

# Object Proposals in Computer Vision

Neelima Chavali

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University  
in partial fulfillment of the requirements for the degree of

Master of Science  
in  
Electrical Engineering

Dhruv Batra, Chair  
Devi Parikh  
Lynn Abbott

22nd July, 2015  
Blacksburg, Virginia

Keywords: Object proposals, evaluation, computer vision  
Copyright 2015, Neelima Chavali

# The PASCAL Visual Object Classes Homepage



## The PASCAL VOC project:

- Provides standardised image data sets for object class recognition
- Provides a common set of tools for accessing the data sets and annotations
- Enables evaluation and comparison of different methods
- Ran challenges evaluating performance on object class recognition (from 2005-2012, now finished)

## Pascal VOC data sets

Data sets from the VOC challenges are available through the challenge links below, and evaluation of new methods on these data sets can be achieved through the [PASCAL VOC Evaluation Server](#). The evaluation server will remain active even though the challenges have now finished.

## News

- **Nov-2014:** A new feature for the [Leaderboards](#) of the PASCAL VOC evaluation server has been added, indicating if the differences between a selected submission and others are statistically significant or not.
- **May-2014:** A new paper covering the 2008-12 years of the challenge, and lessons learnt, is now available:

### **The PASCAL Visual Object Classes Challenge: A Retrospective**

Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J. and Zisserman, A.

*International Journal of Computer Vision*, 111(1), 98-136, 2015

[http://www.pascal2.org/papers/ijcv111.pdf](#)

- **May-2014:** A new paper covering the 2008-12 years of the challenge, and lessons learnt, is now available:

**The PASCAL Visual Object Classes Challenge: A Retrospective**

Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J. and Zisserman, A.

*International Journal of Computer Vision*, 111(1), 98-136, 2015

[Bibtex source](#) | [Abstract](#) | [PDF](#)

- **Feb-2014:** [Leaderboards](#) are now available for the VOC 2010, 2011 and 2012 datasets.

A submission to the Evaluation Server is by default private, but can optionally be "published" to the relevant leaderboard.

The Evaluation Server can now generate an anonymized URL, suitable for inclusion in a conference submission, giving the performance summary for a submitted entry.

- **Nov-2013:** A leaderboard including significance tests will be soon be introduced for new submissions. See [Assessing the Significance of Performance Differences on the PASCAL VOC Challenges via Bootstrapping](#) for a description and a demonstration of the method on VOC2012.

## Pascal VOC Challenges 2005-2012

The VOC series of challenges has now finished. We are very grateful to the hundreds of participants that have taken part in the challenges over the years. The [PASCAL VOC Evaluation Server](#) will continue to run.

### **Mark Everingham**

*It is with great sadness that we report that Mark Everingham died in 2012. Mark was the key member of the VOC project, and it would have been impossible without his selfless contributions. The VOC workshop at ECCV 2012 was dedicated to Mark's memory. [A tribute web page](#) has been set up, and an [appreciation of Mark's life and work](#) published.*

Details of each of the challenges can be found on the corresponding challenge page:

- [The VOC2012 Challenge](#)
- [The VOC2011 Challenge](#)
- [The VOC2010 Challenge](#)
- [The VOC2009 Challenge](#)
- [The VOC2008 Challenge](#)



- [The VOC2008 Challenge](#)
- [The VOC2007 Challenge](#)
- [The VOC2006 Challenge](#)
- [The VOC2005 Challenge](#)

Further details of the challenges may be found in the sections below:

- [Best Practice \(Recommendations on using the training and test data\)](#)
- [History and Background of the VOC Challenge](#)
- [The PASCAL Object Recognition Database Collection](#)
- [Publications relating to the VOC Challenge](#)
- [Organizers](#)
- [Support](#)

## Best Practice

The VOC challenge encourages two types of participation: (i) methods which are trained using only the provided "trainval" (training + validation) data; (ii) methods built or trained using any data except the provided test data, for example commercial systems. In both cases the *test* data must be used strictly for reporting of results alone - it must not be used in any way to train or tune systems, for example by running multiple parameter choices and reporting the best results obtained.

