Semantic Kernel Forests from Multiple Taxonomies

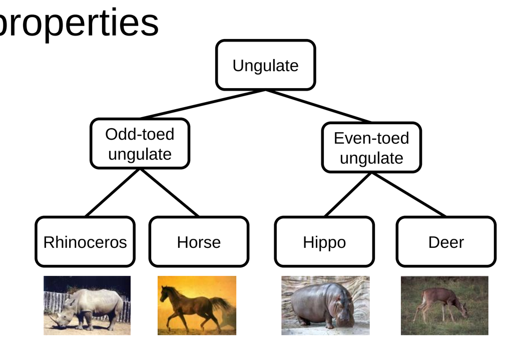
Sung Ju Hwang[1], Kristen Grauman[1], and Fei Sha[2]

[1]University of Texas at Austin, [2]University of Southern California

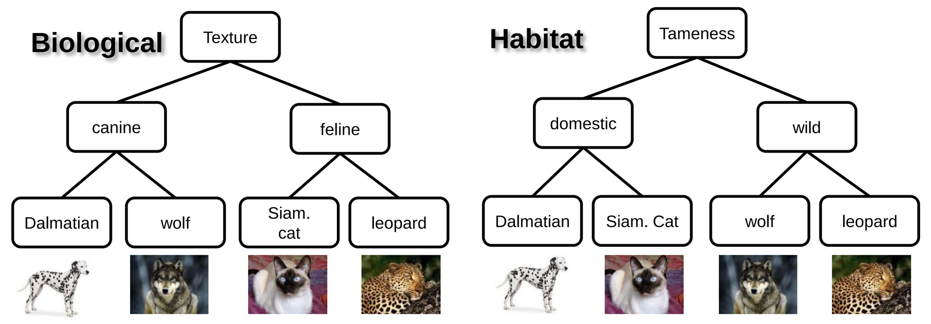
Motivation

The potential use of semantic taxonomies in object categorization is limited

1) Semantic hierarchy **need not align** with visual properties

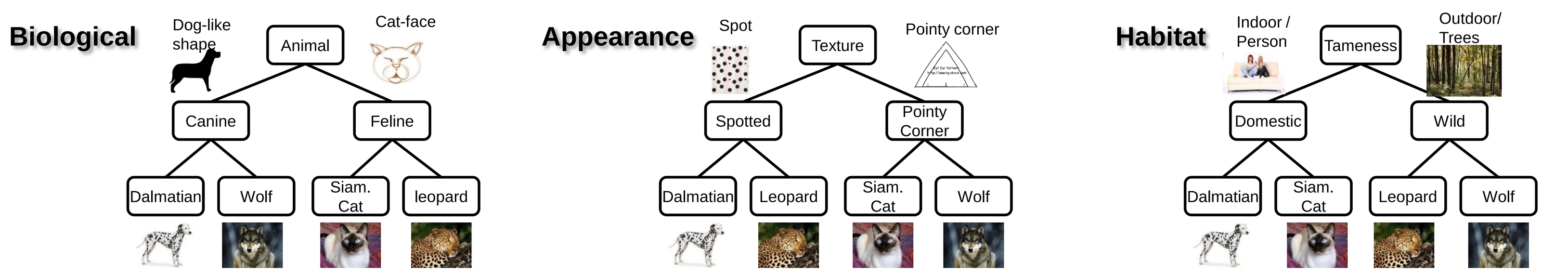


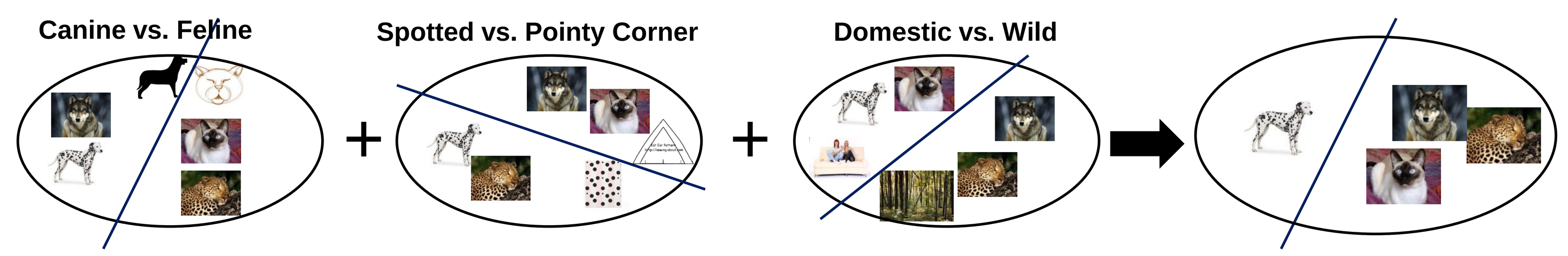
2) There exists no single **optimal** hierarchy



