such ability is not present in current image retrieval systems. We therefore aim to learn fine grained classifiers that are able to detect a novel object category from a single sample.

Humans can generalize concepts from known objects, and new experiences are compared with and differentiated from the seen instances in an existing categorisation framework [Eysenck and Keane, 2005]. In other words, with a single or a few examples, we are able to process a new sample and infer if it belongs to a novel category [Biederman, 1987], whereas machine learning algorithms usually need a large set of training samples to learn a classification model. Inspired by human perceptual and reasoning abilities, and considering that training instances could not be given for new classes of objects, we aim to solve the first aspect of the model updating problem: the detection of novel categories.

Assuming that object categories form a semantic hierarchy where similar categories share the same parent node, we aim at developing techniques that distinguish instances of categories placed outside the current taxonomy, e.g. that create a new internal node in the taxonomy tree, or correspond to unseen objects. The particular focus is on detecting subcategories that belong to a known super-category but were not specialised during training, i.e. classes that originate a new leaf in the tree. In this paper, we investigate different classifier architectures and their associated mechanisms for novel class detection.

Our investigations build up from the method proposed in [Weinshall et al., 2012] and use a hierarchy of classifiers. An incongruence detection method measures disagreement be-

tween classifier outputs at different levels of the hierarchy. Additionally, we propose two alternative methods. The first is based on a flat classifier structure where every concept or its subgroup is considered as a separate class, and novelty is flagged when none of the classifiers detect a positive stimulus. The second is a hybrid between the hierarchical method and the flat method. It uses a general classifier that reduce the number of candidate subclasses and, at the specific level, the novel subclass detectors are based on the flat classification structure. We also evaluate structures that use one-class classifiers.

We initially confine the investigation to a group of concepts studied in [Weinshall et al., 2012], which contains images of different types of motorbikes, and demonstrate the feasibility of the methods and the relative advantages of the classifier architectures investigated here. While the hierarchical structure using an incongruence measure offers better computational efficiency and a better understanding of nuances of novelty, the flat structure exhibits a slightly better detection performance. Most importantly, the results obtained for an extension of the experiments to a larger taxonomy of objects show that the hierarchical approach breaks down when the semantic object categorisation does not map onto a corresponding visual similarity hierarchy. Even if the hybrid approach we introduced leads to a significant improvement in the recognition rate of unknown subcategories, the results suggest that the incongruence-based method advocated in [Weinshall et al., 2012] should be used mainly for taxonomies of concepts closely coupled with the taxonomy of visual appearance. Such taxonomies could be created e.g. by visual appearance clustering.

The rest of the paper is organised as follows. The next section provides a brief review of the related literature. Section 3 defines the problem and describes the classification schemes used in this paper. Section 4 further details the experimental set up and the datasets that we used. Finally Section 5 shows all the experiments we have conducted and analyses the results obtained. The paper then concludes in Section 6.

2. RELATED WORK

Recently, there has been an increased interest in novelty detection, i.e., the ability to detect if new data is of a type (class, model or domain) that has not been seen during training. In [Markou and Singh, 2003a, Markou and Singh, 2003b], comprehensive surveys are offered on novelty detection with the main distinction being made between statistically based approaches and neural network approaches. The reviews also identify various application domains where novelty detection is important. In Computer Vision, novelty detection has recently been approached by [Lampert et al., 2009] and [Weinshall et al., 2012], among others. [Lampert et al., 2009] focuses on detecting unseen categories of objects by using attributes. In order to make predictions about classes with no training data, they learn a representation that goes beyond the class boundaries merging images of the object classes that are characterized by the same attribute. Our approach shares the idea that the knowledge about unseen classes should come to related known classes but we assume that this expertise comes from the global representation of the images, and not from a disjoint set of attributes.

On the other hand, [Weinshall et al., 2012] demonstrate how in a hierarchically organised object class taxonomy, novelty can be identified in terms of disagreement between two classifiers making decisions at different levels of the hierarchy. In particular, a novel object is defined as an input whose probability of belonging to a parent class (general concept) is high but at the same time the probability of membership in any known specific (child) class is low. Despite demonstrating how this framework can be applied to several domains, ranging from detecting novel classes of visual and audio objects, through out-of-vocabulary word detection, to detecting novel patterns of motion, the experiments presented are at proof of concept level.

A broader notion of novelty is anomaly [Chandola et al., 2009], which refers to the problem of finding patterns in data that do not conform to expected behaviour. [Kittler et al., 2014] proposed a taxonomy of anomalies, that include outlier detection, novel class detection and domain change detection. A direct solution to the problem of detecting anomalies is to determine a region in the observation space representing the normal behaviour and classify any object that lies outside this area as an outlier or anomaly. Many approaches propose to identify anomalies using generative methods in a statistical framework [Almajai et al., 2012, Deng et al., 2012, Rodner et al., 2011, Pauwels and Ambekar, 2011]. Although such solutions might be appealing in low dimensional spaces, the problem is very challenging in other (more common) situations.

