Learning to Rank Using Privileged Information

Viktoriia Sharmanska
IST Austria
Klosterneuburg, Austria
viktoriia.sharmanska@ist.ac.at

Novi Quadrianto
University of Cambridge
Cambridge, UK
novi.quadrianto@gmail.com

Christoph H. Lampert
IST Austria
Klosterneuburg, Austria
chl@ist.ac.at

Abstract

Many computer vision problems have an asymmetric distribution of information between training and test time. In this work, we study the case where we are given additional information about the training data, which however will not be available at test time. This situation is called learning using privileged information (LUPI). *We introduce two maximum-margin techniques that are able to make use of this additional source of information, and we show that the framework is applicable to several scenarios that have been studied in computer vision before. Experiments with* attributes, bounding boxes, image tags *and* rationales *as additional information in object classification show promising results.*

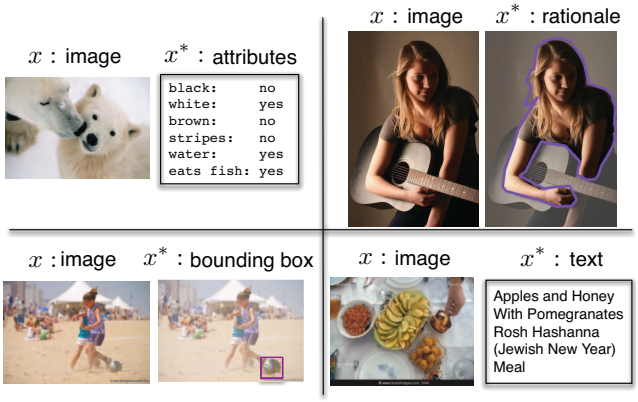

Figure 1: Four different forms of *privileged* information that can help learning better object recognition systems: *attributes*, *annotator rationales*, *object bounding boxes*, and *textual descriptions*.

1. Introduction

In this work we study the problem of *learning using privileged information (LUPI)*, as it was formally introduced by Vapnik in [25]. To learn with privileged information means that for a learning task, e.g. object categorization, one has access not only to input/output training pairs of the task we want to learn, but also to additional information about the training examples. Typically this additional data is more informative about the task at hand than the training data alone, so one would like to use it for better prediction. However, it is not clear how to do so, since at test time there will be no such data source. A possible analogy is human learning with a teacher: when a student learn a concept in school, for example algebra, the teacher can provide additional explanations at any time. Hopefully this will make the student learn faster than if the teacher would only pose questions and give their answers. However, when later in life the student faces an algebra problem, he or she will not be able to rely on the teacher's expertise anymore.

In this work we demonstrate the relevancy of this observation in a variety of computer vision scenarios: we explore four different types of privileged information in the context of object classification: *attributes* that describe semantic properties of an object instead of just visual ones, *bounding boxes* that specify the exact localization of the target object in an image, *image tags* that describe the context of an image in textual form, and *annotator rationales* that provide additional information *why* a training example was annotated the way it was.

Figure 1 illustrates these four modalities. All of them have been studied previously in the computer vision literature, see our discussion in Section 2. However, in each case a separate method was designed to handle the specific additional source of information. One of our contributions in this work is to show that it is possible to handle all these situations in a unified framework: LUPI.

Approach and contribution. At first sight, it is not clear how a data modality that is not available at test time would be useful for classification at all: for example, training a classifier on the privileged data is useless, since there is no way to evaluate the resulting classifier on the test data. LUPI therefore requires an additional step of *information transfer* from the privileged to the original data modality.

At the core of our work in this paper lies the insight that *privileged information allows us to distinguish between* easy *and* hard *examples in the training set*[1]. Assuming that the privileged data is similarly informative about the problem at hand as the original data, one can presume that examples that are easy or hard with respect to the privileged information will also be easy or hard with respect to the original data. Thereby we have gained additional knowledge about the learning problem, and this can guide the training of an image-based predictor to a better solution.

We formalize the above observation in Section 3, where we also introduce two maximum-margin learning techniques for LUPI. The first, SVM+, works in a classification setting and was originally described by Vapnik [25]. The second, *Rank Transfer*, is a new contribution, which targets a ranking setup. In Section 4, we report on experiments in the four privileged information scenarios introduced earlier, and we end with conclusions in Section 5.

2. Related work

In computer vision problems it is common to have access to multiple sources of information. Sometimes all of them are visual, such as when images are represented by color features as well as by texture features. Sometimes, the modalities are mixed, such as for images with text captions. If all modalities are present both at training and at test time, it is rather straight-forward to combine them for better prediction performance. This is studied, e.g., in the fields of *multi-modal* or *multi-view* learning. Methods suggested here range from *stacking*, where one simply concatenates the feature vectors of all data modalities, to complex adaptive methods for early or late data fusions [23], including *multiple kernel learning* [26] and *LP-*β [10].

Situations with an asymmetric distribution of information have also been explored. In *weakly supervised* learning, the annotation available at training time is less detailed than the output one wants to predict. This situation occurs, e.g., when trying to learn an *image segmentation* system using only per-image or bounding box annotation [13]. In *multiple instance* learning, training labels are given not for individual examples, but collectively for groups of examples [16]. The inverse situation also occurs: for example in the PASCAL object recognition challenge, it has become a standard technique to incorporate strong annotation in the form of bounding boxes or per-pixel segmentations, even when the goal is just per-image object categorization [8].

The situation we are interested in occurs when at training time we have an additional data representation compared to test time. Different settings of this kind have appeared in the computer vision literature, but each was studied in a separate way. For clustering with multiple image modalities, it has been proposed to use CCA to learn a shared representation that can be computed from either of representations [3]. Similarly the shared representation is also used for cross-modal retrieval [20]. Alternatively, one can use the training data to learn a mapping from the image to the privileged modality and use this predictor to fill in the values missing at test time [5]. Feature vectors made out of semantic attributes have been used to improve object categorization when very few or no training examples are available [14, 27]. Recently *annotator rationales* [28] have been introduced to the computer vision community. In [7] it was shown that these can act as additional sources of information during training, as long as the rationales can be expressed in the same data representation as the original data (e.g. characteristic regions within the training images).

