CNRS TELECOM ParisTech at ImageCLEF 2015 Scalable Concept Image Annotation Task: Concept Detection with Blind Localization Proposals

Hichem SAHBI

CNRS TELECOM ParisTech
hichem.sahbi@telecom-paristech.fr

Abstract. We introduce our participation at the ImageCLEF 2015 scalable concept detection and localization task. This edition focuses on generating not only annotations (concept detections) but also localizing concepts into a large image collection.
Concept detection part of our runs is based on standard nonlinear support vector machines (SVMs). The localization part is blind and based on a priori learned statistics that generate multiple localization proposals. In spite of its blindness, the performance of this concept localization framework is promising.

Keywords: Support vector machines, histogram intersection kernels, concept detection, blind concept localization proposals

1 Introduction

The general problem of *visual category recognition* generally includes three different tasks: concept detection (also known as image annotation) [1–3], concept localization [4, 5] and object category segmentation [6, 7]. Concept detection consists in inferring a list of keywords that best describes the visual and the semantic content of a given image while localization seeks to find a list of bounding boxes that defines the span of detected concepts. As a variant of concept localization, object category segmentation consists in delimiting the extent of detected concepts with a high precision.

We are interested in this paper in concept detection and localization; we present our solutions submitted to the ImageCLEF 2015 scalable concept image annotation task [8, 9]. This edition focuses on concept localization, which consists in finding all the *occurrences of a list of concepts* into a given test image. This task has been widely studied in different related challenges including Pascal VOC [10], ImageNET [11] and more recently MS-COCO [12]. Existing solutions usually parse images using sliding windows [13], image segmentation and superpixels [14] as well as multiple segmentation proposals [15]. In these methods, detection and segmentation results are scored using machine learning techniques (such as SVM [16], deep networks [17], and decision forests [18]) and

Fig. 1. Sample of pictures taken from the ImageCLEF2015 database.

consolidated using spatial layout and geometric relationships usually described with graphical models (such as conditional and Markov random fields [19, 20]). For more detailed discussions of related work in concept detection and localization, see [21] and references therein.

Among existing object localization (and segmentation) methods those based on region proposals are currently receiving a particular attention. Their general principle consists in defining multiple partitions of test images into sets of blobs that potentially correspond to actual objects. Only few of these partitions are scored and used to annotate and localize concepts in test images. Even though relatively successful, these approaches are highly dependent on the quality of image segmentation, which is known to be challenging especially when no a priori information is used about the statistics of these concept-localization proposals. Our proposed solution, discussed in this paper, avoids image segmentation and it is based on two steps: first, we train SVM classifiers that detect concepts belonging to different test images. Afterwards, we use an priori (trained) statistical model in order to infer their most likely locations, *without* observing the content of these test images. We will show that in spite of the simplicity of this approach, the results are reasonably decent, and very promising, and this opens a new direction towards refining these models and obtaining better performances by combining annotation and concept localization results.

The rest of this paper is organized as follows; first, we describe our concept detection algorithm, based on SVMs and an efficient evaluation of the histogram intersection kernel. Then, we describe our a priori statistical model for blind concept localization, and we present and discuss our ImageCLEF 2015 results. Finally, we conclude the paper, with possible extensions for a future work.

