The joint submission of the TU Berlin and Fraunhofer FIRST (TUBFI) to the ImageCLEF2011 Photo Annotation Task

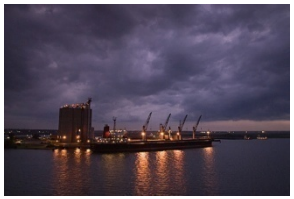
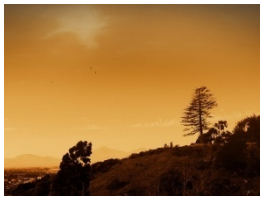
Machine Learning Group
Berlin Institute of Technology
www.ml.tu-berlin.de

www.first.fraunhofer.de

Klaus-Robert Müller

Joint Work with Alexander Binder, Wojciech Samek, Marius Kloft and Motoaki Kawanabe

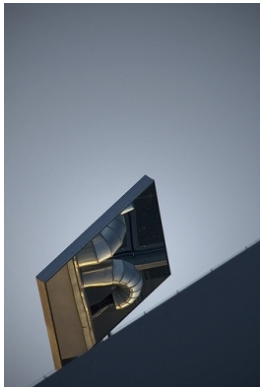

The Task

- Annotate 10000 images into 99 overlapping concepts.
- Training Corpus of 8000 annotated images
- Evaluation by measures like MAP, AUC, EER.
- User Tags available.







Our History

- We started computer vision research from July 2007 (the project THESEUS).
- ImageCLEF2009 Photo Annotation (We were 3rd/4th)
  - Inspired by other front-runners, we prepared many kernels based on various image features (e.g. SIFT Bag-of-Words).
  - We combined them by non-sparse Multiple Kernel Learning (Kloft et al. JMLR 2011).
- ImageCLEF2010 Photo Annotation (1st/2nd/3rd in a post-challenge evaluation)
  - More and Improved image features

ImageCLEF2011 Photo Annotation

- Non-sparse Multiple Kernel Learning (MKL) to combine many kernels from various Bag-of-Words features
  (e.g. multiple color channels as in van de Sande et al. 2010 and spatial tilings as in Lazebnik et al. 2006)
- Multi-Task Learning (MTL) via non-sparse MKL (Samek et al. CAIP 2011) to deal with (semantic) relations between concepts

  >>> the best MAP result in visual task
- Multi-modal classification via kernel combination (Kawanabe et al. WACV 2011)

  >>> the best MAP result in multi-modal task
- Novel random biased sampling
- Modified soft mapping

