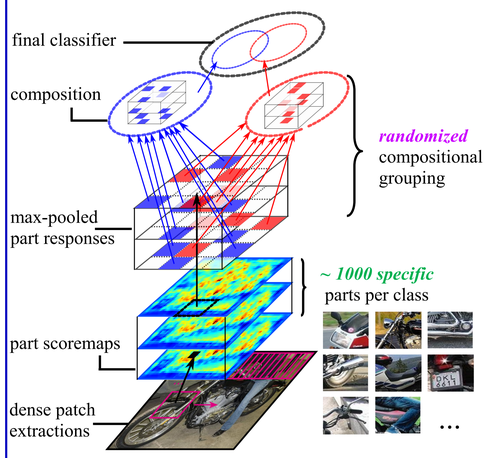
Randomized Max-Margin Compositions for Visual Recognition

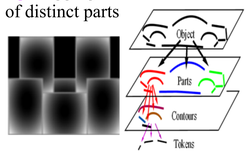
Angela Eigenstetter[1] Masato Takami[1,2] Björn Ommer[1]

[1]IWR & HCI, University of Heidelberg, Germany
[2] Robert Bosch GmbH, Hildesheim, Germany
[aeigenst,mtakami,bommer]@iwr.uni-heidelberg.de

1 Motivation

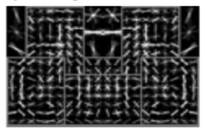
Randomized Max-Margin Compositions RM^{2}C vs. **part based approaches DPM, BoP, HOP ...**





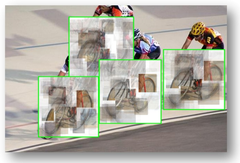
rigid aggregation of distinct parts

small (~8) number of *generic* parts

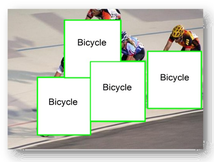


best approaches on PASCAL (comp3)
parts → object

RM^{2}C Parsing vs. **Standard Object Detection**



- Bounding Box
- Class Label
- **Parsing**



- Bounding Box
- Class Label

3 Randomized Parts & Compositions

Learning Parts without Part Annotations

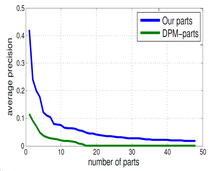
- without extra annotation as difficult as finding the object
- clustering parts based on featuers (e.g HOG) is not reliable
- to avoid incorrect groups we train parts using *single positive* patches which are randomly sampled from the training data

$$\min_{\omega} \frac{1}{2}\|\omega_i\|_2^2 + C_1 \max(0, 1 - h_i(x_p)) + C_2 \sum_{x \in \mathcal{N}} \max(0, 1 + h_i(x))$$

hard negatives

➡ decision function $h_i(x) = \omega_i^T x + \beta_i$

Part Evaluation



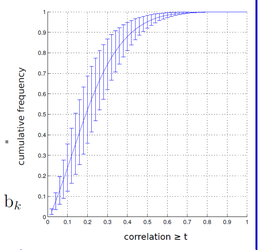
measure individual part performance using keypoint annotations

comparison of the 48 DPM parts with a subset of 48 of our parts

Randomized Max-Margin Compositions RM^{2}C

#parts (~1000) x #sites (21) ➡ > 20000 D ➡ **curse of dimensionality**

parts are **highly uncorrelated** ➡ subspace methods / part grouping are not suitable for dim reduction

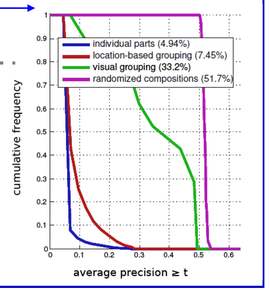


compositional classifier (SVM training)

$$\min_{\mathrm{w}} \frac{1}{2}\|\mathrm{w}\|_2^2 + C \sum_{\mathcal{I} \in \mathcal{T}} \max(0, 1 - y_{\mathcal{I}} f_k(\gamma_k(\mathcal{I})))$$

➡ decision function $f_k(\gamma_k(\mathcal{I})) = \mathrm{w}_k^T \gamma_k(\mathcal{I}) + \mathrm{b}_k$

evaluation of different grouping strategies on the validation set
randomized compositions clearly outperform other grouping strategies



concatenation of the K decision values ➡ composition with representation $F(\cdot) = (f_1(\cdot), ..., f_K(\cdot))^\top$

final non-linear classifier with $\kappa(F(\mathcal{I}), F(\mathcal{I}')) = \exp(-\frac{\|F(\mathcal{I}) - F(\mathcal{I}')\|_2^2}{2\sigma^2})$

optimize $\max_{\alpha} \sum_{\mathcal{I} \in \mathcal{T}} \alpha_{\mathcal{I}} - \frac{1}{2} \sum_{\mathcal{I} \in \mathcal{T}} \sum_{\mathcal{I}' \in \mathcal{T}} \alpha_{\mathcal{I}} \alpha_{\mathcal{I}'} y_{\mathcal{I}} y_{\mathcal{I}'} \kappa(F(\mathcal{I}), F(\mathcal{I}'))$

➡ decision function $g(F(\mathcal{I})) = \sum_{\mathcal{I}' \in \mathcal{T}} \alpha_{\mathcal{I}'} y_{\mathcal{I}'} \kappa(F(\mathcal{I}), F(\mathcal{I}'))$