The problem of novelty detection and modelling new classes also relates to that of zero shot learning, which refers to the ability to recognise classes that were not seen during training, see [Rohrbach et al., 2011] and references therein. Our work can be viewed as an extension of the concept of zero shot learning to a taxonomy of different categories. While zero shot learning aims at modelling a new class, our approach has the ability to (i) first of all, identify if the novel sample belongs to a novel (sub-)class and (ii) define the location of this novel (sub-)class with respect to a known taxonomy, i.e., it indicates how to modify a class hierarchy to accommodate for this new (sub-)class.

3. METHODOLOGY

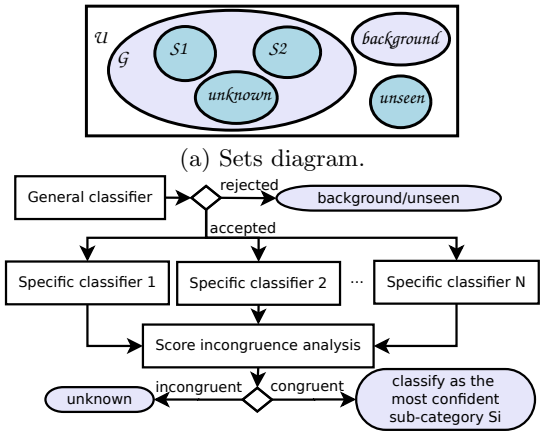
Starting from the assumption that an object from an unknown class belongs to a sibling class of known categories, [Weinshall et al., 2012] define the incongruent or novel event in relation to partial order on a set of classes. The partial order can be represented by a directed graph and subset-superset relations in the graph can be modelled as conjunctive and disjunctive hierarchies. More precisely a conjunctive hierarchy models part-of membership, e.g. head, legs and tail combine to form a dog and can be considered as more general concept than the dog. A disjunctive hierarchy models class membership, where an object can be defined at different levels of generality, e.g. Beagle and Collie are specific concepts of dog. In this paper, we explore disjunctive hierarchies for object classification, where semantically related subcategories of images share the same parent node.

3.1 Problem definition

Following section 3.1 of [Weinshall et al., 2012], we define a general concept $\mathcal{G}$ as a superset of more specific concepts $\{\mathcal{S}_i\}$. However, the union of all the known specific concepts does not form the complete set that comprises all concepts, i.e., $\cup_i\{\mathcal{S}_i\} \subset \mathcal{G}$ (rather than $\cup_i\{\mathcal{S}_i\} = \mathcal{G}$). We also as-

sume that during training, samples are given from the set of known subcategories $\cup_i\{\mathcal{S}_i\}$ and also from a small set of a background class that does not belong to $\mathcal{G}$.

As illustrated in Fig. 1, the input space of the algorithm is therefore defined by the union of the disjoint sets $\mathcal{S}_i$, while the output space is represented by the three possible classification results *Known, Unknown, Background*. In details:



(a) Sets diagram.

(b) Classification flowchart.

Figure 1: Class types shown in a sets diagram (a) and a flowchart that summarises the incongruence detection method for disjunctive hierarchies, proposed in [Weinshall et al., 2012].

- *Known*: samples that belong to the set of subcategories $\cup_i\{\mathcal{S}_i\}$ that are known from the training set.
- *Unknown*: samples that belong to $\mathcal{G}\setminus\cup_i\{\mathcal{S}_i\}$, i.e., they are from a known general category but do not belong to any of the subcategories that were known during training.
- *Background*: samples that are rejected by the general level classifier, i.e., they do not belong to the general concept $\mathcal{G}$ and are detected because the general classifier used background samples at training (they belong to $\mathcal{U}\setminus\mathcal{G}$).

Background samples are collected using images that clearly do not belong to the known general class, such as background regions of images or textures. They certainly do not cover the infinite set of possible object classes that do not belong to the known general category $\mathcal{G}$. Therefore, further to testing with a test split of the *Background* set we also test with another set of samples collected from other foreground object classes that do not belong to $\mathcal{G}$. This set is labelled as the *Unseen* set, as it is not similar to any object category seen at training. Therefore, even though the method produces 3 types of labels (*Known*, *Unknown* subcategory and *Background*), the test set samples have four types of labels (the above plus *Unseen*). Following [Weinshall et al., 2012], we consider that if *Unseen* samples are classified as *Background*, the method has succeeded.

3.2 Classification schemes

3.2.1 Incongruence detection on disjunctive hierarchies with binary SVMs (B-SVMs)

Weinshall et al. proposed to identify novel classes or subclasses of images using the incongruence between classifiers at different levels of a hierarchy. Let $M^{\mathcal{G}}$ be the model learnt on a general concept using samples from $\cup_i\{\mathcal{S}_i\}$ and $M_i^{\mathcal{S}}$ the models learnt on specific concepts or subcategories. The detection of a novel category is based on the disagreement between the predictions of the different models. In other words, a sample is identified as novel when is accepted by $M^{\mathcal{G}}$, but rejected by all $M_i^{\mathcal{S}}$. Conversely, a sample belonging to one of the known categories is accepted both by $M^{\mathcal{G}}$ and one of $M_i^{\mathcal{S}}$, as illustrated in Figure 1(b).