Combining multiple descriptors

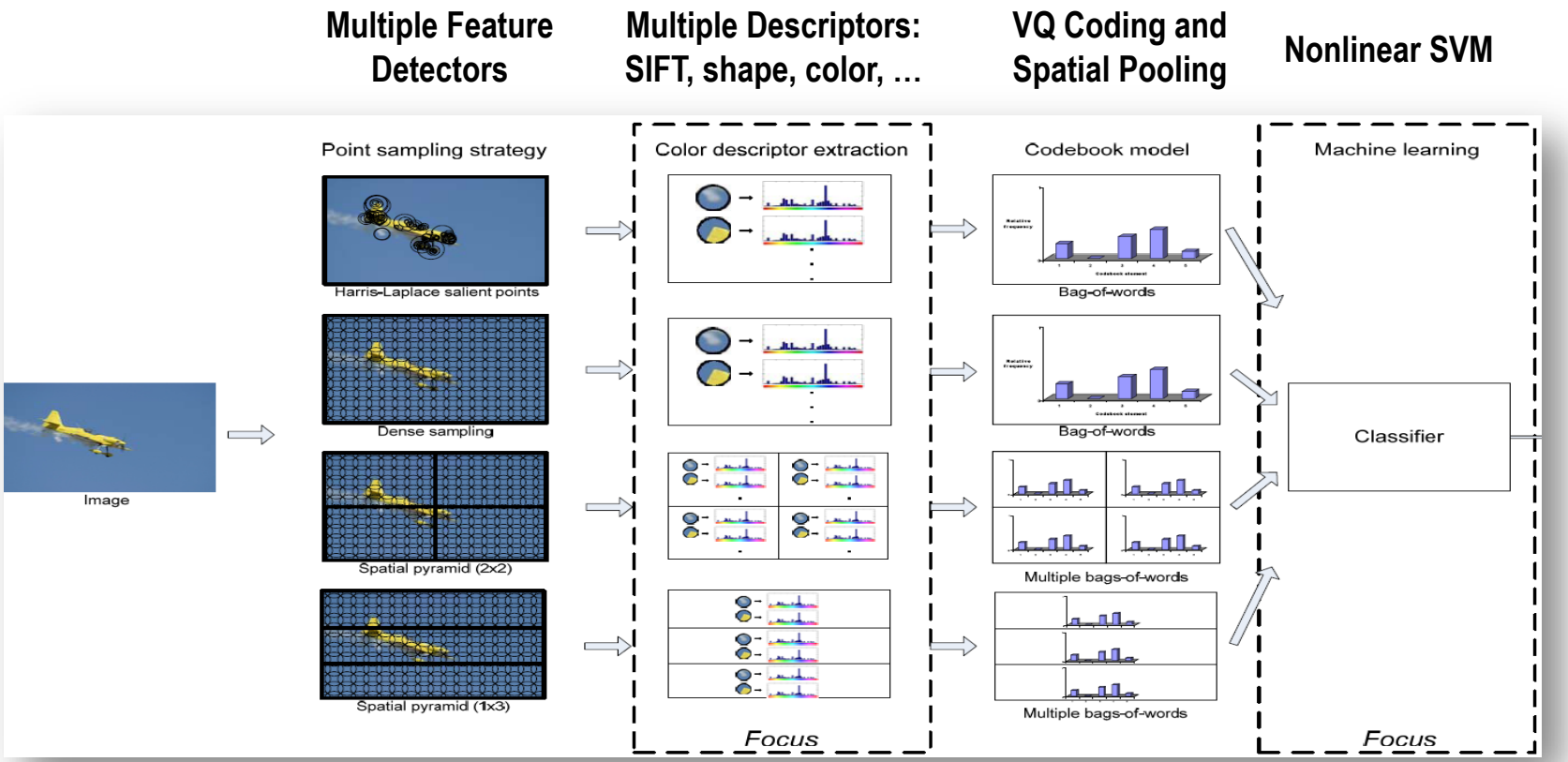
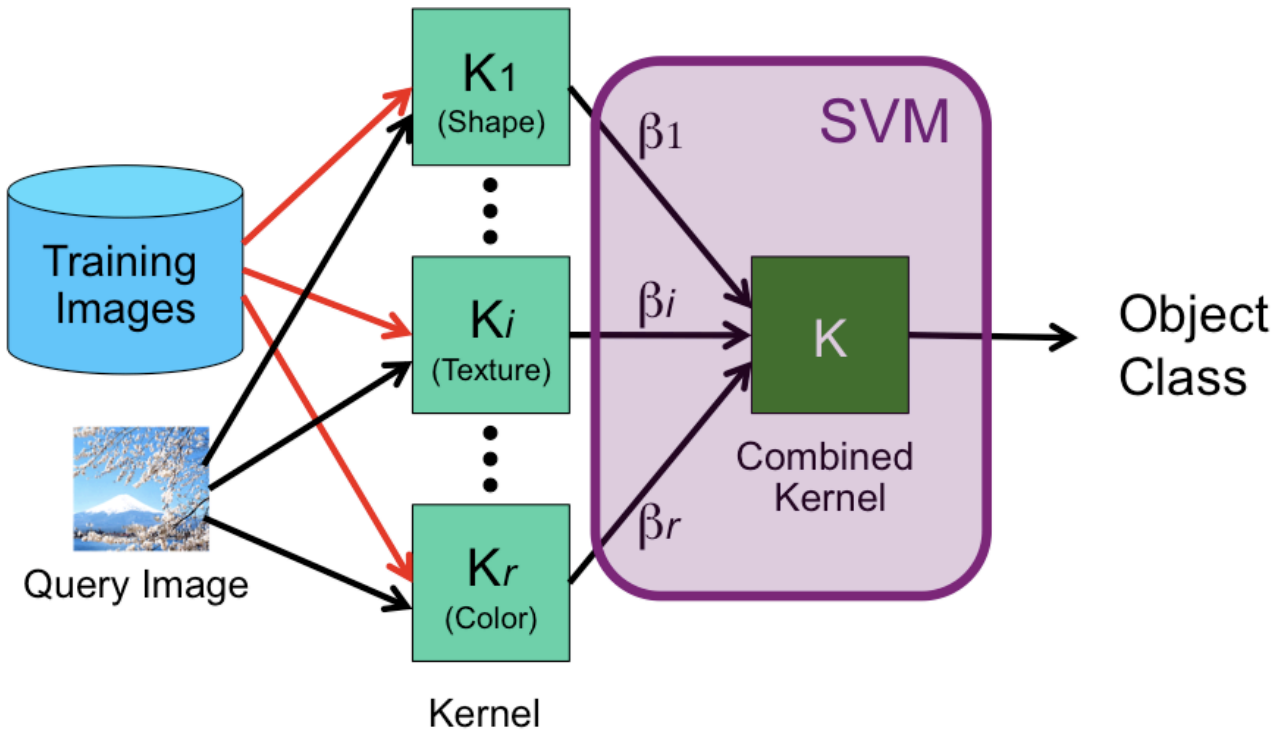