Main Idea

Exploit multiple human-provided taxonomies to learn **complementary** visual features, and **combine** them for discriminative feature learning





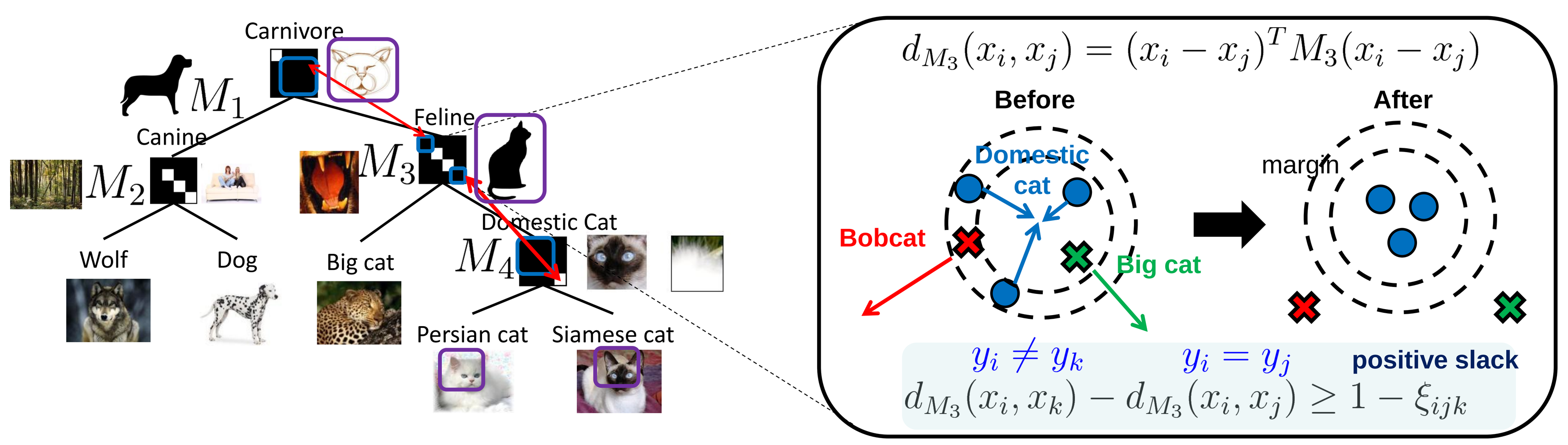
Non-linear feature spaces learned at each **view** and **granularity** Combined feature space

Learning a semantic kernel forest

How can we learn granularity-and view- specific features from multiple taxonomies?

Step 1: Isolating granularity-specific features

Features **useful for superclass** discrimination **not useful for its subclass** discrimination
- e.g. features useful for distinguishing canine and feline should differ from those for Siamese vs. Persian cat

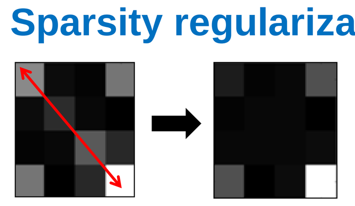


$$d_{M_3}(x_i,x_j) = (x_i - x_j)^T M_3 (x_i - x_j)$$

$$y_i \neq y_k \quad y_i = y_j \quad d_{M_3}(x_i,x_k) - d_{M_3}(x_i,x_j) \geq 1 - \xi_{ijk}$$

Use **Tree of Metrics (ToM)** [1] to capture **granularity-specific** features on each taxonomy
- **Large margin metric** [2] at each node to discriminate between its subclasses
- Isolates **compact, discriminative** features with two **regularizers:**

Sparsity regularization

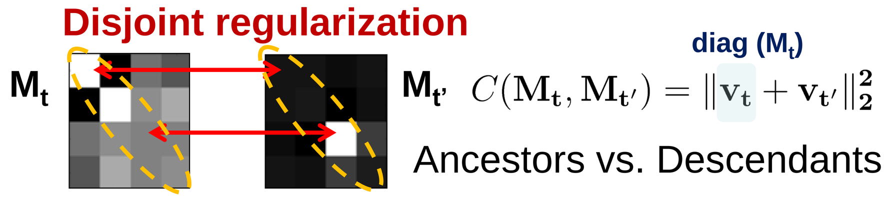


$\lambda Trace[M]$

Feature i vs. Feature j

→ **Compact** metric on informative features.