At the specific level, a decision score $V_i(\mathbf{x})$ is obtained for each sample $\mathbf{x}$ and for each learnt model $M_i^{\mathcal{S}}$. The binary-SVMs method (B-SVMs) uses SVMs for classification at all levels in a one-against-all scheme. Since SVMs are discriminative, [Weinshall et al., 2012] propose to whiten the classification scores as follows:

$$S_i(\mathbf{x}) = \frac{V_i(\mathbf{x}) - V_i^w}{V_i^c - V_i^w} \ , \tag{1}$$

where V_i^c is the average confidence of train or validation examples classified **correctly** using $M_i^{\mathcal{S}}$ and V_i^w is the same for examples classified **wrongly** using $M_i^{\mathcal{S}}$.

Weinshall et al. rely on the assumption that sibling classes semantically grouped in the same super class also have similar feature vectors. This theory is generally accepted and exploited in hierarchical image classification methods, and can also be exploited in the context of novelty detection for classes that were not seen during training. Later in this paper, we will demonstrate that this assumption is not sufficient when a wider taxonomy of images is considered and the visual hierarchy is not trivial.

3.2.2 One-class SVMs (OC-SVMs)

We propose to use the same architecture as Sec. 3.2.1 (Fig. 1(b)), but replacing binary SVMs by OC-SVMs. OC-SVMs [Schölkopf et al., 2001] are usually exploited in the context of outlier detection when only positive training samples are given. They aim to find the hypersphere that best encloses the training data, differently from common binary SVMs that try to find the hyperplane that best separates two training classes, i.e., they are designed for outlier detection. By setting the parameter ν, OC-SVMs can be properly tuned to recognise a fraction of the training samples as outlier and allow for errors and uncertainty in the training set, so there is no need to use (1) to normalise the scores. Similar to common binary SVMs, OC-SVMs can be used in their dual formulation.

3.2.3 B/OC-SVMs

We also evaluated a hybrid combination of binary and one-class SVMs in which a binary SVM was used as the general classifier for $\mathcal{G}$ and OC-SVMs are used as specific subcategory classifiers for $\mathcal{S}_i$. The motivation for this combination is that both positive and negative training samples are given at the general level (i.e. $\mathcal{G}$ and *Background* samples), but for each of the specific level classifiers, only positive samples are given for training ($\mathcal{S}_i$). *Unknown* samples are not given.

3.2.4 Flat model

In contrast to the previous approaches, the class hierarchy is not explored by this method. Instead, it treats the novel subclass as a category of objects that differs from all the

known subclasses *and the background class*. In a problem with N subcategories (regardless of the number of super-categories), a set of $N+1$ one-vs-all binary SVM classifiers is trained: one for each known subcategory and one for the background class. A new object is classified as novel if it is rejected by all the $N+1$ classifiers. Having N subcategories plus the background category makes the normalisation in (1) unnecessary.

3.2.5 B-SVMs/Flat model

This configuration of classifiers combines B-SVMs and Flat models. The model $M^{\mathcal{G}}$ is learnt on a general concept in the same way as Sec. 3.2.1, using samples from $\cup_i\{\mathcal{S}_i\}$. The specific models $M^{\mathcal{S}_i}$ are learnt using instances from $\mathcal{S}_{\hat{i}}$ as positive samples and, for the negative class, all the subcategories $\mathcal{S}_{i,i\neq\hat{i}} \cup \mathcal{S}_j$ and the *background* category, i.e. samples from all the N subcategories that differ from the current one, as explained in Fig. 2(b). This is in contrast to the configuration of 3.2.1, where only samples from $\mathcal{S}_{i,i\neq\hat{i}}$ were used as negative training instances, Fig. 2(a).

The aim of this scheme is to benefit from the advantages of the hierarchical configuration, which reduces the number of candidate subclasses to evaluate for each sample, and to benefit from the classification performance of the Flat structure, which is able to learn a better decision boundary.

4. EXPERIMENTAL SET UP

4.1 Image Representation

In order to build a vectorial representation $\mathbf{x}$ of each image, we used a method that has proven to be state-of-the-art in the benchmark presented in [Chatfield et al., 2011]. Images are represented using Pyramid Histograms Of visual Words (PHOW – based on [Lazebnik et al., 2006]), encoded with Fisher Vectors [Perronnin and Dance, 2007, Perronnin et al., 2010]. More specifically, SIFT descriptors are computed on a dense grid at four different scales defined by setting the width of the spatial bins of SIFT to 4, 6, 8 and 10 pixels. PCA is performed on the obtained local features and the dimensionality is reduced to 80 components.

