# Laboratory Experience 3 - Optional

**Task** This experience focuses on the MKTL transfer learning algorithm [1]. Specifically we will focus on three algorithms:

- the no-transfer baseline
- the prior-features (outputs of prior models as features) baseline
- a simplification of the MKTL algorithm

and one dataset

- the Animals with Attributes dataset [2], containing images of 50 animal classes, obtained by querying internet search engines with animal names as keywords.

The goal of this experience is to reproduce the results reported in section 4.2 of [1].

**Data**

- The precomputed features for the Animals with Attributes dataset, can be downloaded from here:
  http://attributes.kyb.tuebingen.mpg.de/
- You will need the SURF, the Color Histogram and the PHOG features, as well as the Base Package

**A short explanation of the algorithm.** Consider the scenario where we are provided $Z$ black-box prior-knowledge classifiers and we want to use them as experts, to predict the label of a new sample $x$. Together with the predictions of the experts, we want our classifier to directly use also the new sample $x$, in feature space $\psi$. When predicting the score of a new sample w.r.t. the target class $y$, our transfer learning scoring function could thus have the form:

$$s(x, y) = w_y^0 \cdot \psi(x) + \sum_{z=1}^{Z} w_y^z \cdot s_z(x) = \sum_{z=0}^{Z} w^z \cdot \phi^z(x) \tag{1}$$

where $s_z(x)$ is the score assigned by classifier $z$ to sample $x$ and:

$$\phi^z(x) = \begin{cases} \psi(x) & \text{if } z = 0 \\ s_z(x) \text{otherwise.} \end{cases} \tag{2}$$

A classifier of this form can be easily learned with any multiple kernel learning algorithm, considering the kernels $k^z(x_1, x_2) = \phi^z(x_1) \cdot \phi^z(x_2)$, induced by the feature maps $\phi^z(x)$.

**Implementation** Instead of implementing the algorithm described in the paper, which is based on a multi-class multiple kernel learning algorithm [3], you are asked to implement the following simplification:

- Compute a single linear kernel $k^p(x_1, x_2) = \psi^p(x_1) \cdot \psi^p(x_2)$, using as a feature descriptor of each target sample $\psi^p(x)$, the vector of all the decision values produced by the prior-knowledge classifiers on the target image: $\psi^p(x) = \begin{bmatrix} s_1(x) & s_2(x) & \dots & s_Z(x) \end{bmatrix}$
- Compute one kernel on the features extracted from the target images: $k(x_1, x_2) = \psi(x_1) \cdot \psi(x_2)$
- Obtain the final classifier in one of the following ways:
  - **Uniform combination** average the two kernels to obtain a uniformly weighted combination of prior and target knowledges. Make use of a One-VS-All SVM (i.e. the one used for the first mandatory experience) to classify the target images
  - **Weighted combination** make use of a One-VS-ALL Simple-MKL algorithm (i.e. the one used for the second mandatory experience) to learn a weighted combination of the two kernels, as well as the classifier on the target images

The two baselines algorithms can be implemented by training a One-VS-All SVM on:

- **Prior-features** only the linear kernel $k^p(x_1, x_2)$ computed using the prior knowledge
- **No-transfer** only the kernel $k(x_1, x_2)$ computed using the features on the target images

**Experiments** You are asked to perform the following experiment:

- Consider the same 10 target classes used in the dataset paper [2]
- Randomly sample 10, 20 and 40 target training images for each target-class and use the remaining ones for testing
- Use a $\chi^2$ kernel on the PHOG features to describe the target images
- Use the remaining 40 classes to build the prior knowledge classifiers. Randomly sample 90 training images / class
- Use the average of two RBF kernels, computed using Color Histogram and SURF to describe all the prior knowledge images
- Train the prior models using a One-VS-All SVM
- For each algorithm and training-set size compute the multiclass accuracy (as the mean class recognition rate)
- Run the experiment twice, on two random split and plot the average$\pm$std results, as in figure 3 (right) of [1]

# References


[1] L. Jie, T. Tommasi, and B. Caputo. Multiclass transfer learning from unconstrained priors. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1863–1870. IEEE, 2011.

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

[3] F. Orabona, L. Jie, and B. Caputo. Online-batch strongly convex multi kernel learning. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 787–794. IEEE, 2010.