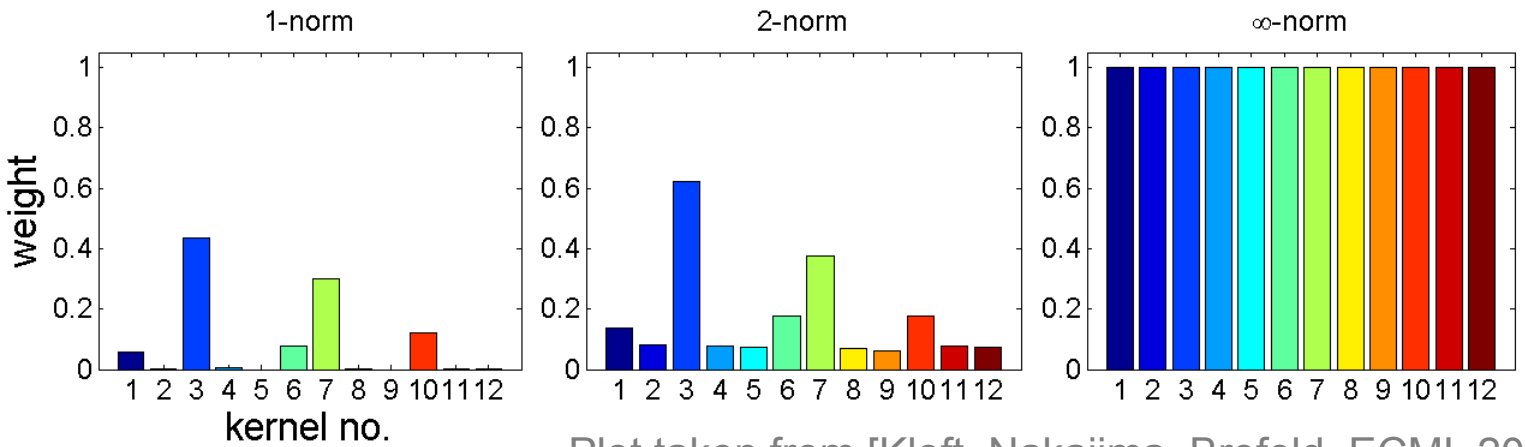
Diagram from SurreyUVA_SRKDA, winner team in PASCAL VOC 2008