Our work follows a different route than the above approaches. We are not looking for task-specific solutions applicable to a specific form of privileged information. Instead, we aim for a generic method that is applicable to any form of privilege information that is given as additional representations of the training data. We show in the following sections that such frameworks do indeed exist, and in Section 4 we illustrate that the individual situations described above can naturally be expressed in these frameworks.

3. Learning using Privileged Information

We assume a situation of supervised binary classification: given a set of N training examples, represented by feature vectors $X = \{x_1, \ldots, x_N\} \subset \mathcal{X} \subset \mathbb{R}^d$, and their label annotation, $Y = \{y_1, \ldots, y_N\} \in \mathcal{Y} = \{+1, -1\}$, the task is to learn a prediction function $f : \mathcal{X} \to \mathbb{R}$ from a space $\mathcal{F}$ of possible functions, e.g. all linear classifiers. In the following, we will think of the examples as images and of their representation as computed from the image content, for example *bag-of-visual-words* histograms [6].

Adopting the LUPI setting, we are given additional information about the training set, which we assume also to be in the form of feature vectors, $X^* = \{x_1^*, \ldots, x_N^*\} \subset \mathcal{X}^* \subset \mathbb{R}^{d^*}$, where any x_i^* encodes the additional information we have about x_i. Note that we do not make further assumption about this *privileged* data. In particular, x_i^* might not be computable from the original image, but rather reflect a very different kind of information, such as explanation provided by a human teacher. Also, in general $\mathcal{X}^*$ will be different from $\mathcal{X}$, so is it not possible, e.g., to apply functions defined on $\mathcal{X}$ to $\mathcal{X}^*$ or vice versa.

The goal of LUPI is to use the privileged data, X^*, to learn a better classifier than one would learn without it. However, it is clear that $f : \mathcal{X} \to \mathbb{R}$ itself cannot rely on the $\mathcal{X}^*$ domain, since this is not available at test time. Therefore, it has to be our *choice* of $f \in \mathcal{F}$ that is influenced by the privileged data.

In this manuscript we rely on the intuition that the priv-

[1] One might also call them *typical* and *atypical*, or *inliers* and *outliers*.

ileged data helps us to distinguish between *easy* and *hard* examples in the training set. This knowledge allows us to identify the relevant aspects of the training data and concentrate the learning step towards these, thereby finding a function of higher prediction quality.

In the following, we explain two maximum-margin methods for learning with privileged information that fit to this interpretation. For simplicity of notation we write all problems in their primal form. Kernelizing and dualizing them is possible using standard techniques [21].

3.1. SVM+

A first model for learning with privileged information, SVM+, was proposed by Vapnik *et al*. [17, 25]. It is based on the insight that training a support vector machine (SVM) would be easier if one had access to a *slack oracle*. Ordinary SVM training is based on the following constrained objective function:

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i \tag{1}$$

subject to, for all $i = 1, \ldots, N$,

$$y_i[\langle w, x_i\rangle + b] \geq 1 - \xi_i \quad \text{and} \quad \xi_i \geq 0. \tag{2}$$

By minimizing over the classifier parameters w, b and the slack variables $\xi_1, \ldots, \xi_N$, we obtain the SVM solution. When the number of training examples increases, this is known to converge with a rate of $\frac{1}{\sqrt{N}}$ to the optimal classifier [24]. However, if we knew the optimal slack values ξ_i in advance, for example from an *oracle*, and perform the optimization with respect to w and b, then the convergence rate improves to $\frac{1}{N}$ [25]. Consequently, such an *OracleSVM* would require fewer training examples to reach a certain prediction accuracy than an ordinary SVM.

An intuitive interpretation of this is that the slack variables tell us which training examples are *easy* and which are *hard*. In the OracleSVM, the training process does not have to infer this from the data and can use all statistical information contained in the training examples to find the actual object of interest: the classifying hyperplane.

The idea of the SVM+ classifier is to use the privileged information as a proxy to the oracle. For this we parameterize $\xi_i = \langle w^*, x_i^*\rangle + b^*$ with unknown w^* and b^*, obtaining the SVM+ training problem:

$$\min_{\substack{w\in\mathbb{R}^d, b\in\mathbb{R} \\ w^*\in\mathbb{R}^{d^*}, b^*\in\mathbb{R}}} \frac{1}{2}\left(\|w\|^2 + \gamma\|w^*\|^2\right) + C\sum_{i=1}^{N}\langle w^*, x_i^*\rangle + b^* \tag{3}$$

subject to, for all $i = 1, \ldots, N$,

$$y_i[\langle w, x_i\rangle + b] \geq 1 - [\langle w^*, x_i^*\rangle + b^*] \tag{4}$$

$$\text{and} \qquad \langle w^*, x_i^*\rangle + b^* \geq 0. \tag{5}$$

Numerical optimization. The SVM+ optimization problem (3)/(4) is convex, but it cannot be solved by off-the-shelf SVM packages, because of the way the weight vectors interact. In [18], suitable sequential minimal optimization (SMO) algorithms were derived, one of which we use for our experiments (Section 4). As it is the case for the original SMO algorithm for SVM training [11], the numeric optimization works with the dual representation and is only applicable to problems with a small to medium size.

3.2. Rank Transfer

To overcome the limitations of the SVM+ setup we introduce a second method for making use of privileged information in this work. Again, the underlying idea is to identify easy and hard cases. However, instead of using the privileged data to identify easy-to-classify and hard-to-classify examples, we adopt a ranking setup and identify easy-to-separate and hard-to-separate example pairs.

Our formulation is based on the learning to rank framework [12], which requires solving the following optimization problem

$$\min_{w\in\mathbb{R}^d,\, \xi_{ij}\in\mathbb{R}} \quad \frac{1}{2}\|w\|^2 + C\sum_{i,j=1}^{N}\xi_{ij} \tag{6}$$

subject to, for all $i, j = 1, \ldots, N$, with $y_i > y_j$,

$$\langle w, x_{ij}\rangle \geq 1 - \xi_{ij} \quad \text{and} \quad \xi_{ij} \geq 0, \tag{7}$$

where $x_{ij} = x_i - x_j$. From the solution vector w we obtain ranking scores $f(x) = \langle w, x\rangle$ for new examples x.

