



Visual Object Classes Challenge 2010 (VOC2010)



[click on an image to see the annotation]

For news and updates, see the [PASCAL Visual Object Classes Homepage](http://host.robots.ox.ac.uk/pascal/VOC/voc2010/)

News

- 09-Mar-11: Test data can now be downloaded from the [PASCAL VOC Evaluation Server](http://host.robots.ox.ac.uk/pascal/VOC/voc2010/#data). You can also use the evaluation server to evaluate your method on the test data.
- 25-Oct-10: Detailed [results](#) of all submitted methods are now online. For summarized results and information about selected methods, please see the [workshop](#) presentations.
- 03-Sep-10: Provisional programme for the [workshop](#) is now online.
- 31-Aug-10: Submission of results is now closed. We are aiming to release results to participants only on 1st September.
- 09-Aug-10: The [evaluation server](#) is now online and results may be [submitted](#).

Dental Monitoring Ex. 2018
Align Technology, Inc., v. Dental Monitoring SAS.
IPR2024-00052

- 09-Aug-10: The deadline for submission of results is extended to [30th August, 23:00 GMT](#). There will be no further extensions.
- 24-May-10: The [test data](#) is now available.
- 08-May-10: The [development kit](#) including training/validation data is now available.
- 03-May-10: A new taster competition on (still image) [action classification](#) has been introduced.
- 08-Apr-10: A new taster competition on [Large Scale Visual Recognition](#) has been introduced in cooperation with [ImageNet](#).
- 01-Mar-10: We are preparing to run the VOC2010 challenge. The [provisional timetable](#) is below.
- 01-Mar-10: The challenge [workshop](#) will be held in conjunction with [ECCV 2010](#), 11th September 2010, Crete.

Contents

- [Introduction](#)
- [Data](#)
- [Development kit](#)
- [Test data](#)
- [Useful software](#)
- [Timetable](#)
- [Submission of results](#)
- [Best practice](#) (Recommendations on using the training and test data)
- [Publication policy](#)
- [Citation](#)
- [Database rights](#)
- [Organizers](#)
- [Acknowledgements](#)
- [Support](#)
- [History and background](#)

Introduction

The goal of this challenge is to recognize objects from a number of visual object classes in realistic scenes (i.e. not pre-segmented objects). It is fundamentally a supervised learning problem in that a training set of labelled images is provided. The twenty object classes that have been selected are:

- *Person*: person
- *Animal*: bird, cat, cow, dog, horse, sheep
- *Vehicle*: aeroplane, bicycle, boat, bus, car, motorbike, train
- *Indoor*: bottle, chair, dining table, potted plant, sofa, tv/monitor

There will be three main competitions: classification, detection, and segmentation; and three "taster" competition: person layout, action classification, and ImageNet large scale recognition:

Classification/Detection Competitions

1. **Classification:** For each of the twenty classes, predicting presence/absence of an example of that class in the test image.
2. **Detection:** Predicting the bounding box and label of each object from the twenty target classes in the test image.

20 classes



Participants may enter either (or both) of these competitions, and can choose to tackle any (or all) of the twenty object classes. The challenge allows for two approaches to each of the competitions:

1. Participants may use systems built or trained using any methods or data excluding the provided test sets.
2. Systems are to be built or trained using only the provided training/validation data.

The intention in the first case is to establish just what level of success can currently be achieved on these problems and by what method; in the second case the intention is to establish which method is most successful given a specified training set.

