Semantic Pooling for Image Categorization using Multiple Kernel Learning

Thibaut Durand [1,2], Nicolas Thome [1], Matthieu Cord [1], David Picard [2]

(1) Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6
(2) ETIS/ENSEA, University of Cergy-Pontoise, CNRS, UMR 8051

ICIP 2014

Outline

Outline

Supervised image classification

Goal

- Predict the label by using the data of the training set

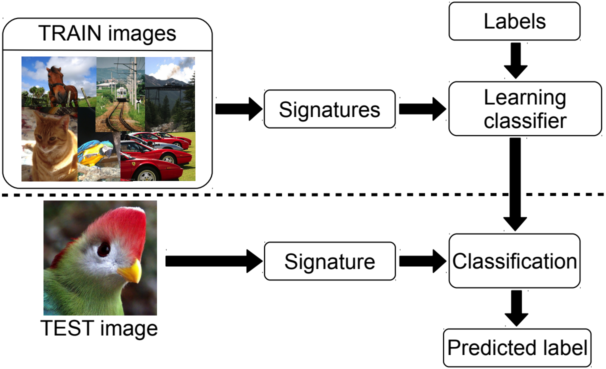
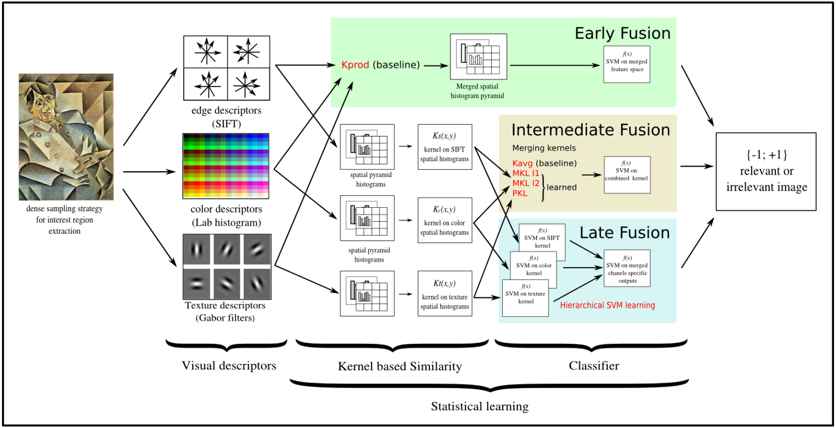


Figure: Standard pipeline

Bag of Words (BoW) model



Drawback

- Spatial information is lost

Integrated geometrical information (1)

Spatial Pyramid

- Good results on scene classification
- SP is not adapted to objects

[ECCV 2012: Russakovsky, Lin, Yu, Fei-Fei. Object-centric spatial pooling for image classification]

- SP does not encode any semantic information

Integrated geometrical information (2)

Spatial Coordinate Coding (SCC)

- Integrate the spatial coordinates of the descriptors into the codebook
- Drawback: lack of invariance with respect to the layout

[ICIP 2011: Koniusz, Mikolajczyk. *Spatial coordinate coding to reduce histogram representations, dominant angle and colour pyramid match*]

Integrated geometrical information (2)

Spatial Coordinate Coding (SCC)

- Integrate the spatial coordinates of the descriptors into the codebook
- Drawback: lack of invariance with respect to the layout

[ICIP 2011: Koniusz, Mikolajczyk. *Spatial coordinate coding to reduce histogram representations, dominant angle and colour pyramid match*]

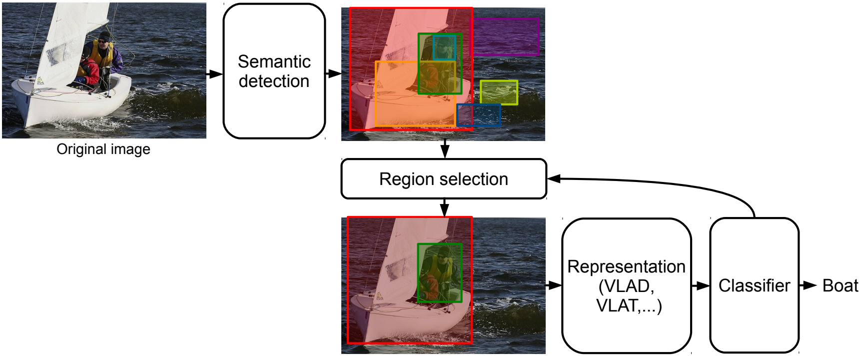
Object detectors

- Use the scores of a set of detectors to compute the signature
- Invariant signatures to the position of the object

[ICIP 2013: Durand, Thome, Cord, Avila. *Image classification using object detectors*]

Contributions

- New image categorization method using semantic pooling regions
- **Semantic pooling region detection**
- **Class-wise selection** (MKL)