The above formulation enforces a difference of at least 1 in ranking score between any pair of examples of different class label. However, it is intuitively clear that some example pairs will be easier to separate than others. Some example pairs might even be impossible to rank correctly in the given data representation. Following the same intuition as above, we hypothesize that knowing *a priori* which example pairs are easy and which are hard to separate and taking this into account during learning should improve the prediction performance.

This consideration leads us to the *Rank Transfer* method, summarized in Algorithm 1. We first train an ordinary ranking SVM on X^*. The resulting ranking function f^* we use to compute the margins achieved between any two training images[2], $\rho_{ij} := f^*(x_i^*) - f^*(x_j^*)$. Example pairs with a large values of ρ_{ij} can be considered easy to separate, whereas small or even negative values of ρ_{ij} indicate hard or even impossible to separate pairs. We then train a ranking SVM on X, aiming for a data-dependent margin ρ_{ij}

[2]Note that we deliberately evaluate the ranking function on the same data it was trained on. The reason is that the quantity we are interested in is how easy it is to separate two examples during training, not by how much one could expect two samples would be separated at test time.

Algorithm 1 Rank Transfer from $\mathcal{X}^*$ to $\mathcal{X}$

Input original data X, privileged data X^*, labels Y
$f^* \leftarrow$ ranking SVM (6)/(7) trained on (X^*, Y)
$\rho_{ij} = f^*(x_i^*) - f^*(x_j^*)$ *(between-sample margins)*
$f \leftarrow$ ranking SVM (8)/(9) trained on (X, Y) using ρ_{ij}
Return $f : \mathcal{X} \to \mathbb{R}$

between any two examples x_i and x_j rather than enforcing a constant margin of 1 between all pairs. The corresponding optimization problem is

$$\min_{w \in \mathbb{R}^d,\, \xi_{ij} \in \mathbb{R}} \quad \frac{1}{2} \|w\|^2 + C \sum_{i,j=1}^{N} \xi_{ij} \tag{8}$$

subject to, for all $i, j = 1, \ldots, N$, with $y_i \geq y_j$ and $\rho_{ij} > 0$,

$$\langle w, x_{ij} \rangle \geq \rho_{ij} - \xi_{ij} \quad \text{and} \quad \xi_{ij} \geq 0. \tag{9}$$

One can see that example pairs with small values of ρ_{ij} have more limited influence on w than in the ordinary ranking SVM. Incorrectly ranked pairs are even completely ignored. Our interpretation is that if it was not possible to correctly rank a pair in the privileged space, it will also be not possible to do so in the, presumably weaker, original space. Forcing the optimization to solve a hopeless tasks would only lead to overfitting and reduced ranking accuracy.

Numeric Optimization. Both learning steps in the *Rank Transfer* method, ranking on X^* and on X, are convex optimization problems. Furthermore, in contrast to SVM+, we can use standard SVM packages to solve them, including efficient methods working in primal representation [4], and solvers based on stochastic gradient descent [22].

For the ranking SVM on X^* this is clear, since the optimization problem (6)/(7) is identical to a binary SVM without bias term, trained on training examples x_{ij} that all have positive labels. For the ranking with data-dependent margin, we achieve the same by a reparameterization: we divide each constraint (9) by the corresponding ρ_{ij}, which is possible since only pairs with $\rho_{ij} > 0$ occur. Changing variables from x_{ij} to $\hat{x}_{ij} = \frac{x_{ij}}{\rho_{ij}}$ and from ξ_{ij} to $\hat{\xi}_{ij} = \frac{\xi_{ij}}{\rho_{ij}}$ we obtain the equivalent optimization problem

$$\min_{w \in \mathbb{R}^d,\, \hat{\xi}_{ij} \in \mathbb{R}} \quad \frac{1}{2} \|w\|^2 + C \sum_{i,j=1}^{N} \rho_{ij} \hat{\xi}_{ij} \tag{10}$$

subject to, for all $i, j = 1, \ldots, N$ with $y_i \geq y_j$ and $\rho_{ij} > 0$,

$$\langle w, \hat{x}_{ij} \rangle \geq 1 - \hat{\xi}_{ij} \quad \text{and} \quad \hat{\xi}_{ij} \geq 0. \tag{11}$$

This corresponds to an ordinary SVM optimization with training examples $\hat{x}_{ij} = \frac{x_i - x_j}{\rho_{ij}}$, where each slack variable has an individual weight $C\rho_{ij}$ in the objective. Many existing SVM packages support such per-sample weights, in our experiments we use LIBLINEAR [9]. Furthermore, in practice we only include example pairs with $\rho_{ij} > 0.1$, thereby preventing numeric instabilities and increasing computational efficiency.

4. Experiments

In our experimental setting we study four different types of privileged information, showing that all of these can be handled in a unified framework, where previously hand crafted methods were used. We consider attribute annotation, bounding box annotation, textual description and rationales as sources of privileged information if these are present at training time but not at test time. As we will see, some modalities are more suitable for transferring the rank than others. We will discuss this in the following subsections. Note that we also include results where the privileged information does not help. Besides scientific honesty, the reason for this is to show that *no negative transfer* occurs.

Methods. We analyze two methods of learning using privileged information: our proposed Rank Transfer method for transferring the rank, and the SVM+ method [18] kindly provided by the authors. We compare the results with ranking SVM and ordinary SVM when learning on the original space $\mathcal{X}$ directly (SVM rank and SVM baselines). We also provide as a reference the performance of SVM rank in the privileged space $\mathcal{X}^*$, as if we had the access to the privileged information during testing.

Evaluation metric. To evaluate the performance of the methods we use average precision (AP), which corresponds to the area under the precision-recall curve. In fact, we report percentage accuracy of AP score (0% to 100%).