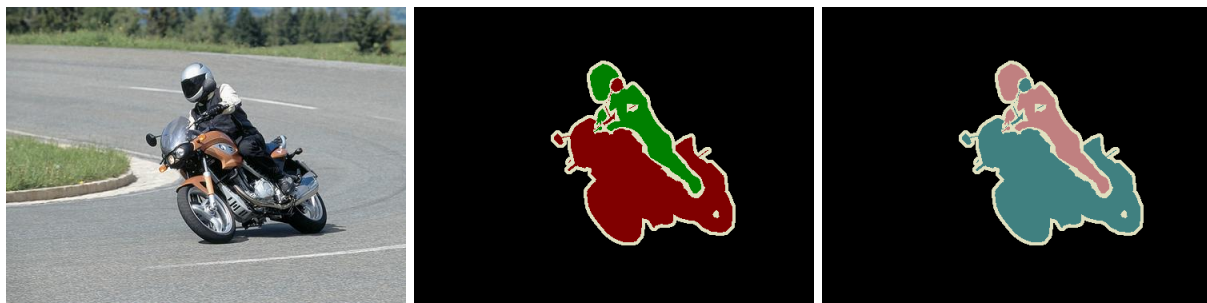
Segmentation Competition

- **Segmentation:** Generating pixel-wise segmentations giving the class of the object visible at each pixel, or "background" otherwise.

Image

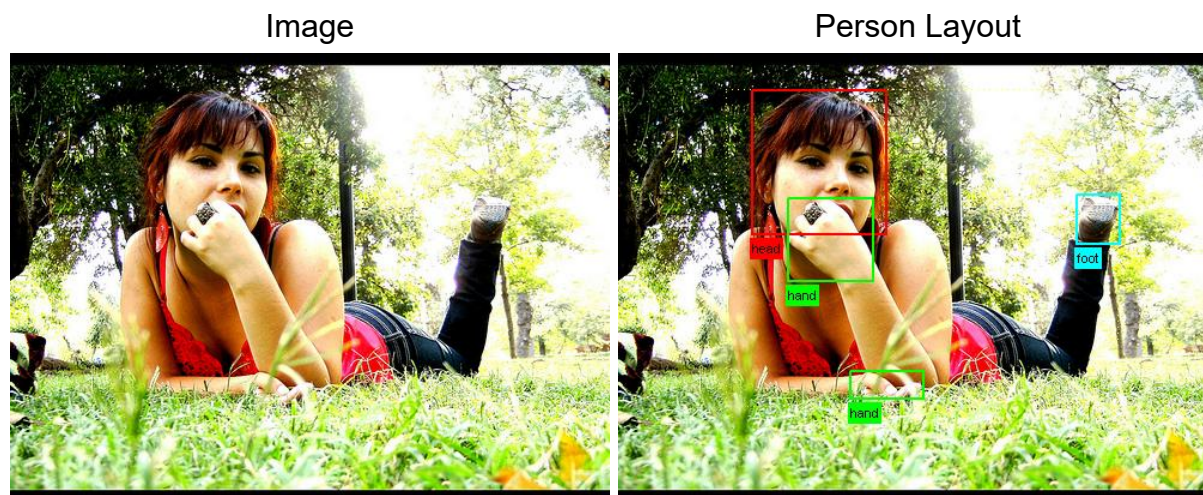
Objects

Class



Person Layout Taster Competition

- **Person Layout:** Predicting the bounding box and label of each part of a person (head, hands, feet).



Action Classification Taster Competition

- **Action Classification:** Predicting the action(s) being performed by a person in a still image.

9 action classes



ImageNet Large Scale Visual Recognition Taster Competition

The goal of this competition is to estimate the content of photographs for the purpose of retrieval and automatic annotation using a subset of the large hand-labeled [ImageNet](#) dataset (10,000,000 labeled images depicting 10,000+ object categories) as training. Test images will be presented with no initial annotation - no segmentation or labels - and algorithms will have to produce labelings specifying what objects are present in the images. In this initial version of the challenge, the goal is only to identify the main objects present in images, not to specify the location of objects.

Further details can be found at the [ImageNet](#) website.

Data

To download the training/validation data, see the [development kit](#).