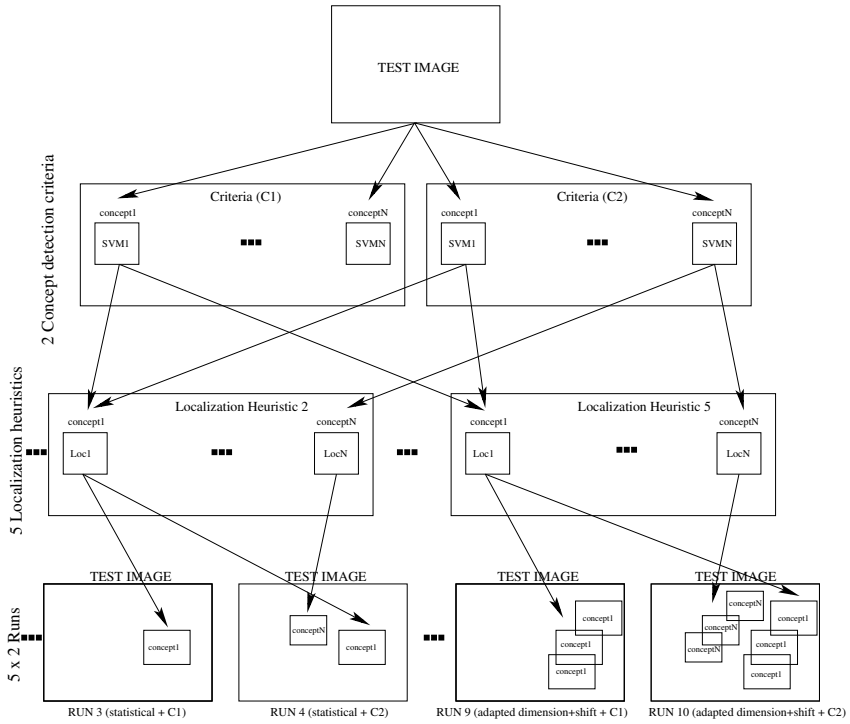

Fig. 2. This figure shows the two-step process used for concept detection and localization

2 Concept Detection with Blind Localization Proposals

Our concept detection and localization results are obtained according to the two following steps (see Fig. 2):

i) Holistic concept detection: this step is achieved using global (holistic) visual and textual features. For that purpose, we train "one versus all" SVMs for each concept, in order to detect whether that concept exists in a given test image (see extra details in Section 2.1).

ii) Blind concept localization proposals: in contrast to concept detection, concept localization is achieved *blindly*, i.e., without observing the content of a given test image. As will be shown subsequently, localization is achieved using a priori knowledge about possible locations of these bounding boxes. These knowledges correspond to learned localization statistics, of bounding boxes, taken from a training/dev set of concepts and their associated bounding boxes (i.e., from the file "imageclef2015.dev.bbox.v20150226"; see extra details in Section 2.2).

2.1 Holistic concept detection: training and classification

We used only the *holistic* features provided in this ImageCLEF task including GIST, Color Histograms, SIFT, C-SIFT, RGB-SIFT, OPPONENT-SIFT, etc. We build 10 gram matrices (9 visual and 1 textual), based on efficient histogram intersection kernel, associated to these features. Then, we linearly combine those matrices into a single one. Notice that this combination does not result from multiple kernel learning but just a convex combination of kernels with uniform weights. We plug the resulting kernel into SVMs for training and testing.

For each concept, we train "one-versus-all" SVM classifiers; we use many random folds (taken from training/dev data in "imageclef2015.dev.bbox.v20150226") for multiple SVM training and we use these SVMs in order to predict the concepts on the test set[1]. We repeat this training process, for each concept, through different random folds from the training set and we take the average scores of the underlying SVM classifiers. This makes classification results less sensitive to the sampling of the training set and also allows us to re-balance classification results mainly for concepts with unbalanced distributions of positive and negative data.

2.2 Blind concept localization proposals

Several heuristics are tried in order to suggest *multiple concept localization proposals*. Given a test image and the list of concepts attached to it (see Section 2.1), concept localization is achieved without consulting the content of the test image (but only its detected concepts). Indeed, concept localization is blind and bounding boxes (BBs) are either fixed (using test image dimensions) or based on statistics estimated offline on the training/dev set (in "imageclef2015.dev.bbox.v20150226") as described subsequently.

In what follows, $p_c = (x, y, w, h)$ denotes the bounding box coordinates of a given detected concept c in a given test image; here (x, y) (resp. (w, h)) corresponds to the center (resp. dimensions; width and height) of the bounding box p_c.

Heuristic 1 (fixed BBs): for a detected concept c in a given test image, its bounding box p_c is set to $(W/2, H/2, W, H)$; here W and H respectively denote the width and the height of the test image. In what follows, we consider that all the test images are re-sized and have the same dimensions (i.e., W, H are constant for all the test images)[2].

In the subsequent heuristics (2–5), we introduce the following notation: given a concept c, we consider $\mathcal{T}_c = \{p_c^i\}_i$ as the union of all the bounding boxes (in the training/dev set "imageclef2015.dev.bbox.v20150226") that belong to c. We also

[1] A given test image is assigned to a given concept, iff the underlying SVM score is positive.

[2] Of course, the actual dimensions of the test images are taken into account in order to re-scale concept localization results.