Model selection. For the LUPI methods, we perform a joint cross validation model selection approach for choosing the regularization parameters in the original and privileged spaces. In the SVM+ method these are C and γ (3), and in the Rank Transfer these are C's in the two-stage procedure (6), (8). For the methods that do not use privileged information there is only a regularization parameter C to be cross validated. In the privileged space we select over 7 parameters $\{10^{-3}, \ldots, 10^{3}\}$. We use the same range in the original space if the data is L_2 normalized, and the range $\{10^{0}, \ldots, 10^{5}\}$ for L_1 normalized data. In all our experiments we use 5 fold cross-validation scheme, the best parameter (or pair of parameters) found is used to retrain the complete training set.

4.1. Attributes as privileged information

Attribute annotation incorporates high-level description of the semantic properties of different objects like shape, color, habitation forms etc. We use the *Animals with Attributes (AwA)* dataset [14]. We focus on the default 10 test

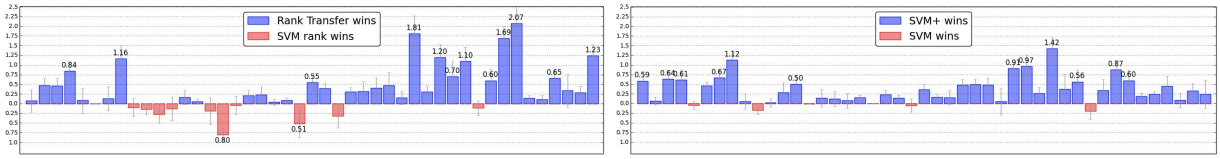

Figure 2: AwA dataset (attributes as privileged information). Pairwise comparison of the methods that utilize privileged information and their baseline counterparts is shown via difference of the AP performance (Rank Transfer versus SVM rank, SVM+ versus SVM). The length of the 45 bars corresponds to relative improvement of the average precision over 45 cases.