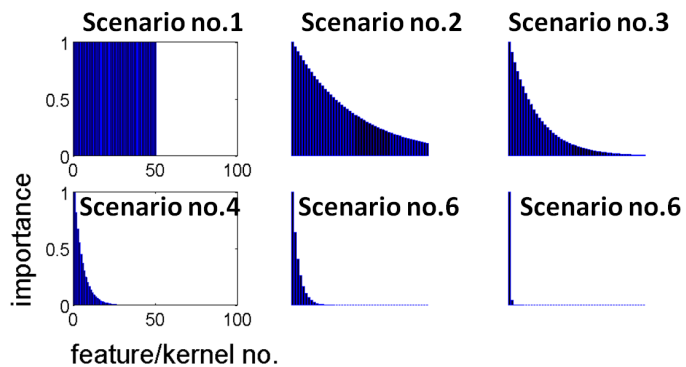
Non-sparse MKL (Min-Max Problem)

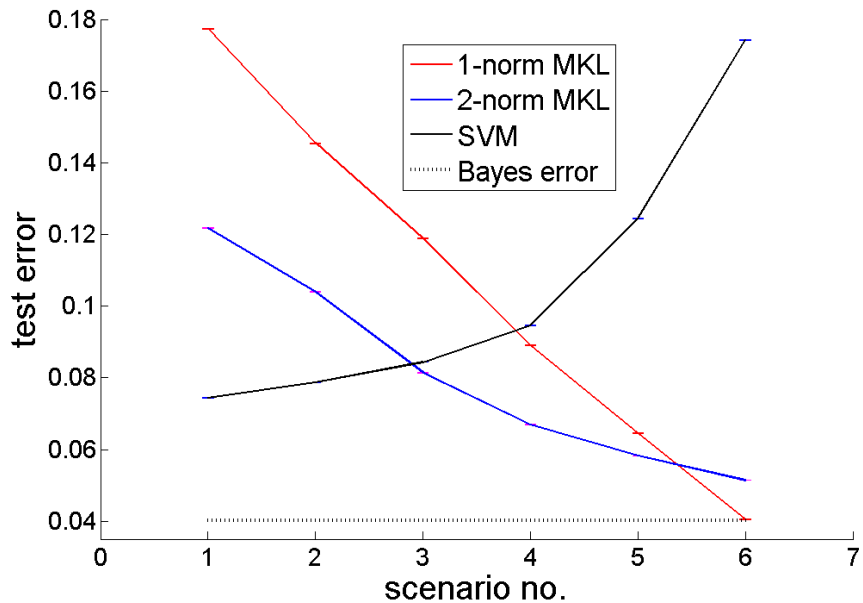
$$\min_{\boldsymbol{\beta}} \max_{\boldsymbol{\alpha}} \quad \boldsymbol{y}^\top \boldsymbol{\alpha} - \frac{1}{2} \boldsymbol{\alpha}^\top \sum_j \beta_j K_j \boldsymbol{\alpha}$$

$$\text{s.t.} \quad \mathbf{0} \leq \boldsymbol{\alpha} \leq \boldsymbol{C}; \quad \boldsymbol{y}^\top \boldsymbol{\alpha} = 1; \quad \boldsymbol{\beta} \geq 0; \quad \boxed{||\boldsymbol{\beta}||_p^p = 1.}$$



Plot taken from [Kloft, Nakajima, Brefeld, ECML 2009]