If using the training data we provide as part of the challenge development kit, all development, e.g. feature selection and parameter tuning, must use the "trainval" (training + validation) set alone. One way is to divide the set into training and validation sets (as suggested in the development kit). Other schemes e.g.  $n$ -fold cross-validation are equally valid. The tuned algorithms should then be run only *once* on the test data.

In VOC2007 we made all annotations available (i.e. for training, validation and test data) but since then we have not made the test annotations available. Instead, results on the test data are submitted to an evaluation server.

Since algorithms should only be run *once* on the test data we strongly discourage multiple submissions to the server (and indeed the number of submissions for the same algorithm is strictly controlled), as the evaluation server should not be used for parameter tuning.

We encourage you to publish test results always on the latest release of the challenge, using the output of the evaluation server. If you wish to compare methods or design choices e.g. subsets of features, then there are two options: (i) use the entire VOC2007 data, where all annotations are available; (ii) report cross-validation results using the latest "trainval" set alone.

## Policy on email address requirements when registering for the evaluation server

In line with the Best Practice procedures (above) we restrict the number of times that the test data can be processed by the evaluation server. To prevent any abuses of this restriction an institutional email address is required when registering for the evaluation server. This aims to prevent one user registering multiple times under different emails. Institutional emails include academic ones, such as name@university.ac.uk, and corporate ones, but not personal ones, such as name@gmail.com or name@123.com.

## History and Background

The main challenges have run each year since 2005. For more background on VOC, the following journal paper discusses some of the choices we made and our experience in running the challenge, and gives a more in depth discussion of the 2007 methods and results:

### The PASCAL Visual Object Classes (VOC) Challenge

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J. and Zisserman, A.  
*International Journal of Computer Vision*, 88(2), 303-338, 2010

[Bibtex source](#) | [Abstract](#) | [PDF](#)

The table below gives a brief summary of the main stages of the VOC development.

| Year                 | Statistics   | New developments  | Notes  |
|----------------------|--|---|--|
| <a href="#">2005</a> | Only 4 classes: bicycles, cars, motorbikes, people. Train/validation/test: 1578 images containing 2209 annotated objects.                            | Two competitions: classification and detection                          | Images were largely taken from existing public datasets, and were not as challenging as the flickr images subsequently used. This dataset is obsolete. |
| <a href="#">2006</a> | 10 classes: bicycle, bus, car, cat, cow, dog, horse, motorbike, person, sheep. Train/validation/test: 2618 images containing 4754 annotated objects. | Images from flickr and from Microsoft Research Cambridge (MSRC) dataset | The MSRC images were easier than flickr as the photos often concentrated on the object of interest. This dataset is obsolete.                          |
|                      | 20 classes:  |   |  |