		SVM rank image	Rank Transfer image+attributes	SVM image	SVM+ image+attributes	Reference (SVM rank attributes)
1	Chimpanzee versus Giant panda	91.76 ± 0.35	91.83 ± 0.37	91.53 ± 0.36	$\mathbf{92.12 \pm 0.40}$	$\mathit{93.34 \pm 0.34}$
2	Chimpanzee versus Leopard	94.33 ± 0.35	$\mathbf{94.80 \pm 0.29}$	94.16 ± 0.35	94.23 ± 0.39	$\mathit{98.58 \pm 0.07}$
3	Chimpanzee versus Persian cat	91.39 ± 0.43	$\mathbf{91.86 \pm 0.38}$	91.09 ± 0.44	91.73 ± 0.38	$\mathit{96.94 \pm 0.24}$
4	Chimpanzee versus Pig	87.75 ± 0.36	$\mathbf{88.59 \pm 0.25}$	87.45 ± 0.33	88.06 ± 0.43	$\mathit{94.02 \pm 0.20}$
5	Chimpanzee versus Hippopotamus	87.49 ± 0.37	87.57 ± 0.42	$\mathbf{87.58 \pm 0.36}$	87.53 ± 0.36	$\mathit{95.67 \pm 0.17}$
6	Chimpanzee versus Humpback whale	98.52 ± 0.18	98.52 ± 0.15	98.12 ± 0.18	$\mathbf{98.57 \pm 0.16}$	$\mathit{99.94 \pm 0.00}$
7	Chimpanzee versus Raccoon	89.41 ± 0.35	89.54 ± 0.29	89.00 ± 0.38	$\mathbf{89.67 \pm 0.35}$	$\mathit{94.11 \pm 0.24}$
8	Chimpanzee versus Rat	87.31 ± 0.51	$\mathbf{88.47 \pm 0.45}$	86.84 ± 0.62	87.96 ± 0.53	$\mathit{96.54 \pm 0.23}$
9	Chimpanzee versus Seal	$\mathbf{92.68 \pm 0.34}$	92.58 ± 0.36	92.53 ± 0.38	92.59 ± 0.35	$\mathit{97.04 \pm 0.16}$
10	Giant panda versus Leopard	$\mathbf{95.26 \pm 0.24}$	95.11 ± 0.21	95.13 ± 0.24	94.95 ± 0.27	$\mathit{98.35 \pm 0.08}$
11	Giant panda versus Persian cat	94.66 ± 0.28	94.38 ± 0.23	94.66 ± 0.28	$\mathbf{94.68 \pm 0.26}$	$\mathit{95.55 \pm 0.21}$
12	Giant panda versus Pig	88.82 ± 0.40	88.69 ± 0.45	88.67 ± 0.46	$\mathbf{88.95 \pm 0.42}$	$\mathit{92.78 \pm 0.23}$
13	Giant panda versus Hippopotamus	92.62 ± 0.44	92.78 ± 0.43	92.35 ± 0.43	$\mathbf{92.85 \pm 0.42}$	$\mathit{96.98 \pm 0.16}$
14	Giant panda versus Humpback whale	98.83 ± 0.18	$\mathbf{98.88 \pm 0.14}$	98.77 ± 0.20	98.76 ± 0.22	$\mathit{99.84 \pm 0.02}$
15	Giant panda versus Raccoon	91.52 ± 0.35	91.33 ± 0.37	91.76 ± 0.34	$\mathbf{91.90 \pm 0.40}$	$\mathit{93.18 \pm 0.15}$
16	Giant panda versus Rat	$\mathbf{91.13 \pm 0.36}$	90.33 ± 0.41	90.50 ± 0.42	90.61 ± 0.47	$\mathit{95.26 \pm 0.20}$
17	Giant panda versus Seal	$\mathbf{93.63 \pm 0.31}$	93.58 ± 0.26	93.33 ± 0.29	93.40 ± 0.24	$\mathit{96.34 \pm 0.21}$
18	Leopard versus Persian cat	95.72 ± 0.21	$\mathbf{95.92 \pm 0.18}$	95.50 ± 0.25	95.65 ± 0.26	$\mathit{98.65 \pm 0.09}$
19	Leopard versus Pig	90.65 ± 0.20	$\mathbf{90.88 \pm 0.25}$	90.40 ± 0.20	90.40 ± 0.18	$\mathit{97.82 \pm 0.10}$
20	Leopard versus Hippopotamus	93.78 ± 0.27	93.81 ± 0.28	93.60 ± 0.28	$\mathbf{93.83 \pm 0.27}$	$\mathit{97.68 \pm 0.12}$
21	Leopard versus Humpback whale	99.08 ± 0.08	99.17 ± 0.08	99.06 ± 0.09	$\mathbf{99.20 \pm 0.07}$	$\mathit{99.95 \pm 0.00}$
22	Leopard versus Raccoon	$\mathbf{83.66 \pm 0.57}$	83.15 ± 0.57	83.23 ± 0.60	83.18 ± 0.64	$\mathit{91.50 \pm 0.19}$
23	Leopard versus Rat	90.43 ± 0.19	$\mathbf{90.98 \pm 0.26}$	90.28 ± 0.24	90.65 ± 0.26	$\mathit{97.19 \pm 0.12}$
24	Leopard versus Seal	95.10 ± 0.22	$\mathbf{95.49 \pm 0.19}$	94.98 ± 0.23	95.14 ± 0.22	$\mathit{98.31 \pm 0.09}$
25	Persian cat versus Pig	$\mathbf{83.71 \pm 0.49}$	83.39 ± 0.58	83.23 ± 0.44	83.38 ± 0.51	$\mathit{84.19 \pm 0.46}$
26	Persian cat versus Hippopotamus	93.11 ± 0.39	$\mathbf{93.41 \pm 0.34}$	92.66 ± 0.38	93.14 ± 0.35	$\mathit{97.50 \pm 0.13}$
27	Persian cat versus Humpback whale	96.94 ± 0.33	$\mathbf{97.26 \pm 0.29}$	96.19 ± 0.39	96.69 ± 0.39	$\mathit{99.82 \pm 0.02}$
28	Persian cat versus Raccoon	90.79 ± 0.41	$\mathbf{91.20 \pm 0.35}$	90.46 ± 0.45	90.94 ± 0.47	$\mathit{93.50 \pm 0.21}$
29	Persian cat versus Rat	69.94 ± 0.52	$\mathbf{70.40 \pm 0.48}$	69.38 ± 0.46	69.43 ± 0.43	$\mathit{73.13 \pm 0.67}$
30	Persian cat versus Seal	86.75 ± 0.64	86.91 ± 0.58	86.06 ± 0.66	$\mathbf{86.97 \pm 0.71}$	$\mathit{94.56 \pm 0.22}$
31	Pig versus Hippopotamus	77.21 ± 0.58	$\mathbf{79.02 \pm 0.63}$	76.45 ± 0.53	77.42 ± 0.54	$\mathit{88.03 \pm 0.45}$
32	Pig versus Humpback whale	97.02 ± 0.22	$\mathbf{97.32 \pm 0.18}$	96.78 ± 0.31	97.04 ± 0.19	$\mathit{99.63 \pm 0.03}$
33	Pig versus Raccoon	80.60 ± 0.56	$\mathbf{81.79 \pm 0.57}$	80.08 ± 0.53	81.50 ± 0.53	$\mathit{91.55 \pm 0.27}$
34	Pig versus Rat	72.98 ± 0.60	$\mathbf{73.68 \pm 0.53}$	72.25 ± 0.58	72.63 ± 0.50	$\mathit{84.16 \pm 0.35}$
35	Pig versus Seal	80.67 ± 0.72	$\mathbf{81.76 \pm 0.65}$	79.76 ± 0.74	80.33 ± 0.68	$\mathit{88.91 \pm 0.46}$
36	Hippopotamus versus Humpback whale	$\mathbf{93.86 \pm 0.33}$	93.75 ± 0.33	93.83 ± 0.28	93.63 ± 0.30	$\mathit{98.88 \pm 0.12}$
37	Hippopotamus versus Raccoon	86.77 ± 0.64	$\mathbf{87.37 \pm 0.61}$	86.49 ± 0.57	86.83 ± 0.68	$\mathit{94.59 \pm 0.21}$
38	Hippopotamus versus Rat	85.68 ± 0.44	$\mathbf{87.37 \pm 0.38}$	85.12 ± 0.44	85.99 ± 0.39	$\mathit{94.82 \pm 0.27}$
39	Hippopotamus versus Seal	73.78 ± 0.67	$\mathbf{75.85 \pm 0.67}$	72.82 ± 0.69	73.41 ± 0.60	$\mathit{80.90 \pm 0.55}$
40	Humpback whale versus Raccoon	97.01 ± 0.24	$\mathbf{97.15 \pm 0.22}$	96.92 ± 0.25	97.11 ± 0.22	$\mathit{99.76 \pm 0.03}$
41	Humpback whale versus Rat	95.43 ± 0.21	$\mathbf{95.53 \pm 0.18}$	95.21 ± 0.21	95.45 ± 0.21	$\mathit{99.66 \pm 0.02}$
42	Humpback whale versus Seal	86.28 ± 0.56	$\mathbf{86.93 \pm 0.47}$	86.44 ± 0.52	86.89 ± 0.52	$\mathit{96.69 \pm 0.14}$
43	Raccoon versus Rat	79.97 ± 0.46	$\mathbf{80.31 \pm 0.56}$	79.59 ± 0.47	79.67 ± 0.44	$\mathit{86.76 \pm 0.30}$
44	Raccoon versus Seal	92.52 ± 0.28	$\mathbf{92.80 \pm 0.24}$	92.22 ± 0.28	92.55 ± 0.23	$\mathit{94.26 \pm 0.21}$
45	Rat versus Seal	81.11 ± 0.62	$\mathbf{82.34 \pm 0.62}$	80.44 ± 0.64	80.68 ± 0.73	$\mathit{92.46 \pm 0.35}$

Table 1: AwA dataset (attributes as privileged information). The numbers are mean and standard error of the AP performance over 20 runs. The best result is highlighted in **boldface**, which in total is **7** for SVM rank, **27** for Rank Transfer, **1** for SVM, and **10** for SVM+. Highlighted blue indicates significant improvement of the methods that utilize privileged information (Rank Transfer and/or SVM+) over the methods that do not (SVM rank and SVM). We used a paired Wilcoxon test with 95% confidence level as a reference. Additionally, we also provide the SVM rank performance on $\mathcal{X}^*$ (last column).