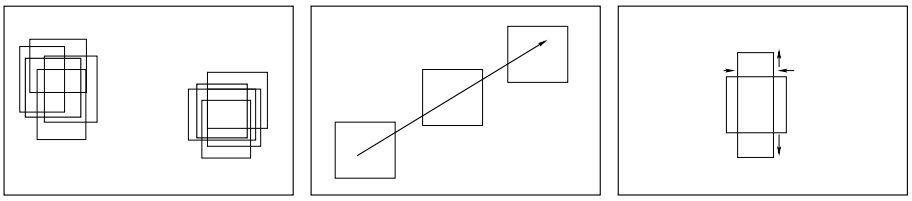

Fig. 3. This figure shows: (left) BB clustering process and the union of bounding boxes in the training set that belong to a given concept, (middle) BB shift using the first principal direction and (right) BB re-scale using the first principal direction.

consider N_c as the average number of bounding boxes (par image) associated to c; N_c is evaluated from the training set. Prior to use the following heuristics (2–5), we consider an offline step that clusters the coordinates in $\mathcal{T}_c$ (using k-means) with a number of clusters fixed to N_c (see Fig. 3, left).

Heuristic 2 (concept-dependent BBs): for a detected concept c in a given test image, we generate N_c bounding boxes whose coordinates correspond to the cluster centers obtained after applying k-means on $\mathcal{T}_c$.

In the remaining three heuristics (3–5), we update the coordinates of the bounding boxes, by manipulating i) their dimensions in heuristic 3, ii) their centers in heuristic 4, and iii) both their centers and dimensions in heuristic 5.

Heuristic 3 (re-scaled concept-dependent BBs): each bounding box $p_c = (x, y, w, h)$ generated in heuristic 2, is replaced by re-scaled BB. First, principal component analysis (PCA) is applied offline to the BB dimensions $\{(w_i, h_i)\}_i$ in the training set that also belong to concept c, afterwards, the dimensions (w, h) of p_c are moved towards the first principal component of PCA[3], with an amplitude proportional to its eigenvalue (and this corresponds a re-scale of the dimensions of p_c). In this heuristic (x, y) remains unchanged (see Fig. 3, right).

Heuristic 4 (shifted concept-dependent BBs): for each bounding box $p_c = (x, y, w, h)$ generated in heuristic 2, we generate two extra BBs, with shifted coordinates. Again, PCA is applied offline to the BB coordinates $\{(x_i, y_i)\}_i$ in the training set that also belong to concept c, afterwards, the (x, y) coordinates of p_c are shifted towards two opposite directions corresponding to the first principal component of PCA. In this heuristic (w, h) remains unchanged (see Fig. 3, middle).

Heuristic 5 (shifted and re-scaled concept-dependent BBs): this heuristic corresponds to the combination of the two heuristics 3 and 4.

[3] i.e., the eigenvector with the largest eigenvalue.

3 ImageCLEF 2015 Evaluation

The targeted task is, again, concept detection and localization: given a picture, the goal is to predict which concepts (classes) are present into that picture and a proposal of bounding boxes surrounding these concepts.

3.1 ImageCLEF 2015 Collection

A very large amount of images was gathered by the organizers, and using associated web pages, tags and meta-data were also provided. This set includes 500k images with only 2k images with known ground truth (i.e., labels and bounding boxes are given). These images belong to 251 concepts (see example in Fig. 1). Each image is again described with nine holistic visual features provided by the organizers, and we compute one extra textual feature using a normalized vector space model; first, a vocabulary of keywords $\mathcal{V}$ is defined[4] in order to query the associated meta-data that include 500k textual descriptions. For each keyword $\omega \in \mathcal{V}$, only images whose textual descriptions include ω have their ω vector entry set to non-zeros.

3.2 Submitted Runs

All our submitted runs are based on SVM training and classification with the same kernel function (i.e., histogram intersection kernel), and the differences reside in the used decision criteria for concept detection and localization. Our ten submitted runs correspond to the combination of the five concept localization heuristics described earlier (see Section 2.2) and the two following concept detection criteria

i) Criterion 1 (C1): the first concept detection results are obtained by following the setting in Section 2.1.

ii) Criterion 2 (C2): the second set of concept detection results is obtained using a slightly different criterion; more precisely, if an image has no detected concepts, i.e., all the SVM scores are negatives for all concepts, then we select the top 3 concepts (i.e., with the highest negative SVM scores) as annotations for that image. This makes it possible to increase the recall, with a possible impact on the precision.