Outline

Semantic Pooling with MKL

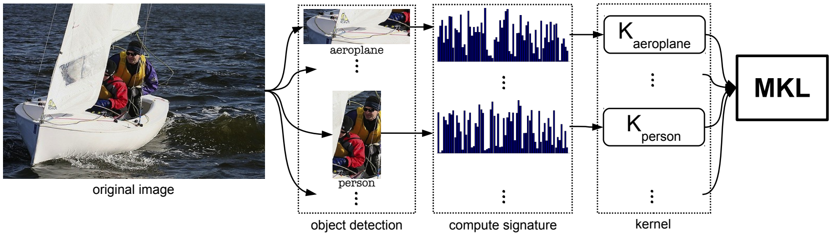


Figure: SemanticMKL pipeline

1. Object detection
2. Image representation
3. Region selection and classification

1 - Object detection

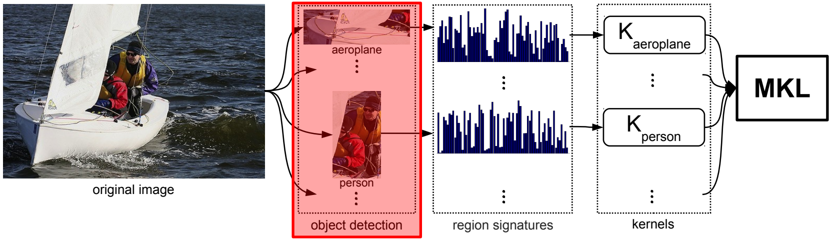


Figure: SemanticMKL pipeline

1 - Object detection: Latent SVM object detector

- Sliding window approach
- Works as a classifier: predict if an object is present in a certain position and scale in an image

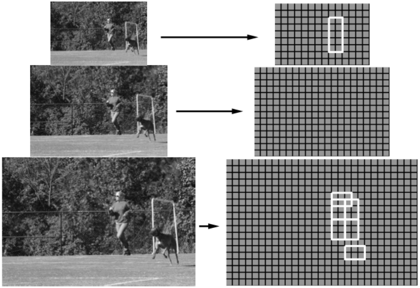


Image pyramid Feature pyramid

[PAMI 2010 : Felzenszwalb, Girshick, McAllester, Ramanan. *Object detection with discriminatively trained part based models*]

1 - Object detection: Latent SVM object detector

Spatial Pyramid

Semantic pooling regions

Selected semantic pooling regions *(details part 3)*

Boat
Car
Horse
Motorbike
Person
Pottedplant
Sofa

Figure: Examples of pooling regions

1 - Object detection: Latent SVM object detector

Spatial Pyramid

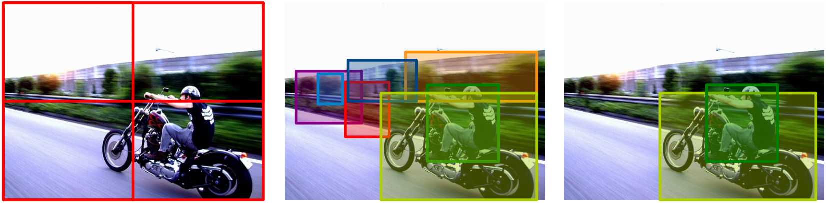
Semantic pooling regions

Selected semantic pooling regions *(details part 3)*

Boat	Motorbike	Sofa
Car	Person	
Horse	Pottedplant	

Figure: Examples of pooling regions

1 - Object detection: Latent SVM object detector



Spatial Pyramid

Semantic pooling regions

Selected semantic pooling regions *(details part 3)*

Boat
Car
Horse
Motorbike
Person
Pottedplant
Sofa

Figure: Examples of pooling regions

2 - Image representation

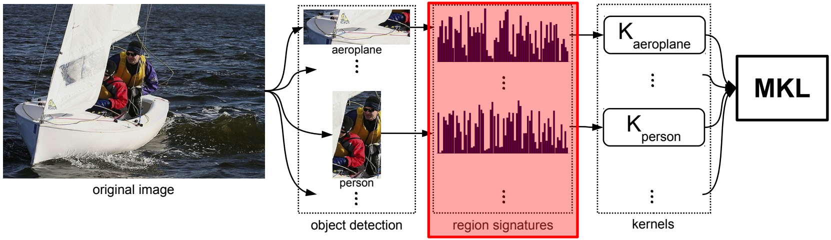


Figure: SemanticMKL pipeline

2 - Image representation: VLAT [ICIP 2011]

- Extension of the VLAD approach
- Vector image representation based on the aggregation of tensor products of local descriptors