|                             |  |   |  |
|-----------------------------|--|---|--|
| <p><a href="#">2007</a></p> | <p>20 classes:</p> <ul style="list-style-type: none"> <li>• <i>Person</i>: person</li> <li>• <i>Animal</i>: bird, cat, cow, dog, horse, sheep</li> <li>• <i>Vehicle</i>: aeroplane, bicycle, boat, bus, car, motorbike, train</li> <li>• <i>Indoor</i>: bottle, chair, dining table, potted plant, sofa, tv/monitor</li> </ul> <p>Train/validation/test: 9,963 images containing 24,640 annotated objects.</p> | <ul style="list-style-type: none"> <li>• Number of classes increased from 10 to 20</li> <li>• Segmentation taster introduced</li> <li>• Person layout taster introduced</li> <li>• Truncation flag added to annotations</li> <li>• Evaluation measure for the classification challenge changed to Average Precision. Previously it had been ROC-AUC.</li> </ul>   | <p>This year established the 20 classes, and these have been fixed since then. This was the final year that annotation was released for the testing data.</p>                                  |
| <p><a href="#">2008</a></p> | <p>20 classes. The data is split (as usual) around 50% train/val and 50% test. The train/val data has 4,340 images containing 10,363 annotated objects.</p>  | <ul style="list-style-type: none"> <li>• Occlusion flag added to annotations</li> <li>• Test data annotation no longer made public.</li> <li>• The segmentation and person layout data sets include images from the corresponding VOC2007 sets.</li> </ul>  |  |
| <p><a href="#">2009</a></p> | <p>20 classes. The train/val data has 7,054 images containing 17,218 ROI annotated objects and 3,211 segmentations.</p>  | <ul style="list-style-type: none"> <li>• From now on the data for all tasks consists of the previous years' images augmented with new images. In earlier years an entirely new data set was released each year for the classification/detection tasks.</li> <li>• Augmenting allows the number of images to grow each year, and means that test results can be compared on the previous years' images.</li> <li>• Segmentation becomes a standard challenge (promoted from a taster)</li> </ul> | <ul style="list-style-type: none"> <li>• No difficult flags were provided for the additional images (an omission).</li> <li>• Test data annotation not made public.</li> </ul>                 |
| <p><a href="#">2010</a></p> | <p>20 classes. The train/val data has 10,103 images containing 23,374 ROI annotated objects and 4,203 segmentations.</p>   | <ul style="list-style-type: none"> <li>• Action Classification taster introduced.</li> <li>• Associated challenge on large scale classification introduced based on ImageNet.</li> <li>• Amazon Mechanical Turk used for early stages of the annotation.</li> </ul>   | <ul style="list-style-type: none"> <li>• Method of computing AP changed. Now uses all data points rather than TREC style sampling.</li> <li>• Test data annotation not made public.</li> </ul> |

|                      |   |   |  |
|----------------------|---|---|--|
| <a href="#">2010</a> | 20 classes. The train/val data has 10,103 images containing 23,374 ROI annotated objects and 4,203 segmentations. | <ul style="list-style-type: none"> <li>• Associated challenge on large scale classification introduced based on ImageNet.</li> <li>• Amazon Mechanical Turk used for early stages of the annotation.</li> </ul>             | <ul style="list-style-type: none"> <li>• Method of computing AP changed. Now uses all data points rather than TREC style sampling.</li> <li>• Test data annotation not made public.</li> </ul> |
| <a href="#">2011</a> | 20 classes. The train/val data has 11,530 images containing 27,450 ROI annotated objects and 5,034 segmentations. | <ul style="list-style-type: none"> <li>• Action Classification taster extended to 10 classes + "other".</li> </ul>  | <ul style="list-style-type: none"> <li>• Layout annotation is now not "complete": only people are annotated and some people may be unannotated.</li> </ul>                                     |
| <a href="#">2012</a> | 20 classes. The train/val data has 11,530 images containing 27,450 ROI annotated objects and 6,929 segmentations. | <ul style="list-style-type: none"> <li>• Size of segmentation dataset substantially increased.</li> <li>• People in action classification dataset are additionally annotated with a reference point on the body.</li> </ul> | <ul style="list-style-type: none"> <li>• Datasets for classification, detection and person layout are the same as VOC2011.</li> </ul>  |

### Organizers

- Mark Everingham (University of Leeds)
- Luc van Gool (ETHZ, Zurich)
- Chris Williams (University of Edinburgh)
- John Winn (Microsoft Research Cambridge)
- Andrew Zisserman (University of Oxford)

with major contributions from

- Yusuf Aytar (University of Oxford)
- Ali Eslami (Microsoft Research Cambridge)
- Alexander Sorokin (University of Illinois at Urbana-Champaign)

### Support

The preparation and running of this challenge is supported by the EU-funded [PASCAL2](#) Network of Excellence on Pattern Analysis, Statistical Modelling and Computational Learning.