4 Experimental Results

Object Detection Results PASCAL VOC 2007 (comp3)

	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	**mean**
DPM rel5 [12]	45.6	49.0	11.0	11.6	27.2	50.5	43.1	23.6	17.2	23.2	10.7	20.5	42.5	44.5	41.3	8.7	29.0	18.7	40.0	34.5	29.6
Poselets [3]	33.2	**51.0**	8.5	8.2	**34.8**	39.0	**48.8**	22.2	-	20.6	-	18.5	**48.2**	44.1	**48.5**	9.1	28.0	13.0	22.5	33.0	-
BCP [8]	44.3	35.2	9.7	10.1	15.1	44.6	32.0	**35.3**	4.4	17.5	**15.0**	**27.6**	36.2	42.1	30.0	5.0	13.7	18.8	34.4	28.6	25.0
AOT [30]	44.6	48.5	10.8	12.9	26.3	47.5	41.6	21.6	**17.3**	23.6	11.5	22.9	40.9	45.3	37.9	9.6	**30.4**	25.3	39.0	31.2	29.4
RM^{2}C	**49.8**	50.6	**15.1**	**15.5**	28.5	**51.1**	42.2	30.5	**17.3**	**28.3**	12.4	26.0	45.6	**51.8**	41.4	**12.6**	**30.4**	**26.1**	**44.0**	**37.6**	**32.8**

Object Detection Results PASCAL VOC 2010 (comp3)

	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	**mean**
DPM rel5 [12]	33.2	60.3	10.2	16.1	27.3	54.3	**58.2**	23.0	20.0	24.1	26.7	12.7	58.1	48.2	**43.2**	12.0	21.1	36.1	46.0	43.5	33.7
LHS [32]	29.4	55.8	9.4	14.3	28.6	44.0	51.3	21.3	20.0	19.3	25.2	12.5	50.4	38.4	36.6	15.1	19.7	25.1	36.8	39.3	29.6
AOT [30]	35.3	60.2	9.4	16.6	29.5	53.0	57.1	23.0	22.9	27.7	28.6	13.1	58.9	49.9	41.4	**16.0**	22.4	37.2	48.5	42.4	34.7
RM^{2}C w/o obj	37.0	58.3	12.0	14.7	22.9	51.3	51.7	23.7	21.7	25.0	29.0	20.6	51.4	46.1	36.3	12.7	22.3	35.1	43.9	41.8	32.9
RM^{2}C	**37.7**	**61.4**	**12.7**	**17.6**	**29.9**	**55.1**	56.3	**29.5**	**24.6**	**28.2**	**30.7**	**21.2**	**59.5**	**51.5**	40.3	14.3	**23.9**	**41.6**	**49.2**	**46.0**	**36.6**

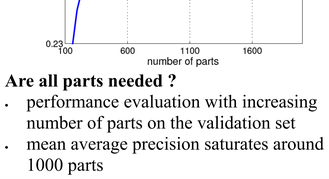
Results on the MIT Indoor Dataset

Method	Acc. (%)	Mean AP
Object Bank [20]	37.60	-
RBoW [26]	37.93	-
DPM+GIST-color+SP [25]	43.10	-
Patches+GIST+SP+DPM [25]	49.40	-
IFV+BoP [16]	63.10	63.18
Mid-Level Patches [29]	38.10	-
BoP [16]	46.10	43.55
RM^{2}C	51.34	46.70

Project Website

parts for all datasets are available for download here :

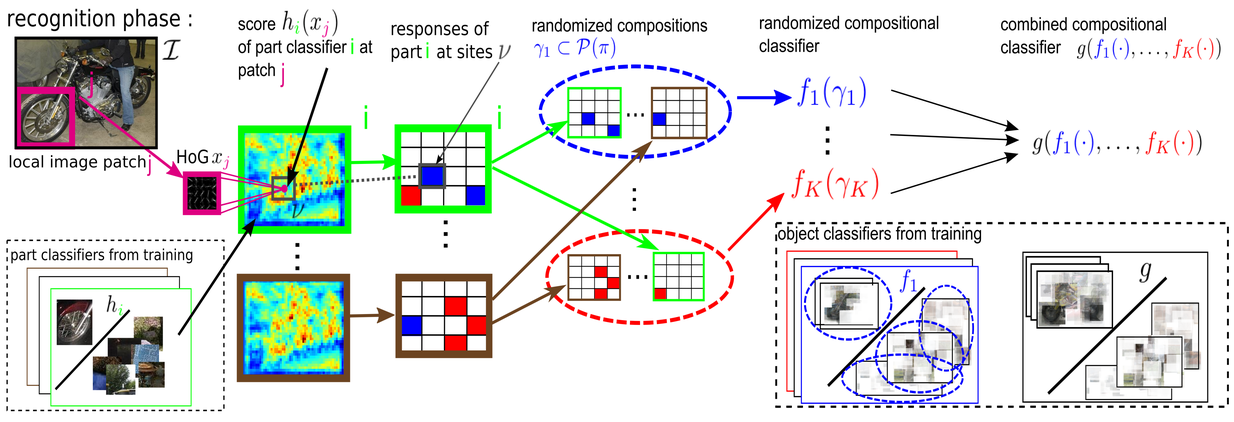
Part Selection (PASCAL VOC 2010)



Are all parts needed ?

- performance evaluation with increasing number of parts on the validation set
- mean average precision saturates around 1000 parts

2 Detection Procedure



5 Object Parsing Results