Fisher Vectors (FV) are built by concatenating Gaussian gradient vectors $\mathbf{x} = [\cdots, \mathcal{F}^F_{\mu,i}, \mathcal{F}^F_{\sigma,i}, \cdots]$ w.r.t. mean μ_i and standard deviation σ_i (the variables are assumed independent), for each Gaussian i in a GMM that models all training features $\mathbf{f}$, where

$$\mathcal{F}^F_{\mu,i} = \frac{1}{T\sqrt{\omega_i}} \sum_{t=1}^{T} \gamma_t(i) \left(\frac{\mathbf{f}_t - \mu_i}{\sigma_i} \right) \tag{2}$$

and

$$\mathcal{F}^F_{\sigma,i} = \frac{1}{T\sqrt{2\omega_i}} \sum_{t=1}^{T} \gamma_t(i) \left[\frac{(\mathbf{f}_t - \mu_i)^2}{\sigma_i^2} - 1 \right] \tag{3}$$

where $\gamma_t(i)$ represents the soft assignment of the descriptor of patch $\mathbf{f}_t$ to the Gaussian i and F is the set of T descriptors $\mathbf{f}_i$ of an image region.

This is done in each region of the spatial pyramid, which was set up combining regions in this configuration: 1×1, 2×2 and 3×1 and the FVs of each of these regions are concatenated for each image. This results in a vectorial representation $\mathbf{x}$ of $D = M \times 2G \times R$ dimensions per image, where $M = 80$ is the local feature dimensionality (after PCA), $G = 256$ is number of Gaussians in the mixture and $R = 8$ is the number of pyramid regions.

For the above, we used the implementation publicly available in the VLFeat toolbox [Vedaldi and Fulkerson, 2008].

4.2 Kernels

In all the experiments we used the Hellinger (or Bhattacharyya) kernel, which is an additive kernel. [Vedaldi and Zisserman, 2012] state how additive kernels usually yield classification results similar to non-linear kernels while being at the same time efficient to compute. Additive kernels are in the form $K(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{D} k(x_i, y_i)$ where k is itself a kernel (and $\mathbf{x},\mathbf{y} \in \mathbb{R}^D$). In particular the Hellinger kernel can be computed for non-negative vectors as $k(x,y) = \sqrt{xy}$. This can be easily extended to arbitrary vectors: $k'(x,y) = sign(xy)k(|x|,|y|)$. The interesting advantage of these kernels is to allow to perform an explicit embedding of the data and then learn a linear classifier in the new space. For example for the Hellinger kernel we can define a feature map as $\varphi(x_i) = \sqrt{x_i}$ and then

$$K(\mathbf{x},\mathbf{y}) = \langle \varphi(\mathbf{x}), \varphi(\mathbf{y}) \rangle = \sum_{i=1}^{D} \sqrt{x_i}\sqrt{y_i} = \sum_{i=1}^{D} \sqrt{x_i y_i}. \tag{4}$$

4.3 Datasets

Caltech256 - Motorbikes.

The first evaluation setting is based on a small sub-hierarchy of the Caltech256 dataset [Griffin et al., 2007], which was used in the novelty detection experiments of [Weinshall et al., 2012]. The category *Motorbikes* is chosen as the general concept and the hierarchy is represented by the three more specific subclasses: *Cross*, *Road* and *Sport*. Finally the *Clutter* class images are used as negative examples for the general level classifier and twenty two object classes (different from *Motorbikes*) are sampled to serve as *Unseen* objects. Fig. 3 shows the structure of the taxonomy.

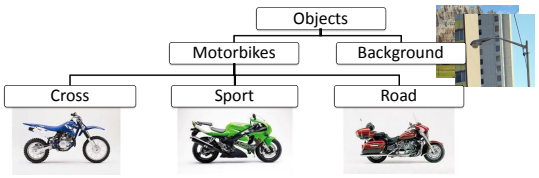
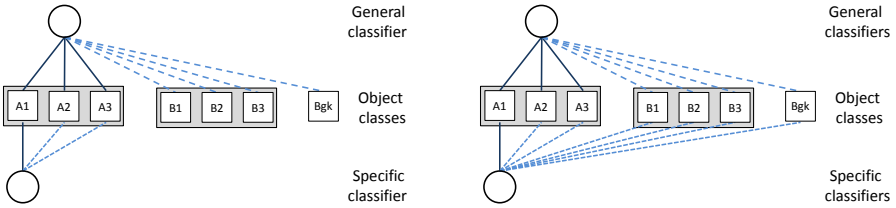


Figure 3: Samples of the *Caltech256 - Motorbikes* dataset in the taxonomy of [Weinshall et al., 2012]. For information about the copyright of the images shown in the panels, please contact the authors of [Griffin et al., 2007].

Caltech256 - Transportation.

In addition, we evaluate a more extensive hierarchy of images using the transportation hierarchy in Caltech256, and specifically *Air* and *Ground transportation*, Fig. 4(a). These two super-categories are respectively divided in *Blimps*, *Fighter Jets*, *Helicopters*, *Airplanes* and *Fire Trucks*, *Motorbikes*, *Car Sides*, *School Bus*. As in the *Motorbikes* dataset, the *Clutter* category is used as negative class at the general level and a set of samples of twenty-two different classes are used as *Unseen* samples.