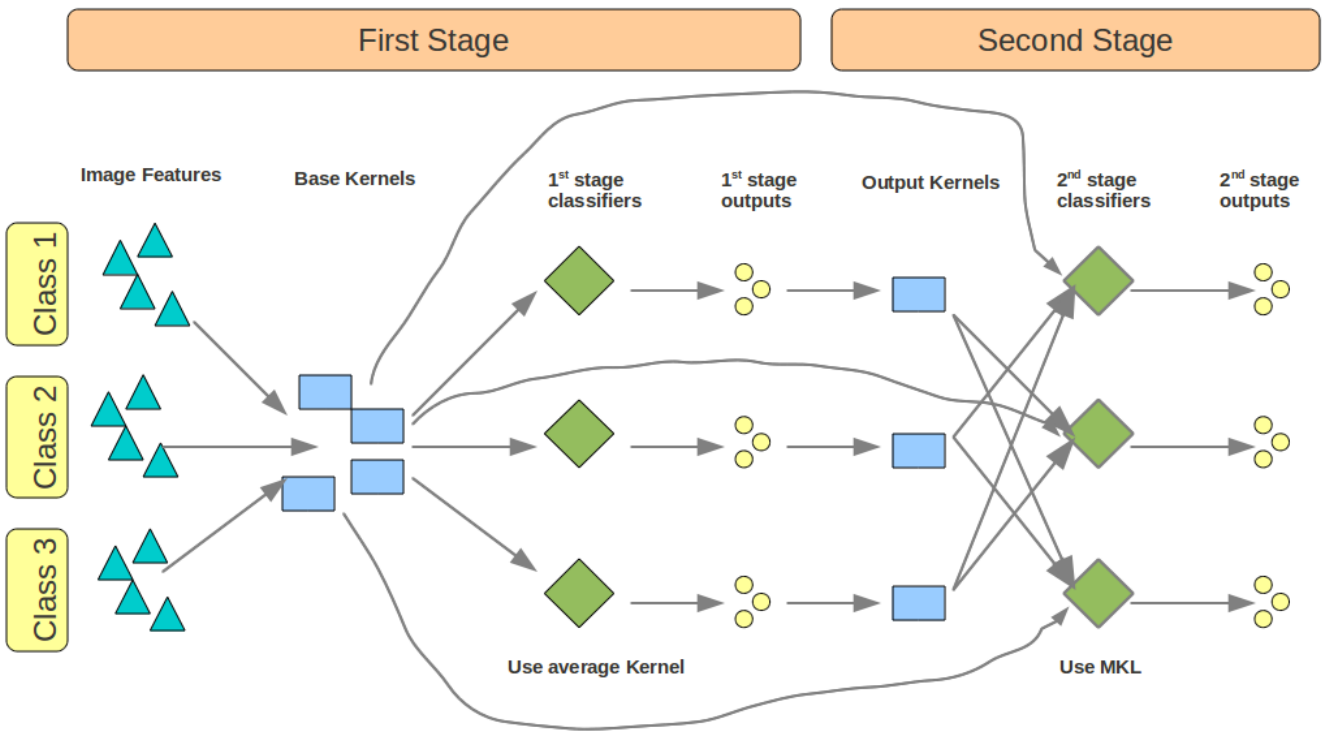




Multi-Task Learning via MKL (Samek et al. CAIP2011)

- Output kernels measuring similarity between classification scores of other concepts

$$\widetilde{K}_c(\mathbf{x}_i, \mathbf{x}_j) = \exp\left[-\left\{s_c(\mathbf{x}_i) - s_c(\mathbf{x}_j)\right\}^2\right]$$