Our runs are summarized in table 1. For all the submitted runs, performances are evaluated, by the organizers, using a variant of the Jaccard measure; the latter is defined as the intersection over union of bounding boxes provided in the submitted runs and those in the ground truth. Mean average precision (MAP) measures based on different percentages of bounding box overlaps are given for each concept and also averaged through different concepts (see our results in

[4] Including relevant keywords that are used in concept definitions.

Table 1: This table shows the definition of the ten runs submitted to the ImageCLEF 2015 challenge.

	Heuristic 1	**Heuristic 2**	**Heuristic 3**	**Heuristic 4**	**Heuristic 5**
Criterion 1 (C1)	Run 1	Run 3	Run 5	Run 7	Run 9
Criterion 2 (C2)	Run 2	Run 4	Run 6	Run 8	Run 10

Table 2: Performances (in %) of our different concept detection and localization proposal heuristics sorted from the highest to the lowest (taken from ImageCLEF 2015 results).

Runs # \ Overlap	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%
5 (heuristic 3+C1)	**30.73**	**27.64**	**25.11**	**22.47**	**19.80**	**16.92**	**14.68**	**12.46**	**10.05**	**07.85**
9 (heuristic 5+C1)	**30.73**	27.21	24.87	21.69	19.01	16.58	14.08	11.65	09.32	07.62
3 (heuristic 2+C1)	**30.73**	26.13	24.48	21.52	18.72	15.82	13.01	10.30	08.03	06.32
7 (heuristic 4+C1)	**30.73**	25.73	23.50	20.58	17.77	14.61	11.80	09.38	07.38	05.55
1 (heuristic 1+C1)	**30.73**	26.11	20.79	16.81	13.29	10.49	08.53	06.81	04.99	03.17
6 (heuristic 3+C2)	19.40	17.39	15.83	14.08	12.44	10.63	09.25	08.00	06.56	05.20
10(heuristic 5+C2)	19.40	17.21	15.71	13.83	12.10	10.38	08.90	07.52	06.10	05.01
4 (heuristic 2+C2)	19.40	16.35	15.31	13.56	11.98	10.11	08.33	06.87	05.40	04.15
8 (heuristic 4+C2)	19.40	16.23	14.91	13.17	11.41	09.48	07.83	06.16	04.92	03.68
2 (heuristic 1+C2)	19.40	16.30	13.05	10.53	08.44	06.73	05.53	04.51	03.34	02.21

Tables 2, 3). Details about these measures can be found in the ImageCLEF 2015 website[5].

From tables 2 and 3, we observe the following issues

- Different methods for "concept localization proposals" provide much better results when concept detection is relatively successful (see runs 1, 3, 5, 7, 9 vs runs 2, 4, 6, 8, 10 in table 2 for different overlap ratios). Following the spirit of our two-step method, these results clearly corroborate the fact that concept detection could be decoupled from localization as long as concept detection is achieved with a relative success. This clearly opens a direction towards enhancing the performances of localization proposals by further improving concept detection results.

- From table 2, heuristic 3 (BB re-scaling) provides the best overall performances; indeed, even though shifting is important, it has less impact on performances compared to re-scaling. This is mainly due to the variability and non-rigidity of many concepts (such as animals), that require an adaptation of the dimensions of BBs, while shifting is already well captured by the statistical model (in heuristics 2, 4, 5); see again k-mean clustering in Section 2.2.

[5] http://www.imageclef.org/2015/annotation.

Table 3: Some "concept-by concept" performances (in %) of our different concept detection and localization proposal heuristics (taken from ImageCLEF 2015 results).