(a) B-SVMs classification scheme

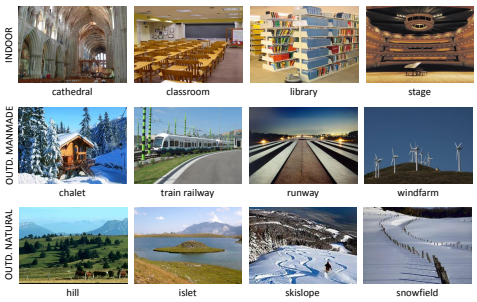
(b) B-SVMs/Flat classification scheme

Figure 2: The B-SVMs and B-SVMs/FLAT schemes. Object classes belonging to the same general category are grouped in gray boxes. Connections from classifier to categories with continuous lines indicate that the object category is used to provide positive training samples, connections with dotted lines indicate negative training samples.

SUN397.

In order to further evaluate the proposed framework we used a subset of the SUN397 Scene Categorization dataset [Xiao et al., 2010]. This dataset has a hierarchical division of the scenes in *Indoor*, *Outdoor natural*, *Outdoor man-made*, Fig. 4(b). We sampled four more specific classes for each one of the three super-categories. Specifically Indoor scenes are divided in *Cathedral*, *Classroom*, *Library* and *Stage*; Outdoor natural scenes are divided in *Hill*, *Islet*, *Skislope* and *Snowfield* and finally Outdoor man-made are divided in *Chalet*, *Train railway*, *Runway* and *Windfarm*. Differently from Caltech256, this dataset does not contain a category that can be used as negative example for the general level classifier, e.g. the *Clutter* category. For this reason, and because the chosen taxonomy covers all the three super categories of which the dataset is composed, we decided to focus only on the detection on *Novel* subcategories disregarding the *Unseen* classes.

(a) Caltech 256 - Transportation.
For information about the copyright of these images, please contact the authors of [Griffin et al., 2007].



(b) SUN.
For information about the copyright of these images, please contact the authors of [Xiao et al., 2010].

Figure 4: Sample images of the taxonomies from *Caltech256-Transportation* and *SUN*.

5. RESULTS

Using the two datasets explained in Sec. 4.3 the experiments were repeated with a leave-one-class-out approach on the subcategories to simulate the novel class. The training data of the general level classifier consists of the combination of the known subcategories for positive samples and the clutter class for negative samples. The specific level classifiers were trained using, as positive samples, objects from one of the subclasses used to train the general level SVM and, as negative samples, objects from the other subclasses depending on the approach used. For each subcategory, 39 images were chosen randomly for training and 20 for testing, as done in [Weinshall et al., 2012]. The experiments were repeated 25 times, sampling different train and test samples.

For classification, we used the open source *LibSvm* library with parameters optimised using cross-validation. Table 1 shows the average detection scores obtained on the three datasets for each classification scheme and compares our results with those of [Weinshall et al., 2012] on the motorbikes dataset. For the evaluation of the **B-SVMs/Flat** model, since the positive and negative classes were unbalanced, we used a SVM implementation with weighted cost functions, i.e. the cost parameters C were set to $w_+ \times C$ and $w_- \times C$, with $w_+ \neq w_-$ for positive and negative training samples.

Table 1: Correct detection rates for Known subcategories, Novel subcategories and Unseen classes. The first row (B-SVMs*) shows results from [Weinshall et al., 2012] and the remaining rows show our results.

Data	Method	Known	Subcat.	Unseen
Motorbikes	***B-SVMs****	0.57	0.71	0.74
	B-SVMs	0.73	0.95	0.95
	OC-SVMs	0.67	0.49	0.64
	B/OC-SVMs	0.71	0.68	0.97
	FLAT	0.84	0.86	0.95
Trns.	***B-SVMs***	0.66	0.20	0.40
	B-SVMs/FLAT	0.67	0.39	0.62
SUN	***B-SVMs***	0.65	0.46	-
	B-SVMs/FLAT	0.68	0.57	-

5.1 Experiments on Caltech256 - Motorbikes

We exploited this hierarchy to evaluate all the classification schemes detailed in Sec. 3. As expected, our implementation of **B-SVMs** gave significantly better results than

[Weinshall et al., 2012] thanks to the better image representation we adopted[1]. This scheme yields good novelty detection rates, as shown in Fig. 5(a), but has the main drawback of strongly relying on a threshold on the score values normalized with Eq. (1). In [Weinshall et al., 2012], the authors fixed this threshold to 0.5. Here we used 0. This threshold directly controls the number of elements classified as *Known* or *Unknown*: with a value of the threshold near 0 almost all unknown objects are classified correctly but also a relatively high percentage of known objects is classified as unknown, while moving the threshold to 0.5 the effect is the opposite.