Disjoint regularization



$C(\mathbf{M}_t, \mathbf{M}_{t'}) = \|\mathbf{v}_t + \mathbf{v}_{t'}\|_2^2$

Ancestors vs. Descendants

→ **Disjoint** features at each node

Step 2: Constructing a semantic kernel forest

ToMs on multiple taxonomies → **A set of view- and granularity-specific kernels**

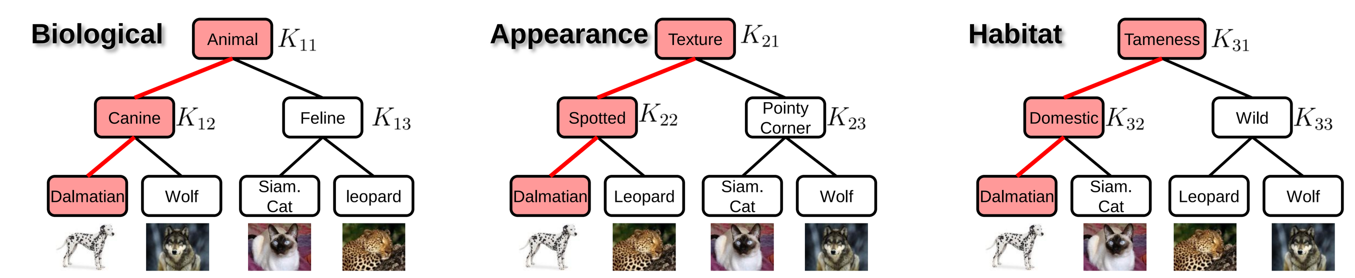
$$K_{tn}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\{-\gamma_{tn} d^2_{M_{tn}}(\boldsymbol{x}_i, \boldsymbol{x}_j)\}$$ Mahalanobis kernel

Discriminatively learning class-specific kernels

How can we then combine the learned semantic kernel forest for optimal discrimination?

Basic setting

To learn each class-specific kernel, we take kernels on its **tree-path** and find the additive kernel combination weights **β** through **multiple kernel learning.**

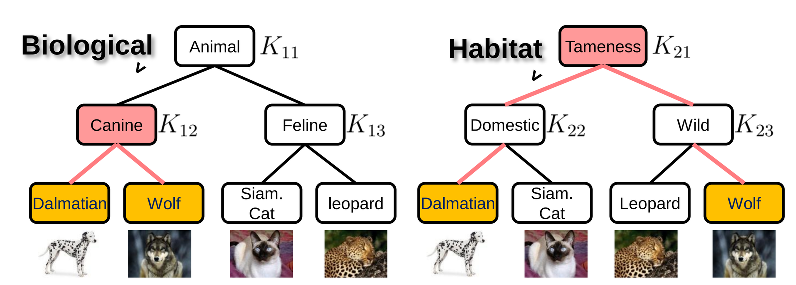


$$K_{dalmatian} = \beta_{11}K_{11} + \beta_{12}K_{12} + \beta_{21}K_{21} + \beta_{22}K_{22} + \beta_{31}K_{31} + \beta_{32}K_{32}$$

Consider only the **small fraction of kernels that are relevant**
– *O(TlogN)* kernels (*T* = # of taxonomies, *N* = # of nodes in each taxonomy)