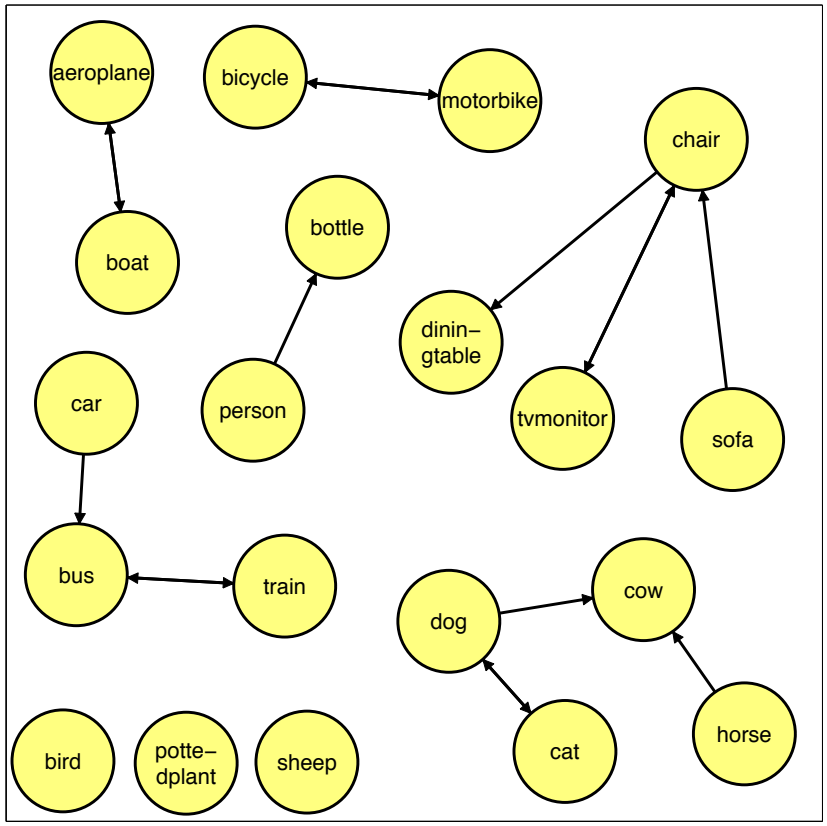


Multi-Task Learning via MKL (Samek et al. CAIP2011)

- Advantages over classical MTL (e.g. Evgeniou et al. 2005):

1. Sample size scalability: multiple N x N kernels instead of single (CN) x (CN) kernel (C:=#Tasks, N:=#Samples)
2. Asymmetric interactions: bad performing concepts do not spoil better performing ones

We included the output kernels of the following concepts: *animals*, *food*, *no_persons*, *outdoor*, *indoor*, *building_sights*, *landscape_nature*, *single_person*, *sky*, *water*, *sea*, *trees*



Directional Interactions of the VOC 2009 Task

Multi-modal classification using visual and tag kernels (Kawanabe et al. WACV2011)

- User tags have potential to provide higher-order semantic information, but they are very noisy (e.g. personal variation)
- Smoothing tag BoW feature by Markov Random Walks (MRW) based on co-occurrence graph between selected tags
- Combining visual kernels and MRW tag kernels with different steps (smoothing levels) with uniform weights (infinity-norm MKL)

Visual		Proposed		Tags	
		✓	MusicalInstrument	guitar	bass
×	No_Person	✓	Small_Group	live	music
		✓	male	rockstorm	metal
				maldives	. . .

Results

Results by MAP for the best three submissions.

Submission	Modality	MAP on test data
BPACAD 3	T(extual)	**34.56**
IDMT 1	T	32.57
MLKD 1	T	32.56
TUBFI 1	V(isual)	38.27
TUBFI 2	V	37.15
TUBFI 4	V	**38.85**
TUBFI 5	V	38.33
CAEN 2	V	38.24
ISIS 3	V	37.52
TUBFI 3	V+T	**44.34**
LIRIS 5	V+T	43.70
BPACAD 5	V+T	43.63

Multi-modal classification using visual and tag kernels (Kawanabe et al. WACV2011)

ImageCLEF2010

ImageCLEF2011 Photo Annotation

- Non-sparse Multiple Kernel Learning (MKL) to combine many kernels from various Bag-of-Words features
  (e.g. multiple color channels as in van de Sande et al. 2010 and spatial tilings as in Lazebnik et al. 2006)
- Multi-Task Learning (MTL) via non-sparse MKL (Samek et al. CAIP 2011) to deal with (semantic) relations between concepts

  >>> the best MAP result in visual task
- Multi-modal classification via kernel combination (Kawanabe et al. WACV 2011)

  >>> the best MAP result in multi-modal task
- Novel random biased sampling
- Modified soft mapping

End

Thank You

TUBFI Submissions(1)

- all: tackle diversity of concepts by diversity of methods
- TUBFI 1: SVM over one big set of kernels
- TUBFI 2: TUBFI 1+ SVMs over subsets of TUBFI 1 kernel set
- TUBFI 5: TUBFI 2+ non-sparse Multiple Kernel Learning (MKL) *heuristically down-scaled*
- TUBFI 4: TUBFI 5+ Multi-task Learning (MTL) using heuristically down-scaled non-sparse MKL
- TUBFI 3: multimodal: TUBFI 1+ (BoW) over User Tags