Despite **OC-SVMs** being theoretically well suited for outlier detection, our experiments demonstrate their limitations in this context, where the data points lie in a high dimensional space. Its detection rate was certainly better than random for each category (known/unknown/unseen) but the overall performance is significantly lower than other approaches, as shown in Fig. 5(b).

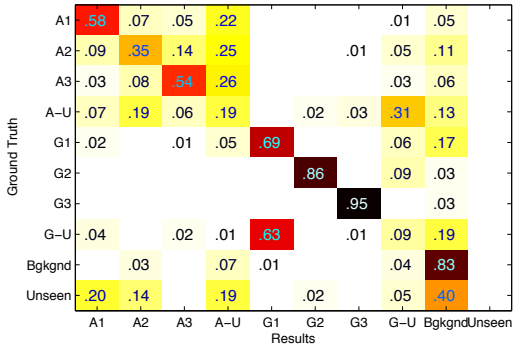
Fig. 5(c) shows the average classification rates for the hybrid scheme **B/OC-SVMs**. The performance in this case was worse than the **B-SVMs** configuration, but this method has the advantage of not requiring score normalisation. Finally Fig. 5(d) shows the results obtained with the **Flat** model. It can be observed that these results are similar to the ones obtained with the **B-SVMs** scheme. There is an improvement in the detection rate of the *known* subcategories, which is relevant when it is preferable to have a lower number of misclassified known objects. One disadvantage is that this scheme can not be exploited in larger hierarchies because novel objects can only be identified if they are rejected by all classifiers. It is therefore unable to detect subclasses belonging to different super-categories. The **B-SVMs/Flat** model has not been evaluated in this set of experiments because dealing with only one general category reduces this scheme to the initial **B-SVMs** approach.

5.2 Experiments on Caltech256 - Transportation

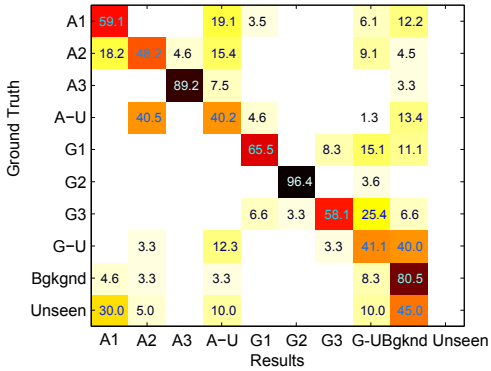
We explored the possibility of extending this framework to more complex hierarchies of images using the taxonomy **Caltech 256 - Transportation**. Based on the results discussed in the previous section, we decided to restrict these experiments to using the **B-SVMs** scheme, which was the best performing hierarchical method in the previous experiment. We extended the evaluation to the **B-SVMs/Flat** model, to benefit from the Flat model at the specific level of classification and because the Flat method alone is unable to deal with multiple super-categories. Noting that the score normalisation of (1) makes the framework sensitive to the threshold, we decided not to use that normalisation in these experiments. In the previous settings with only two known subcategories, it was necessary to normalise the classifiers score, otherwise the two specific level SVMs would become the same classifier with swapped output signals, i.e., trained on opposite labels. When more than two subcategories are known, the normalisation of (1) becomes unnecessary (using the one-against-all setting) and does not produce any performance improvement.

The best results are obtained with the **B-SVMs/Flat** model. The confusion matrices in Fig. 6 show that most of the known sub-categories were correctly classified. The

[1]This is evident from Tab. 1 and by comparing Fig. 5(a) with Fig. 3 of [Weinshall et al., 2012].



(a) Novel Subcategories: Airplanes - School Bus



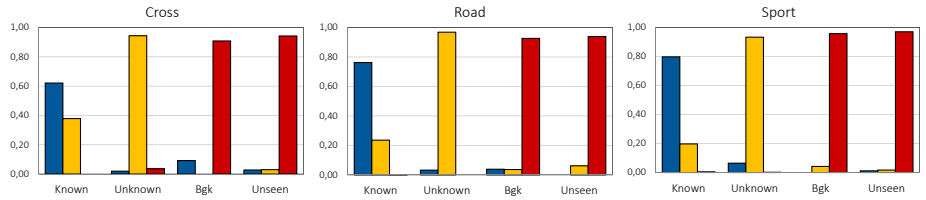
(b) Novel Subcategories: Helicopters - Car Sides

Figure 6: Confusion matrices (in %) obtained on Caltech 256 with the B-SVMs/Flat scheme by removing these subcategories from the training set (a) *Airplanes* and *School Bus*, (b) *Helicopters* and *Car Sides*. 'A' and 'G' indicates *Air* and *Ground* transportation super-category, respectively. '-U' indicates the unknown subcategory. Note that 'Unseen' is not a label in the training set and unseen samples are expected to be classified as background.