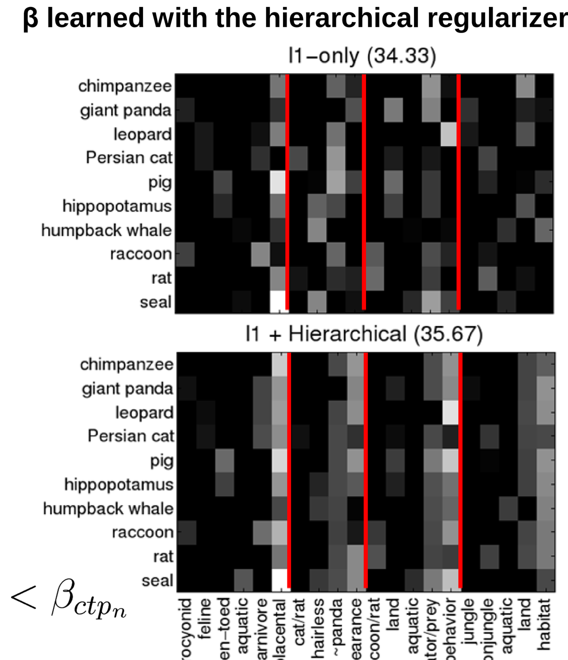
Proposed hierarchical regularization

Multiple taxonomies introduce **redundant** kernels
- need interleaved selection of kernels.



c : category
t : taxonomy
n : node
p_n : parent of n

β learned with the hierarchical regularizer



$$\Omega(\beta_c) = \lambda \sum_{t,n\sim c} \beta_{ctn} + \mu \sum_{t,n\sim c} \max(0, \beta_{ctn} - \beta_{ctp_n} + 1)$$

Sparsity Regularization Hierarchical regularization

A **hinge-loss** regularizer that favors **higher-level kernels** $\beta_{ctn} + 1 < \beta_{ctp_n}$
- More robust as trained with more training examples
- Discriminates the target class with more categories

Optimization

We minimize the sum of the MKL objective + regularization terms

$$\min_{\beta_c} \max_{\alpha_c} \sum_i \alpha_{ci} - \frac{1}{2}\sum_i \sum_j \alpha_{ci}\alpha_{cj} q_{ci} q_{cj} F_c(x_i, x_j) + \lambda \sum_{t,n\sim c} \beta_{ctn} + \mu \sum_{t,n\sim c} \max(0, \beta_{ctn} - \beta_{ctp_n} + 1)$$

MKL objective Sparsity regularization Hierarchical regularization

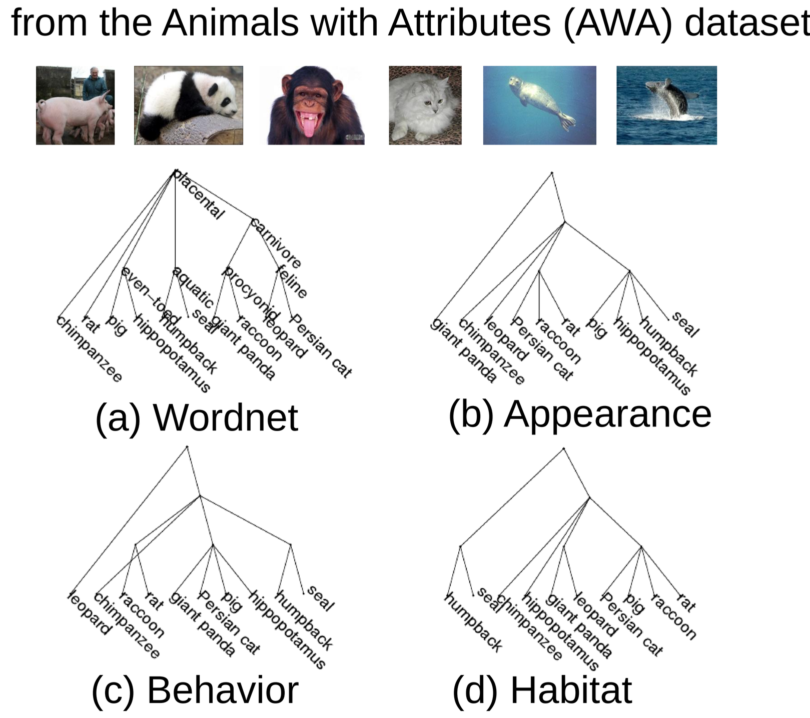
$$\text{s.t.} \quad \sum_i \alpha_{ci} q_{ci} = 0, \quad 0 \leq \alpha_{ci} \leq C$$

Convex, but **nonsmooth** due to the regularization terms.- Use **Projected subgradient** to optimize