concepts	description	*(run 5) statistical+rescale*	*(run 9) statistical+rescale+shift*	*(run 3) statistical*	*(run 7) statistical+shift*	*(run 1) fixed*	*(run 6) statistical+rescale*	*(run 10) statistical+rescale+shift*	*(run 4) statistical*	*(run 8) statistical+shift*	*(run 2) fixed*
n01639765	frog	18.18	**36.36**	**36.36**	27.27	18.18	18.18	27.27	45.45	27.27	18.18
n01896031	feather	20.00	20.00	**40.00**	**40.00**	20.00	20.00	20.00	**40.00**	**40.00**	20.00
n02084071	dog	**50.00**	**50.00**	**50.00**	**50.00**	**50.00**	33.33	33.33	33.33	33.33	33.33
n02114100	wolf	22.86	20.00	**28.57**	25.71	20.00	22.86	20.00	**28.57**	25.71	20.00
n02129165	lion	0	0	0	0	0	**02.22**	**02.22**	**02.22**	**02.22**	**02.22**
n02131653	bear	0	0	0	0	**50.00**	0	0	0	0	**50.00**
n02206856	bee	**66.67**	**66.67**	**66.67**	**66.67**	**66.67**	50.00	50.00	50.00	50.00	50.00
n02330245	mouse	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	50.00	50.00	50.00	50.00	50.00
n02395406	hog	40.00	40.00	**52.00**	**52.00**	36.00	40.00	40.00	**52.00**	**52.00**	36.00
n02411705	sheep	**52.94**	**52.94**	35.29	41.18	47.06	**52.94**	**52.94**	23.53	29.41	47.06
n02416519	goat	**50.00**	**50.00**	0	0	**50.00**	33.33	33.33	0	0	33.33
n02430045	deer	**25.00**	**25.00**	**25.00**	**25.00**	0	**25.00**	**25.00**	**25.00**	**25.00**	0
n02484322	monkey	57.14	57.14	**64.29**	**64.29**	50.00	52.94	52.94	58.82	58.82	47.06
n02503517	elephant	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	28.57	28.57	28.57	28.57	14.29
n02512053	fish	60.00	60.00	**70.00**	60.00	60.00	46.67	46.67	53.33	46.67	40.00
n02691156	airplane	0	**100.0**	0	**100.0**	0	0	0	33.33	33.33	0
n02709367	anchor	63.16	**73.68**	42.11	47.37	52.63	63.64	63.64	45.45	59.09	54.55
n02774152	bag	**10.00**	**10.00**	0	0	**10.00**	07.69	07.69	0	0	07.69
n02778669	ball	**16.67**	**16.67**	**16.67**	**16.67**	0	12.50	12.50	12.50	12.50	0
n02782093	balloon	0	**05.26**	0	**05.26**	0	0	**05.26**	0	**05.26**	0
n02800213	baseball	0	0	0	0	0	**01.49**	**01.49**	**01.49**	**01.49**	**01.49**
n02828884	bench	20.00	20.00	**25.00**	**25.00**	20.00	20.00	20.00	**25.00**	**25.00**	20.00
n02834778	bicycle	10.00	05.00	10.00	05.00	05.00	**13.16**	05.26	10.53	05.26	07.89
n02839910	bin	30.43	30.43	30.43	30.43	30.43	30.43	30.43	30.43	30.43	30.43
n02883344	box	0	0	0	0	0	**04.35**	**04.35**	0	**04.35**	0
n02909870	bucket	46.67	**53.33**	**53.33**	**53.33**	**53.33**	41.18	47.06	47.06	47.06	47.06
n02933112	cabinet	60.00	60.00	**80.00**	**80.00**	60.00	24.24	21.21	27.27	24.24	21.21
n02942699	camera	0	05.26	0	05.26	0	05.41	05.41	02.70	**08.11**	05.41
n02984061	cathedral	36.76	**37.50**	13.97	08.82	34.56	36.76	**37.50**	15.44	16.18	34.56
n02990373	ceiling	**36.36**	**36.36**	**36.36**	**36.36**	**36.36**	23.26	23.26	23.26	25.58	18.60
n03001627	chair	0	0	0	0	0	10.26	10.26	10.26	**15.38**	0
n03046257	clock	**11.32**	09.43	07.55	07.55	05.66	09.43	09.43	09.43	07.55	05.66
n03135532	cross	**25.00**	0	**25.00**	0	0	20.00	0	20.00	0	0

– From table 3, for almost all the concepts, statistical bounding box estimation (i.e., heuristics 2, 3, 4, 5) is very helpful in order to improve the quality of localization; for some concepts such as "frog", re-scaling and shifting are important, as this category is highly non-rigid while for other categories such as "bear", adaptation does not improve performances as "bear" localization is less predictable. Note also that for rigid (and man-made) objects, such as "cathedral" and "bicycle", re-scaling is more important than shifting as the proportions of the w-h dimensions, in these concepts, are very changing while for others (including natural objects and also some other man-made objects such as "airplane", "balloon", "bucket", "camera"), the adaptation of shift is more important than scale; as the variability of w-h proportions is small in these concepts. In sum, bounding box re-scaling and shifting is important for some concepts and less for others. This suggests, as a future extension, to mix different heuristics for different concepts (and we already observe this improvement in the "concept-by-concept" results).