The training data provided consists of a set of images; each image has an annotation file giving a bounding box and object class label for each object in one of the twenty classes present in the image. Note that multiple objects from multiple classes may be present in the same image. Some [example images](#) can be viewed online. A subset of images are also annotated with pixel-wise segmentation of each object present, to support the segmentation competition. Some [segmentation examples](#) can be viewed online.

Annotation was performed according to a set of [guidelines](#) distributed to all annotators.

The data will be made available in two stages; in the first stage, a development kit will be released consisting of training and validation data, plus evaluation software (written in MATLAB). One purpose of the validation set is to demonstrate how the evaluation software works ahead of the competition submission.

In the second stage, the test set will be made available for the actual competition. As in the VOC2008/VOC2009 challenges, no ground truth for the test data will be released.

The data has been split into 50% for training/validation and 50% for testing. The distributions of images and objects by class are approximately equal across the training/validation and test sets. In total there are 21,738 images. Further [statistics](#) are online.

Example images

Example images and the corresponding annotation for the classification/detection/segmentation tasks, and and person layout taster can be viewed online:

- [Classification/detection example images](#)
- [Segmentation example images](#)
- [Person Layout taster example images](#)
- [Action Classification taster example images](#)

Development Kit

The development kit consists of the training/validation data, MATLAB code for reading the annotation data, support files, and example implementations for each competition.

The development kit will be available according to the [timetable](#).

- Download the [training/validation data](#) (1.3GB tar file)
- Download the [development kit code and documentation](#) (290KB tar file)
- Download the [PDF documentation](#) (200KB PDF)
- Browse the [HTML documentation](#)
- View the [guidelines](#) used for annotating the database

Test Data

The test data is now available. Note that the only annotation in the data is for the layout/action taster competitions. As in 2008/2009, there are no current plans to release full annotation - evaluation of results will be provided by the organizers.

The test data can now be downloaded from the [evaluation server](#). You can also use the evaluation server to evaluate your method on the test data.

Useful Software

Below is a list of software you may find useful, contributed by participants to previous challenges.

- [Discriminatively Trained Deformable Part Models](#)
Pedro Felzenszwalb, Ross Girshick, David McAllester, Deva Ramanan.
- [Color Descriptors](#)
Koen van de Sande, Theo Gevers, Cees Snoek.

Timetable

- 3 May 2010: Development kit (training and validation data plus evaluation software) made available.
- 31 May 2010: Test set made available.
- **30 August 2010, 23:00 GMT. Deadline for submission of results. There will be no further extensions.**
- 11 September 2010: [Workshop](#) in association with [ECCV 2010](#), Crete.

Submission of Results

Participants are expected to submit a single set of results per method employed. Participants who have investigated several algorithms may submit one result per method. Changes in algorithm parameters do *not* constitute a different method - all parameter tuning must be conducted using the training and validation data alone.

Results must be submitted using the automated evaluation server:

- [PASCAL VOC Evaluation Server](#)

It is essential that your results files are in the correct format. Details of the required file formats for submitted results can be found in the development kit documentation. The results files should be collected in a single archive file (tar/tgz/tar.gz). An example of the correct format is available:

- [Example results file](#) (4MB gzipped tar file)

The format of your results **must** match that specified in the development kit and in the example file, including both **file names** and **directory structure**. If you are not entering a particular competition, just omit the corresponding files.

Dental Monitoring Ex. 2018

Participants submitting results for several different methods (noting the definition of different methods above) should produce a **separate** archive for each method.

In addition to the results files, participants will need to additionally specify:

- contact details and affiliation
- list of contributors
- brief description of the method

If you would like to submit a more detailed description of your method, for example a relevant publication, this can be included in the results archive.

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.

Publication Policy

The main mechanism for dissemination of the results will be the challenge webpage.

The detailed output of each submitted method will be published online e.g. per-image confidence for the classification task, and bounding boxes for the detection task. The intention is to assist others in the community in carrying out detailed analysis and comparison with their own methods. The published results will not be anonymous - by submitting results, participants are agreeing to have their results shared online.