Datasets

AWA-4 and AWA-10

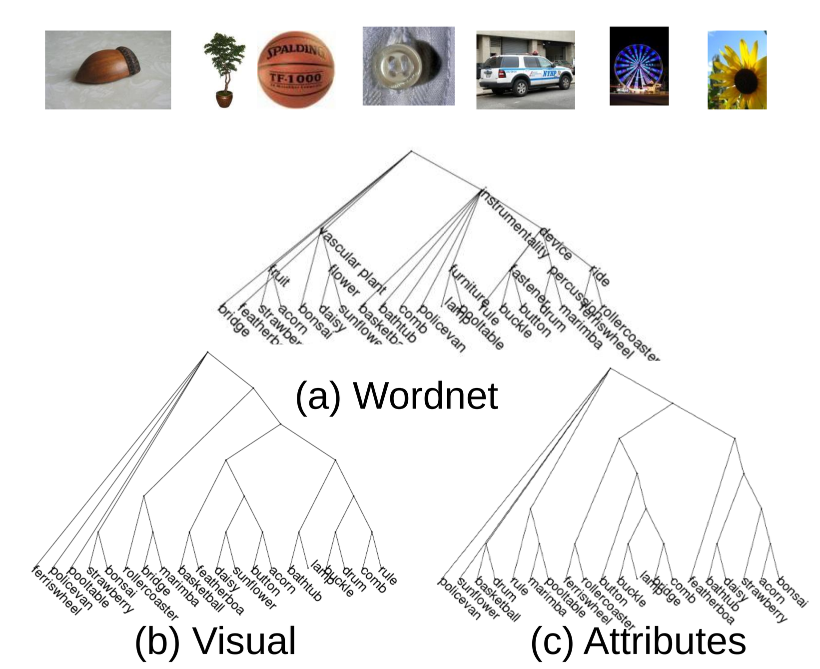
- **AWA-4**: 2,228 images of 4 example classes
- **AWA-10**: 6,180 images of 10 animal classes from the Animals with Attributes (AWA) dataset.



(a) Wordnet (b) Appearance (c) Behavior (d) Habitat

ImageNet-20

28,957 images of 20 non-animal classes from the ImageNet dataset.