4 Conclusion

We discussed in this paper, our participation at the ImageCLEF 2015 Scalable Concept Image Annotation Task. Our runs are based on a two-step process that decouples concept detection from localization. The former is achieved using SVMs trained with linear combination of elementary histogram intersection kernels, while the latter is accomplished blindly using a simple statistical model that allows us to generate multiple localization proposals (without image segmentation). Observed results show that i) the accuracy of concept detection has an impact on the performance of localization, and ii) the adaptation of scale and shift of concept localization is essential to improve performances mainly for concepts with a large variability in their extents.
A future possible extension, of this work, is to make concept localization non-blind and also coupled with concept detection. Another possible extension is to mix and select different localization heuristics for different concepts.

Acknowledgments. This work is supported in part by a grant from the French Research Agency ANR (Agence Nationale de la Recherche) under the MLVIS project.

References

1. V. Lavrenko, R. Manmatha, and J. Jeon, "A model for learning the semantics of pictures," *In: Proc. of NIPS*, 2004.
2. Jia Li and James Ze Wang, "Real-time computerized annotation of pictures," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 6, pp. 985–1002, 2008.
3. H. Sahbi and X. Li, "Context based support vector machines for interconnected image annotation (the saburo tsuji best regular paper award)," *In the Asian Conference on Computer Vision (ACCV)*, 2010.

4. Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 580–587.
5. Olga Barinova, Victor Lempitsky, and Pushmeet Kholi, "On detection of multiple object instances using hough transforms," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 9, pp. 1773–1784, 2012.
6. X. Li and H. Sahbi, "Superpixel based object class segmentation using conditional random fields," *In the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2011.
7. Jian Yao, Sanja Fidler, and Raquel Urtasun, "Describing the scene as a whole: Joint object detection, scene classification and semantic segmentation," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 702–709.
8. Andrew Gilbert, Luca Piras, Josiah Wang, Fei Yan, Emmanuel Dellandrea, Robert Gaizauskas, Mauricio Villegas, and Krystian Mikolajczyk, "Overview of the ImageCLEF 2015 Scalable Image Annotation, Localization and Sentence Generation task," in *CLEF2015 Working Notes*, Toulouse, France, September 8-11 2015, CEUR Workshop Proceedings, CEUR-WS.org.
9. Mauricio Villegas, Henning Müller, Andrew Gilbert, Luca Piras, Josiah Wang, Krystian Mikolajczyk, Alba García Seco de Herrera, Stefano Bromuri, M. Ashraful Amin, Mahmood Kazi Mohammed, Burak Acar, Suzan Uskudarli, Neda B. Marvasti, José F. Aldana, and María del Mar Roldán García, "General Overview of ImageCLEF at the CLEF 2015 Labs," Lecture Notes in Computer Science. Springer International Publishing, 2015.
10. Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision*, vol. 88, no. 2, pp. 303–338, 2010.
11. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 248–255.
12. Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick, "Microsoft coco: Common objects in context," in *Computer Vision–ECCV 2014*, pp. 740–755. Springer, 2014.
13. Paul Viola and Michael Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. IEEE, 2001, vol. 1, pp. I–511.
14. Xuming He, Richard S Zemel, and MA Carreira-Perpindn, "Multiscale conditional random fields for image labeling," in *Computer vision and pattern recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE computer society conference on*. IEEE, 2004, vol. 2, pp. II–695.
15. Pekka Rantalankila, Juho Kannala, and Esa Rahtu, "Generating object segmentation proposals using global and local search," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 2417–2424.
16. Tomasz Malisiewicz, Abhinav Gupta, and Alexei A Efros, "Ensemble of exemplar-svms for object detection and beyond," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 89–96.
17. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

18. Juergen Gall and Victor Lempitsky, "Class-specific hough forests for object detection," in *Decision Forests for Computer Vision and Medical Image Analysis*, pp. 143–157. Springer, 2013.
19. Xuming He and Stephen Gould, "An exemplar-based crf for multi-instance object segmentation," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 296–303.
20. Stan Z Li, *Markov random field modeling in Image Analysis (was: Markov random field modeling in computer vision).*, 2011.
21. *ImageNET webpage. http://image-net.org/about-publication.*