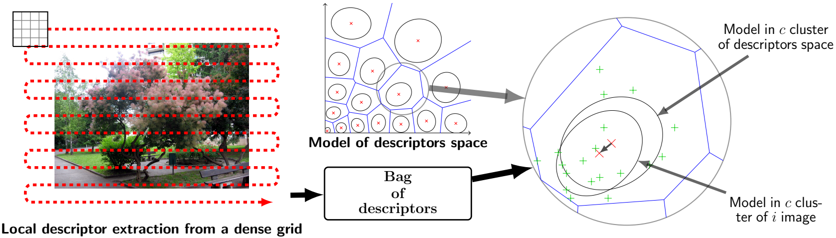


Figure: VLAT pipeline

[ICIP 2011: Picard, Gosselin. *Improving Image Similarity With Vectors of Locally Aggregated Tensors*]

3 - Region selection and classification

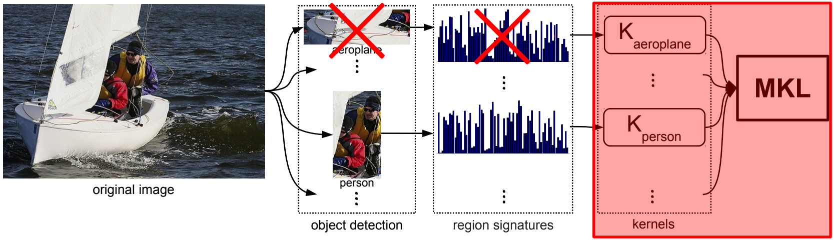


Figure: SemanticMKL pipeline

3 - Kernel selection

Figure: Semantic pooling region for object *sofa*

- Many signatures represent “noise”
- Aggregation of background local descriptors
- Selection of the relevant regions

3 - Kernel selection: ℓ_1-Multiple Kernel Learning (MKL)

Similarity between two regions:

- $\mathcal{R}$ pooling region
- $\phi_{\mathcal{R}}$ function computing a signature (VLAT) for region $\mathcal{R}$
- Definition of an explicit kernel function $k_{\mathcal{R}}(\cdot,\cdot)$ measuring the similarity between two images i and j:

$$k_{\mathcal{R}}(i,j) = \langle \phi_{\mathcal{R}}(\mathbf{B}_{\mathcal{R}i}), \phi_{\mathcal{R}}(\mathbf{B}_{\mathcal{R}j}) \rangle \tag{1}$$

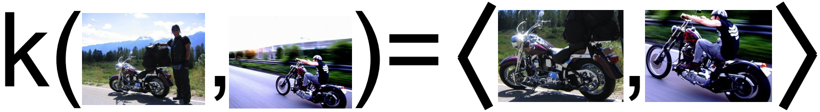


Figure: Kernel for $\mathcal{R}$ = motorbike

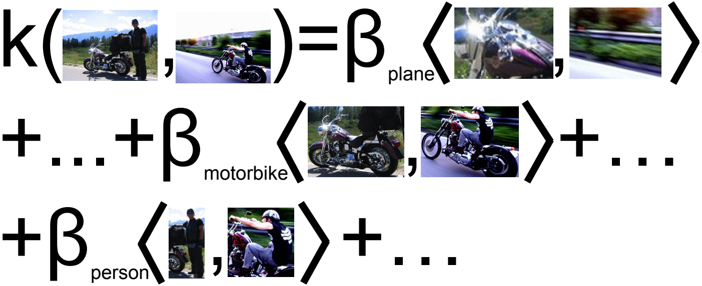
3 - Kernel selection: ℓ_1-Multiple Kernel Learning (MKL)

Similarity between two images:

- Linear combination of the kernel corresponding to the associated pooling regions:

$$k(i,j) = \sum_{\mathcal{R}} \beta_{\mathcal{R}} k_{\mathcal{R}}(i,j) \tag{2}$$

$\beta_{\mathcal{R}}$ the weights associated with each pooling region $\mathcal{R}$



3 - Kernel selection: ℓ_1-Multiple Kernel Learning (MKL)

- Learn the weights associated with each kernel using MKL
- SimpleMKL algorithm
- ℓ_1 norm constraint enforces sparsity $\rightarrow$ **kernel selection**
- Learning jointly the classifier and the kernel combination

Optimization problem

$$\min_{\beta} \max_{\alpha} \sum_{i} \alpha_i - \frac{1}{2} \sum_{i,i} \alpha_i \alpha_i y_i y_i \sum_{\mathcal{R}} \beta_{\mathcal{R}} k_{\mathcal{R}}(i,j) \tag{3}$$

$$\text{s.t.} \quad \forall \mathcal{R}, \quad \beta_{\mathcal{R}} \geq 0, \quad \sum_{\mathcal{R}} \beta_{\mathcal{R}} = 1, \quad \forall i, \quad 0 \leq \alpha_i y_i \leq C \tag{4}$$