(a) Wordnet (b) Visual (c) Attributes

Results

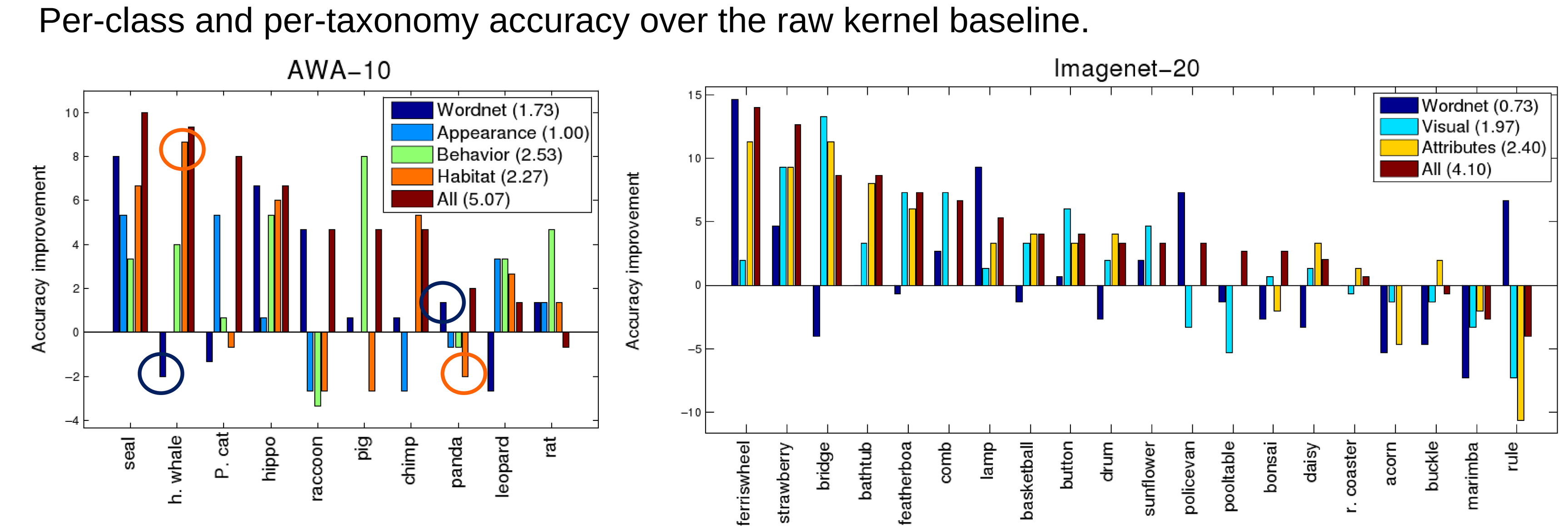
Multiclass classification

Method	AWA-4	AWA-10	Imagenet-20
Raw feature kernel	47.67 ± 2.22	30.80 ± 1.36	28.20 ± 1.45
Raw feature kernel + MKL	48.50 ± 1.89	31.13 ± 2.31	27.57 ± 1.50
Perturbed semantic kernel tree + MKL-H	N/A	31.53 ± 2.07	28.20 ± 2.02
Semantic kernel tree + Average	47.17 ± 2.40	31.92 ± 1.21	28.97 ± 1.61
Semantic kernel tree + MKL	48.89 ± 1.06	32.43 ± 1.93	29.74 ± 1.26
Semantic kernel tree + MKL-H	50.06 ± 1.12	32.68 ± 1.79	29.90 ± 0.70
Semantic kernel forest + MKL	49.67 ± 1.11	34.60 ± 1.78	30.97 ± 1.14
Semantic kernel forest + MKL-H	52.83 ± 1.68	35.87 ± 1.22	32.30 ± 1.00

1) Semantic kernel tree better than perturbed kernel tree – **semantic knowledge useful**
2) Multiple taxonomies better than using a single taxonomy – **complementary information**
3) Hierarchical regularizer improves accuracy significantly – **semantic structure useful**

Per-class and per- taxonomy results

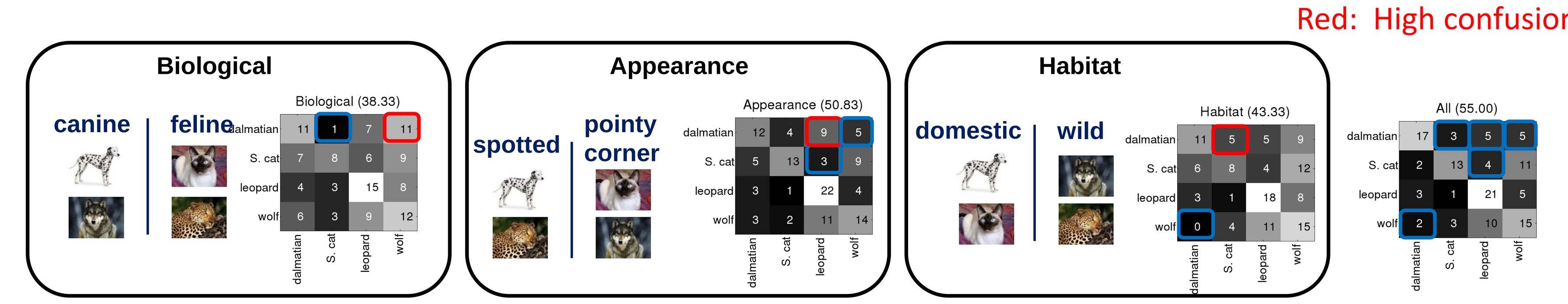
Per-class and per-taxonomy accuracy over the raw kernel baseline.



1) Our method improves accuracy on **9/10** classes for AwA-10, and **16/20** classes for Imagenet-20
2) A single semantic tree useful for some classes, but degenerates performance on others.

Confusion matrices

Blue: Low confusion
Red: High confusion



Each taxonomy-derived kernel specializes in discriminating between different sets of classes.
– Using all achieves better performance (55.00) than the best performing tree (50.83)

Conclusion and Future Work

We proposed a method to learn combined feature space from **multiple semantic taxonomies** that,
- exploits **disjoint sparsity** between parent and child classes in a taxonomy
- leverages **complementary information** from multiple **semantic views**.
- uses a novel **hierarchical** MKL regularizer to favor more **high-level** semantic grouping/splits**.**

Future Work: learn **non-additive** and **per-instance** semantic kernel combinations

References

[1] S. J. Hwang, K, Grauman, and F. Sha, Learning a Tree of Metrics with Disjoint Visual Features, NIPS 2011
[2] K.Q. Weinberger, J. Blitzer and L. K. Saul, Distance Metric Learning for Large Margin Nearest Neighbor Classification, NIPS 2006
[3] F. Bach, Exploring large feature space with hierarchical multiple kernel learning, NIPS 2008