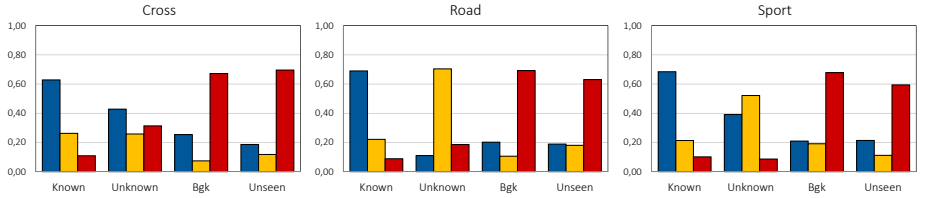
roughly block-diagonal structure of the matrices show that the majority of the samples were classified to the correct super-category. Most of the unseen samples were correctly detected as background. On the other hand, most of the mistakes were either false *background* detections or false *unknown* subcategory detections. The true positive rates for *unknown* (novel) subcategory are substantially improved with respect to the **B-SVMs** scheme, where the obtained rates were disappointing, as most of those samples were either misclassified as other sub-categories or as background. The gain of nearly 20% is quantified in Table 1. This shows that the incongruence-based method of [Weinshall et al., 2012] breaks down when the taxonomy of the concepts is not strictly related with the visual hierarchy (i.e. the structure in the feature space), while the approach we proposed is stronger and yields better results.

5.3 Experiments on SUN 397

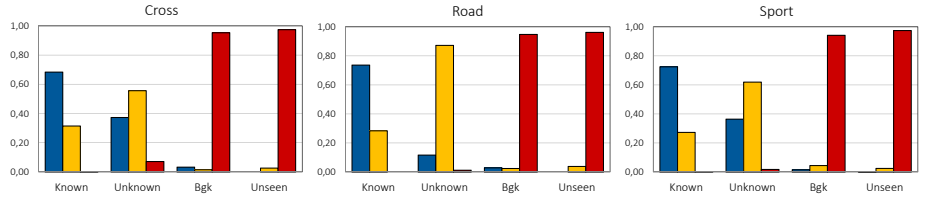
We finally evaluated our framework on a taxonomy built over the **SUN397** dataset for scene recognition. Similarly to the previous experiments we restricted the evaluation to the **B-SVMs** and **B-SVMs/Flat** schemes. Using the taxonomy described in Sec. 4.3 we iteratively sampled one of



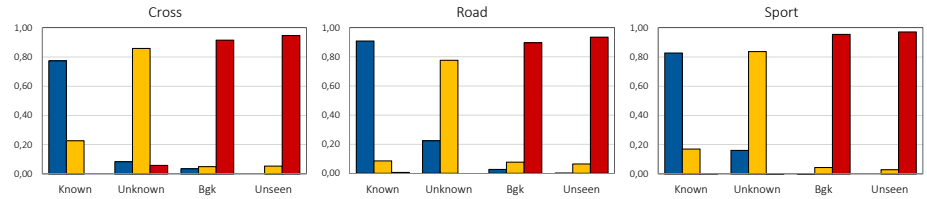
(a) B-SVMs configuration



(b) OC-SVMs configuration



(c) B/OC-SVMs configuration



(d) Flat configuration

Figure 5: Results leaving out one subcategory of motorbikes from the training set (Cross, Road and Sport, from left to right) for each classification scheme. The x-axis represents the ground truth subcategory type and the y-axis is detection rate. Blue, yellow and red bars correspond respectively to Known, Unknown, Background category type detection (see Sec. 3.1).

the four subcategories as *Unknown* and we used the other three to train the classifiers. The average results are shown in Table 1. As already mentioned in Sec. 4.3 in this case we limited our aim to the detection of novel subcategories without focusing on *unseen* classes. The detection rate on *known* categories is similar to the one obtained for Caltech256 - Motorbikes taxonomy, while the *unknown* detection rate is significantly better than the previous one. Also in this case the average values demonstrate that the proposed framework, despite the satisfactory results, needs improvements when the visual hierarchy is not trivial to avoid the misclassification of the novel samples.

6. CONCLUSIONS

We explored methods to detect novel categories and subcategories in hierarchies of images using a novelty detection framework which leverages the incongruence between classifiers at different levels of the hierarchy. We evaluated this framework on three datasets for object and scene classification, and using different classification schemes with Binary and One-Class SVMs in different parts of the hierarchy. We also evaluated a flat classification scheme, which only works in datasets with a unique super-category. Finally we evaluated a hybrid approach between the hierarchical and the flat. To our surprise, the results showed that binary SVMs outperform OC-SVMs for novelty detection. The hierarchical method achieves satisfactory novelty detection rates for small taxonomies and when the semantic hierarchy matches the appearance hierarchy of image classes, but breaks down when these assumptions are not satisfied. The hybrid approach benefits from the advantages of the flat method at the specific level and from the reduced number of candidate subcategories given by the hierarchical structure. It lead to our best results on Caltech256-Transportation and SUN dataset datasets. For future work, we intend to exploit these methods as staring point to built stronger models of the novel classes using a semi-supervised learning approach.

7. ACKNOWLEDGMENTS

The authors are grateful for the support of EPSRC grants EP/F069421/1, EP/F02827X/1 and EP/K014307/1 (UK).