[JMLR 2008: Rakotomamonjy, Bach, Canu, Grandvalet. SimpleMKL]

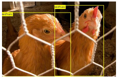
Outline

1. Context
2. Model
3. Experiments

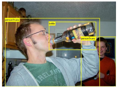
Dataset - Pascal VOC 2007 - 20 classes

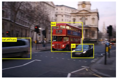

aeroplane


bicycle


bird

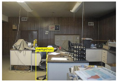

boat

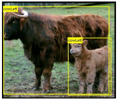

bottle

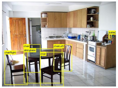

bus

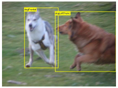

car

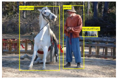

cat


chair

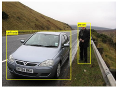

cow

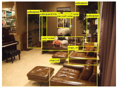

diningtable

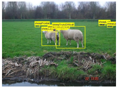

dog


horse


motorbike

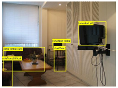

person


pottedplant


sheep


sofa


train


tvmonitor

Setup

- HOG descriptors sampled every 3 pixels at 4 scales
- Visual codebook: 64 visual words
- **Compressed VLAT**: final dimension 8192 (code available at `www.vlat.fr`)
- Spatial pooling: $1 \times 1, 2 \times 2, 3 \times 1$
- Semantic pooling: detectors trained on the *trainval* set of Pascal VOC 2007 (20 detectors = 20 classes)
- ℓ_1-MKL using **JKernelMachines** (code available on github)

[JMLR 2013: Picard, Thome, Cord, *JKernelMachines: A simple framework for kernel machines*]

Results

- Without kernel selection

	VLAT	pVLAT	sVLAT
mAP (%)	57.9	59.0	58.4

Table: Results VOC 2007 mean Average Precision (mAP)

VLAT : VLAT without spatial pyramid
pVLAT : VLAT with spatial pyramid $1 \times 1, 2 \times 2, 3 \times 1$ (concatenation)
sVLAT : VLAT with semantic pooling (concatenation)

- Straightforward combination of the signatures does not work
- Many signatures represent “noise”

Results

	Without selection			With selection		
Method	VLAT	pVLAT	sVLAT	pMKL	sMKL	spMKL
mAP (%)	57.9	59.0	58.4	59.7	63.2	**64.0**

Table: Resultats VOC 2007 mean Average Precision (mAP)

VLAT : VLAT without spatial pyramid
pVLAT : VLAT with spatial pyramid $1 \times 1, 2 \times 2, 3 \times 1$
sVLAT : VLAT with semantic pooling
pMKL: MKL with spatial pooling $1 \times 1, 2 \times 2, 3 \times 1$
sMKL: MKL with semantic pooling
spMKL: MKL with spatial and semantic pooling

- Filter out the objects which are uncorrelated with the considered category

Results

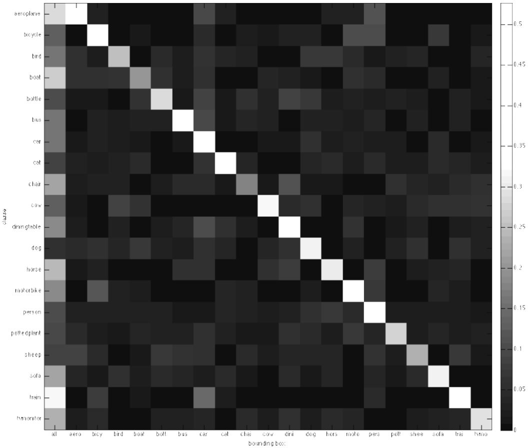


Figure: Learned semanticMKL weights (row → category, column → region, first column → whole image)

Results

- Correlation between classes

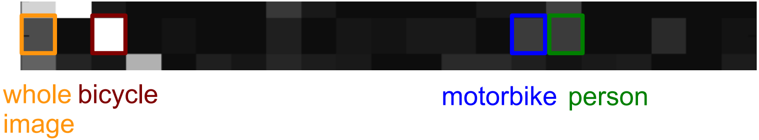


Figure: Learned semanticMKL weights for bicycle category

Conclusion

- New image categorization system based on a **semantic pooling regions**
- Take into account the layout of the images
- **Selection** of the relevant detectors with respect to a specific category

Thank you for your attention!

Questions?

Thibaut Durand(1,2) thibaut.durand@lip6.fr
Nicolas Thome(1) nicolas.thome@lip6.fr
Matthieu Cord(1) matthieu.cord@lip6.fr
David Picard(2) picard@ensea.fr

(1) Sorbonne Universités, UPMC Univ Paris 06, UMR 7606, LIP6
(2) ETIS/ENSEA, University of Cergy-Pontoise, CNRS, UMR 8051

Code available on demand