classes, for which the attribute annotation is provided together with the dataset. The 10 classes are *chimpanzee, giant panda, leopard, persian cat, pig, hippopotamus, humpback whale, raccoon, rat, seal*, and contain 6180 images in total. The attributes capture 85 properties of the animals, color, texture, shape, body parts, behavior among others. We use L_1 normalized 2000 dimensional SURF descriptors [1] as original features, and 85 dimensional predicted attributes as the privileged information. The values of the predicted attributes are obtained from DAP model [14] and correspond to probability estimates of the binary attributes in the images. We train 45 binary classifiers for each pair

of the 10 classes with varying size of training data: 50, 100 images per class. We use 200 samples per class for testing. To get better statistics of the performance we repeat the procedure of train/test split 20 times. Due to space constraints, we only include the results with $N = 100$ training samples per class here. Please, refer to the supplementary material for the case $N = 50$.

Results. As we can see from the Figure 2, utilizing attributes as privileged information for object classification task is useful. Rank Transfer outperforms SVM rank in 34 out of 45 cases, and SVM+ outperforms SVM in 39 out of 45 cases. Noticeably, the Rank Transfer model is able to utilize privileged information better than the SVM+. We observe partial overlap of cases where Rank Transfer and SVM+ are not able to utilize privileged information (location of the red bars). Full comparison of the AP performance of all methods is shown in the Table 1. In general, we observe that ranking-based models are superior to the non-ranking ones, and in particular, we can see clear advantage of the Rank Transfer model over all other baselines. We also notice, that the gain of the Rank Transfer method is higher in the regime when the problem is hard, i.e. when AP performance is below 90%. We obtain very similar results with $N = 50$ training samples per class. As a further analysis, we also check the hypothetical performance of SVM rank in the privileged space $\mathcal{X}^*$. The privileged information has consistently higher AP performance than SVM rank in $\mathcal{X}$. In most cases, higher AP performance in the privileged space than in the original translates to positive effect in rank transfer. We also analyze the data ranking in the original space, privileged space, and in the original space with transferred rank. Typically the data ranking in the privileged space of attributes is well spread out comparing to the original space. In this case the distinction between easy-to-separate and hard-to-separate pairs is feasible in the privileged space and we can potentially benefit from it by transferring the rank.

4.2. Bounding box as privileged information

Bounding box annotation is designed to capture the exact location of an object in the image. When performing image-level object recognition, knowing the exact location of the object in the training data is privileged information. We use a subset of the categories from the ImageNet 2012 challenge (ILSVRC2012) for which bounding box annotation is available[3]. We define two groups of interest: group with variety of snakes, and group with balls in different sport activities. The group of snakes has 17 classes: *thunder snake, ring-neck snake, hognose snake, green snake, king snake, garter snake, water snake, vine snake, night snake, boa constrictor, rock python, indian cobra, green mamba, sea snake, horned viper, diamondback, sidewinder*, and has 8254 images in total, on average 500 samples per class. We ignore few images with too small bounding box region, and use 8227 images for further analysis. The group balls has 6 classes: *soccer ball, croquet ball, golf ball, ping-pong ball, rugby ball, tennis ball*, and has 3259 images in total, on average 500 samples per class. Here, we also ignore images with uninformative bounding box annotation and use 3165 images instead. We consider one-versus-rest scenario for each group separately. We use L_2 normalized 4096-dimensional Fisher vectors [19] extracted from the whole images as well as from only the bounding box regions, and we use the former as the original data representation and the latter as privileged information. We train one binary classifier for each class, 17 in the first group and 6 in the second group. For training we use 100 images from the desired class and 100 samples randomly drawn from the remaining classes. For testing we use the rest of the images in the desired class and the same amount from the other categories. To get better statistics of the performance we repeat the train/test split 10 times. Due to space constraints, we only include the results with the group of snakes here. Please, refer to the supplementary material for the group of balls.

Results. As we can see from Table 2, utilizing bounding box annotation as privileged information for fine-grained classification is useful. We show the pairwise difference in performance of the methods that utilize privileged information and that do not in the bar plot on the right of Table 2. Rank Transfer clearly outperforms SVM rank (image) in 10 cases, and SVM+ outperforms SVM in 12 cases out of 17. In this experiment, the SVM+ method is able to exploit the privileged information better than the Rank Transfer method (in 13 out of 17 cases). And overall we observe that non-ranking models are superior to the ranking ones. In the group of balls, both LUPI methods outperform non-LUPI baselines in 4 out of 6 cases (refer to the supplementary). Noticeably, SVM rank performs worse than all other methods, where as standard SVM is a competitive baseline. Interestingly, the performance in the privileged space is not superior to the original data space, sometimes it is even worse, especially in the group of balls. However the LUPI methods are able to exploit easy and hard samples in both spaces. We credit this to the fact that in this experiment, both spaces are of the same modality, i.e. the privileged information is obtained from a subset of the same image features that are used for the original data representation. Thus, our underlying assumption that the same examples are easy and hard in both modalities is fulfilled.

4.3. Textual description as privileged information

A textual description provides complementary view to a visual representation of an object. This can be used as privileged information in object classification task. We use *IsraelImages* dataset introduced in [2]. The dataset has 11