Citation

If you make use of the VOC2010 data, please cite the following reference (to be prepared after the challenge workshop) in any publications:

```
@misc{pascal-voc-2010,  
  author = "Everingham, M. and Van~Gool, L. and Williams, C. K. I. and Winn, J. and Zisserman, A.",  
  title = "The {PASCAL} {V}isual {O}bject {C}lasses {C}hallenge 2010 {(VOC2010)} {R}esults",  
  howpublished = "http://www.pascal-network.org/challenges/VOC/voc2010/workshop/index.html"}  

```

Database Rights

The VOC2010 data includes images obtained from the "flickr" website. Use of these images must respect the corresponding terms of use:

- ["flickr" terms of use](#)

For the purposes of the challenge, the identity of the images in the database, e.g. source and name of owner, has been obscured. Details of the contributor of each image can be found in the annotation to be included in the final release of the data, after completion of the challenge. Any queries about the use or ownership of the data should be addressed to the [organizers](#).

Organizers

- Mark Everingham (University of Leeds), m.everingham@leeds.ac.uk
- Luc van Gool (ETHZ, Zurich)
- Chris Williams (University of Edinburgh)
- John Winn (Microsoft Research Cambridge)
- Andrew Zisserman (University of Oxford)

Acknowledgements

We gratefully acknowledge the following, who spent many long hours providing annotation for the VOC2010 database:

Dental Monitoring Ex. 2018

Page 9 of 12

Yusuf Aytar, Jan Hendrik Becker, Patrick Buehler, Miha Drenik, Chris Engels, Ali Eslami, Adrien Gaidon, Sam Johnson, Jyri Kivinen, Lubor Ladicky, Markus Mathias, Alastair Moore, Glenn Sheasby, Paul Sturges, David Tingdahl, Josiah Wang.

We also thank Alexander Sorokin for production and support of the annotation systems for Mechanical Turk, Ivan Laptev for development of the action classification task, and Marcin Eichner, Ali Eslami, Lubor Ladicky, Marcin Marszalek, Arpit Mittal and Andrea Vedaldi for testing of the development kit and evaluation server. We are also most grateful to Yusuf Aytar for further development and administration of the evaluation server.

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.

We are grateful to Alyosha Efros for providing additional funding for annotation on Mechanical Turk.

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
2005	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.

2006	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.
2007	20 classes: • <i>Person</i>: person • <i>Animal</i>: bird, cat, cow, dog, horse, sheep • <i>Vehicle</i>: aeroplane, bicycle, boat, bus, car, motorbike, train • <i>Indoor</i>: bottle, chair, dining table, potted plant, sofa, tv/monitor Train/validation/test: 9,963 images containing 24,640 annotated objects.	• Number of classes increased from 10 to 20 • Segmentation taster introduced • Person layout taster introduced • Truncation flag added to annotations • Evaluation measure for the classification challenge changed to Average Precision. Previously it had been ROC-AUC.	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.
2008	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.	• Occlusion flag added to annotations. • Test data annotation no longer made public. • The segmentation and person layout data sets include images from the corresponding VOC2007 sets.	

2009	20 classes. The train/val data has 7,054 images containing 17,218 ROI annotated objects and 3,211 segmentations.	• 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. • Augmenting allows the number of images to grow each year, and means that test results can be compared on the previous years' images. • Segmentation becomes a standard challenge (promoted from a taster)	• No difficult flags were provided for the additional images (an omission). • Test data annotation not made public.
2010	20 classes. The train/val data has 10,103 images containing 23,374 ROI annotated objects and 4,203 segmentations.	• Action Classification taster introduced. • Associated challenge on large scale classification introduced based on ImageNet. • Amazon Mechanical Turk used for early stages of the annotation.	• Method of computing AP changed. Now uses all data points rather than TREC style sampling. • Test data annotation not made public.