8. REFERENCES

[Almajai et al., 2012] Almajai, I., Yan, F., deCampos, T., Khan, A., Christmas, W., Windridge, D., and Kittler, J. (2012). Anomaly detection and knowledge transfer in automatic sports video annotation. In *Studies in Computational Intelligence*. Springer.

[Biederman, 1987] Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Rev.*, 97:115–147.

[Chandola et al., 2009] Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3).

[Chatfield et al., 2011] Chatfield, K., Lempitsky, V., Vedaldi, A., and Zisserman, A. (2011). The devil is in the details: an evaluation of recent feature encoding methods. In *British Machine Vision Conference*.

[Deng et al., 2012] Deng, J., Krause, J., Berg, A. C., and Fei-Fei, L. (2012). Hedging your bets: Optimizing accuracy-specificity trade-offs in large scale visual recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[Douze et al., 2011] Douze, M., Ramisa, A., and Schmid, C. (2011). Combining attributes and fisher vectors for efficient image retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 745–752.

[Duchenne et al., 2011] Duchenne, O., Joulin, A., and Ponce, J. (2011). A graph-matching kernel for object categorization. In *IEEE International Conference on Computer Vision*.

[Everingham et al., 2010] Everingham, M., Gool, L. J. V., Williams, C. K. I., Winn, J. M., and Zisserman, A. (2010). The PASCAL visual object classes (VOC) challenge. *International Journal of Computer Vision*, 88(2).

[Eysenck and Keane, 2005] Eysenck, M. W. and Keane, M. T. (2005). *Cognitive psychology: a student's handbook*. Psychology Press.

[Griffin et al., 2007] Griffin, G., Holub, A. D., and Perona, P. (2007). The Caltech 256 object category data-set. Technical Report CNS-TR-2007-001, California Institute of Technology.

[Kittler et al., 2014] Kittler, J., Christmas, W., deCampos, T., Windridge, D., Yan, F., Illingworth, J., and Osman, M. (2014). Domain anomaly detection in machine perception: A system architecture and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[Lampert et al., 2009] Lampert, C. H., Nickisch, H., and Harmeling, S. (2009). Learning to detect unseen object classes by betweenclass attribute transfer. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[Lazebnik et al., 2006] Lazebnik, S., Schmid, C., and Ponce, J. (2006). Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2169–2178. IEEE Computer Society.

[Markou and Singh, 2003a] Markou, M. and Singh, S. (2003a). Novelty detection: A review - part 1: Statistical approaches. *Journal of Signal Processing*, 83.

[Markou and Singh, 2003b] Markou, M. and Singh, S. (2003b). Novelty detection: A review - part 2: Neural network based approaches. *Journal of Signal Processing*, 83.

[Pauwels and Ambekar, 2011] Pauwels, E. J. and Ambekar, O. (2011). One class classification for anomaly detection: Support vector data description revisited. In *Advances in Data Mining. Applications and Theoretical Aspects*, volume 6870 of *Lecture Notes in Computer Science*. Springer.

[Perronnin and Dance, 2007] Perronnin, F. and Dance, C. (2007). Fisher kernels on visual vocabularies for image categorization. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[Perronnin et al., 2010] Perronnin, F., Sanchez, J., and Mensink, T. (2010). Improving the Fisher kernel for large-scale image classification. In *IEEE European Conference on Computer Vision*.

[Rodner et al., 2011] Rodner, E., Wacker, E. S., Kemmler, M., and Denzler, J. (2011). One-class classification for anomaly detection in wire ropes with gaussian processes in a few lines of code. In *IAPR Conference on Machine Vision Applications*.

[Rohrbach et al., 2011] Rohrbach, M., Stark, M., and Schiele, B. (2011). Evaluating knowledge transfer and zero-shot learning in a large-scale setting. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[Schölkopf et al., 2001] Schölkopf, B., Platt, J., Shawe-Taylor, J., Smola, A., and Williamson, R. (2001). Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471.

[Vedaldi and Fulkerson, 2008] Vedaldi, A. and Fulkerson, B. (2008). VLFeat: An open and portable library of computer vision algorithms.

[Vedaldi and Zisserman, 2012] Vedaldi, A. and Zisserman, A. (2012). Efficient additive kernels via explicit feature maps. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(3).

[Wang et al., 2012] Wang, G., Hoiem, D., and Forsyth, D. (2012). Learning image similarity from Flickr groups using fast kernel machines. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2177–2188.

[Weinshall et al., 2012] Weinshall, D., Zweig, A., Hermansky, H., Kombrink, S., Ohl, F. W., Anemuller, J., Bach, J. H., Gool, L. V., Nater, F., Pajdla, T., Havlena, M., and Pavel, M. (2012). Beyond novelty detection: Incongruent events, when general and specific classifiers disagree. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(10).

[Xiao et al., 2010] Xiao, J., Hays, J., Ehinger, K., Oliva, A., and Torralba, A. (2010). Sun database: Large-scale scene recognition from abbey to zoo. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3485–3492.