[3] http://www.image-net.org/challenges/LSVRC/2012/index

	SVM rank image	Rank Transfer image+bbox	SVM image	SVM+ image+bbox	Reference (SVM rank bbox)
Thunder snake	66.48 ± 0.72	66.23 ± 0.73	66.51 ± 0.72	$\mathbf{67.52 \pm 0.37}$	*68.06 ± 0.64*
Ringneck snake	73.33 ± 0.63	73.32 ± 0.68	$\mathbf{73.71 \pm 0.82}$	73.51 ± 0.59	*74.12 ± 0.66*
Hognose snake	72.33 ± 0.60	72.67 ± 0.61	72.54 ± 0.42	$\mathbf{72.89 \pm 0.61}$	*75.12 ± 0.40*
Green snake	76.91 ± 0.66	$\mathbf{77.22 \pm 0.66}$	77.01 ± 0.70	76.25 ± 0.97	*77.30 ± 0.56*
King snake	85.99 ± 0.27	86.22 ± 0.36	85.44 ± 0.34	$\mathbf{86.67 \pm 0.26}$	*87.87 ± 0.23*
Garter snake	$\mathbf{83.74 \pm 0.61}$	83.51 ± 0.60	81.57 ± 0.68	83.41 ± 0.89	*86.86 ± 0.53*
Water snake	72.07 ± 0.57	71.92 ± 0.50	$\mathbf{73.03 \pm 0.57}$	72.01 ± 0.86	*68.49 ± 0.58*
Vine snake	85.24 ± 0.51	85.21 ± 0.51	$\mathbf{85.81 \pm 0.51}$	85.06 ± 0.56	*85.99 ± 0.50*
Night snake	57.69 ± 1.37	57.64 ± 1.25	58.17 ± 1.39	$\mathbf{58.39 \pm 1.06}$	*58.27 ± 0.92*
Boa constrictor	81.44 ± 0.71	81.59 ± 0.69	79.88 ± 0.80	$\mathbf{82.15 \pm 0.72}$	*82.25 ± 0.58*
Rock python	65.56 ± 1.14	65.92 ± 1.18	64.16 ± 1.35	$\mathbf{66.94 \pm 0.83}$	*67.17 ± 1.22*
Indian cobra	65.90 ± 0.95	65.89 ± 1.02	66.20 ± 0.96	$\mathbf{66.38 \pm 0.44}$	*65.96 ± 0.57*
Green mamba	75.30 ± 0.25	75.62 ± 0.32	$\mathbf{76.18 \pm 0.46}$	76.07 ± 0.42	*76.75 ± 0.43*
Sea snake	87.70 ± 0.45	87.91 ± 0.48	87.86 ± 0.38	$\mathbf{88.26 \pm 0.37}$	*84.00 ± 0.50*
Horned viper	77.00 ± 0.47	77.36 ± 0.45	77.09 ± 0.51	$\mathbf{77.84 \pm 0.59}$	*81.56 ± 0.40*
Diamondback	83.69 ± 0.70	84.19 ± 0.60	82.00 ± 0.50	$\mathbf{84.29 \pm 0.52}$	*85.66 ± 0.17*
Sidewinder	75.03 ± 0.68	$\mathbf{75.90 \pm 0.67}$	74.56 ± 1.10	75.47 ± 0.94	*77.94 ± 0.82*

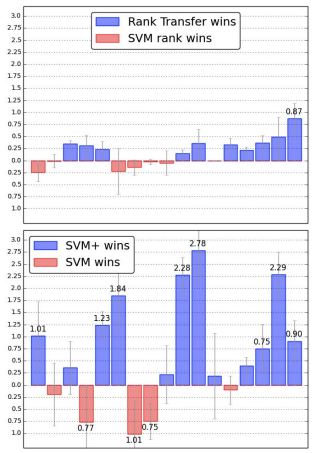

Table 2: ImageNet dataset, group of snakes (bounding box annotation as privileged information). The numbers are mean and standard error of the AP performance over 10 runs. The best result is highlighted in **boldface**. Highlighted blue indicates significant improvement of the methods that utilize privileged information (Rank Transfer and/or SVM+) over the methods that do not (SVM rank and SVM). We used a paired Wilcoxon test with 95% confidence level as a reference. Additionally, we also provide the SVM rank performance on $\mathcal{X}^*$ (last column). The bar plots on the right show advantage of the LUPI methods over non-LUPI (Rank Transfer versus SVM rank, SVM+ versus SVM). The length of the 17 bars corresponds to relative improvement of the average precision over 17 snake classes.

classes, 1823 images in total, with a textual description (up to 18 words) attached to each of the image. The number of samples per class is relatively small, around 150 samples, and varies from 96 to 191 samples. We merge the classes into three groups: nature (birds, trees, flowers, desert), religion (christianity, islam, judaism, symbols) and urban (food, housing, personalities), and perform binary classification on the pairs of groups. We use L_2 normalized 4096-dimensional Fisher vectors [19] extracted from the images as the original data representation and bag-of-words representation of the text data as privileged information. We use 100 images per group for training and all the rest for testing. We repeat the train/test split 20 times.

Results. As we can see from Table 3, utilizing textual privileged information as provided in the *IsraelImages* dataset does not help. All four methods have near equal performance, and there is no signal of privileged information being utilized in both LUPI methods. This might seem contradictory to the high performance of the reference baseline in the text domain, $\mathcal{X}^*$. However high accuracy in the privileged space does not necessarily mean that the privileged information is helpful. For example, assume we used the labels themselves as privileged modality: classification would be trivial, but it would provide no additional information to transfer. In the *IsraelImages*, the textual descriptions of the images are sparse and contain duplicates. For samples with identical scores there is no information in their relative ranking. Therefore, ranking the samples in the privileged space does not capture the relation between objects and mainly preserves the class separation only. The performance does not degrade nevertheless.

4.4. Rationales as privileged information

Rationale annotation was introduced in [7] as a way to capture additional information *why* an annotator makes the decision about the given image. This information is provided in the form of the most informative hand-annotated region in the image, and is privileged in the LUPI framework. We use the *Hot or Not* dataset[4], which is designed for binary classification, whether male and female people in the images are *hot*. Following the setting of [7], we use 108 and 104 images for female and male classes accordingly with $+1$(hot) and -1(not) label. The original data and the privileged data are represented with L_1 normalized 500 dimensional densely sampled SIFT features, extracted from the whole image and from the rationales accordingly. We train two binary classifiers for females and males separately with varying size of training data 50 and 100 samples. We use the remaining data for testing. We repeat the train/test split 100 times.

Results. First we would like to mention that the *Hot or Not* dataset is very challenging: the AP performance is relatively low for all four methods and the standard deviation is high. In case of $N = 50$, the data contains too little signal for any method to work with, thus it is hard to draw a conclusion. In case of $N = 100$ (see Table 4), we can not make statistical summary of the result due to small test sample size (8 and 4 samples accordingly). Nevertheless we perform this experiment to position our result with respect to the original work [7]. For *male* class the authors report classification accuracy 60.01%, and 57.07% for *female* class.

[4] http://vision.cs.utexas.edu/projects/rationales/

	SVM rank image	Rank Transfer image+text	SVM image	SVM+ image+text	Reference (SVM rank text)
Nature vs Religion	89.06 ± 0.34	89.28 ± 0.24	89.51 ± 0.27	89.41 ± 0.26	96.94 ± 0.12
Religion vs Urban	71.82 ± 0.66	71.71 ± 0.59	72.04 ± 0.56	72.11 ± 0.40	96.20 ± 0.11
Nature vs Urban	88.56 ± 0.23	88.94 ± 0.22	88.85 ± 0.24	88.92 ± 0.23	94.41 ± 0.17

Table 3: Israeli dataset (textual description as privileged information). The numbers are mean and standard error of the AP performance over 20 runs. As reference we also provide the SVM rank performance on the $\mathcal{X}^*$ (last column).

	SVM rank image	Rank Transfer image+rationale	SVM image	SVM+ image+rationale	Reference (SVM rank rationale)
Female N=100	58.06 ± 1.40	56.58 ± 1.34	57.58 ± 1.39	57.06 ± 1.49	75.65 ± 1.52
Male N=100	72.33 ± 1.82	75.50 ± 1.97	72.25 ± 1.75	73.58 ± 1.81	79.91 ± 1.94

Table 4: HotOrNot dataset (rationale as privileged information). The numbers are mean and standard error of the AP performance over 100 runs. As reference we also provide the SVM rank performance on $\mathcal{X}^*$ (last column).

Because of the different performance measures (AP versus accuracy) we can not directly compare our results, but the numbers are 'in the same ballpark'.

5. Conclusion

We have studied the setting of learning using privileged information (LUPI) in visual object classification tasks. We showed how it can be applied to several situations that previously were handled by hand-crafted separate methods. Our experiments show that prediction performance often improves when utilizing the privileged information. When it does not, at least no negative transfer occurs. We have studied two approaches for solving the LUPI task: SVM+ and the proposed Rank Transfer method. Rank Transfer shows comparable performance to the SVM+ algorithm and can easily be applied using standard SVM solvers.

In future work, we plan to further analyze the potential of both approaches, also in light of recent results that SVM+ classifiers can be reformulated as a special forms of example-weighted binary SVMs [15].

Acknowledgments. This work was in parts funded by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no 308036. NQ is supported by the Newton International Fellowship. We thank Dmitry Pechyony for SVM+ code.

References

[1] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (SURF). *CVIU*, 2008.

[2] R. Bekkerman and J. Jeon. Multi-modal clustering for multimedia collections. In *CVPR*, 2007.

[3] M. B. Blaschko and C. H. Lampert. Correlational spectral clustering. In *CVPR*, 2008.

[4] O. Chapelle. Training a support vector machine in the primal. *Neural Computation*, pages 1155–1178, 2007.

[5] C. M. Christoudias, R. Urtasun, and T. Darrell. Multi-view learning in the presence of view disagreement. In *UAI*, 2008.

[6] C. Dance, J. Willamowski, L. Fan, C. Bray, and G. Csurka. Visual categorization with bags of keypoints. In *ECCV*, 2004.

[7] J. Donahue and K. Grauman. Annotator rationales for visual recognition. In *ICCV*, 2011.

[8] M. Everingham, L. J. V. Gool, C. K. I. Williams, J. M. Winn, and A. Zisserman. The Pascal VOC challenge. *IJCV*, 2010.

[9] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. LIBLINEAR: A library for large linear classification. *JMLR*, 2008.

[10] P. Gehler and S. Nowozin. On feature combination for multiclass object classification. In *ICCV*, 2009.

[11] T. Joachims. Making large-scale SVM learning practical. In *Advances in kernel methods*, pages 169–184, 1999.

[12] T. Joachims. Optimizing search engines using clickthrough data. In *KDD*, 2002.

[13] D. Kuettel, M. Guillaumin, and V. Ferrari. Segmentation propagation in ImageNet. In *ECCV*, 2012.

[14] C. H. Lampert, H. Nickisch, and S. Harmeling. Attribute-based classification for zero-shot visual object categorization. *PAMI*, 2013.

[15] M. Lapin, M. Hein, and B. Schiele. Learning using privileged information: SVM+ and weighted SVM, 2013. arXiv:1306.3161 [stat.ML].

[16] O. Maron and A. L. Ratan. Multiple-instance learning for natural scene classification. In *ICML*, 1998.

[17] D. Pechyony and V. Vapnik. On the theory of learning with privileged information. In *NIPS*, 2010.

[18] D. Pechyony and V. Vapnik. Fast optimization algorithms for solving SVM+. In *Stat. Learning and Data Science*, 2011.

[19] F. Perronnin, J. Sánchez, and T. Mensink. Improving fisher kernel for large-scale image classification. In *ECCV*, 2010.

[20] N. Quadrianto and C. H. Lampert. Learning multi-view neighborhood preserving projections. In *ICML*, 2011.

[21] B. Schölkopf and A. Smola. *Learning with Kernels*. MIT Press, 2002.

[22] S. Shalev-Shwartz, Y. Singer, and N. Srebro. Pegasos: Primal estimated sub-gradient solver for SVM. In *ICML*, 2007.

[23] C. G. Snoek, M. Worring, and A. W. Smeulders. Early versus late fusion in semantic video analysis. In *ACM MM*, 2005.

[24] V. Vapnik. *The nature of statistical learning theory*. Springer, 1999.

[25] V. Vapnik and A. Vashist. A new learning paradigm: Learning using privileged information. *Neural Networks*, 2009.

[26] A. Vedaldi, V. Gulshan, M. Varma, and A. Zisserman. Multiple kernels for object detection. In *ICCV*, 2009.

[27] Y. Wang and G. Mori. A discriminative latent model of object classes and attributes. In *ECCV*, 2010.

[28] O. F. Zaidan and J. Eisner. Modeling annotators: A generative approach to learning from annotator rationales. In *Conference on Empirical Methods in Natural Language Processing*, 2008.