

VOC2011 Annotation Guidelines

Guidelines on what and how to label.

What to label	<p>All objects of the defined categories, unless:</p> <ul style="list-style-type: none"> • you are unsure what the object is. • the object is very small (at your discretion). • less than 10-20% of the object is visible, <i>such that you cannot be sure what class it is</i>. e.g. if only a tyre is visible it may belong to car or truck so cannot be labelled car, but feet/faces can only belong to a person. <p>If this is not possible because too many objects, mark image as bad.</p>
Viewpoint	<p>Record the viewpoint of the 'bulk' of the object e.g. the body rather than the head.</p> <p>Allow viewpoints within 10-20 degrees.</p> <p>If ambiguous, leave as 'Unspecified'. Unusually rotated objects e.g. upside-down people should be left as 'Unspecified'.</p>
Bounding box	<p>Mark the bounding box of the visible area of the object (<i>not</i> the estimated total extent of the object).</p> <p>Bounding box should contain all visible pixels, except where the bounding box would have to be made excessively large to include a few additional pixels (<5%) e.g. a car aerial.</p>
Truncation	<p>If more than 15-20% of the object lies outside the bounding box mark as Truncated.</p> <p>The flag indicates that the bounding box does not cover the total extent of the object.</p>
Occlusion	<p>If more than 5% of the object is occluded within the bounding box, mark as Occluded.</p> <p>The flag indicates that the object is not totally visible within the bounding box.</p>
Image quality/illumination	<p>Images which are poor quality (e.g. excessive motion blur) should be marked bad.</p> <p>However, poor illumination (e.g. objects in silhouette) should not count as poor quality unless objects cannot be recognised.</p> <p>Images made up of multiple images (e.g. collages) should be marked bad.</p>
Clothing/mud/snow etc.	<p>If an object is 'occluded' by a close-fitting occluder e.g. clothing, mud, snow etc., then the occluder should be treated as part of the object.</p>
Transparency	<p>Do label objects visible through glass, but treat reflections on the glass as occlusion.</p>
Mirrors	<p>Do label objects in mirrors.</p>
Pictures	<p>Label objects in pictures/posters/signs only if they are photorealistic but not if cartoons, symbols etc.</p>

Guidelines on categorisation

Aeroplane	Includes gliders but not hang gliders or helicopters
Bicycle	Includes tricycles, unicycles
Bird	All birds
Boat	Ships, rowing boats, pedaloes but not jet skis
Bottle	Plastic, glass or feeding bottles
Bus	Includes minibus but not trams
Car	<p>Includes cars, vans, large family cars for 6-8 people etc.</p> <p>Excludes go-carts, tractors, emergency vehicles, lorries/trucks etc.</p> <p>Do not label where only the vehicle interior is shown.</p> <p>Include toys that look just like real cars, but not 'cartoony' toys.</p>
Cat	Domestic cats (not lions etc.)
Chair	<p>Includes armchairs, deckchairs but not stools or benches.</p> <p>Excludes seats in buses, cars etc.</p> <p>Excludes wheelchairs.</p>
Cow	All cows
Dining table	<p>Only tables for eating at.</p> <p>Not coffee tables, desks, side tables or picnic benches</p>
Dog	Domestic dogs (not wolves etc.)
Horse	Includes ponies, donkeys, mules etc.
Motorbike	Includes mopeds, scooters, sidecars
People	Includes babies, faces (i.e. truncated people)
Potted plant	Indoor plants excluding flowers in vases, or outdoor plants clearly in a pot.
Sheep	Sheep, not goats
Sofa	Excludes sofas made up as sofa-beds

Train	Includes train carriages, excludes trams
TV/monitor	Standalone screens (not laptops), not advertising displays

Guidelines on segmentation

What to segment	Objects whose bounding boxes have been labelled according to the above guidelines. You may need to exclude backpacks, handbags etc. which were included in the bounding box. You may also need to include hands, chair legs etc. which were <i>outside</i> the bounding box.
Accuracy	Segment within 5 pixels. Labelled pixels MUST be the object; pixels outside the 5-pixel border area MUST be background. Border pixels can be either. Use the tri-map displayed by the segmentation tool to ensure these constraints hold. This may involve labelling pixels outside the bounding box.
Mixed pixels/transparency	Pixels which are mixed e.g. due to transparency, motion blur or the presence of a border should be considered to belong to the object whose colour contributes most to the mix.
Thin structures	Aim to capture thin structures where possible, within the accuracy constraints. Structures of around one pixel thickness can be ignored e.g. wires, rigging, whiskers.
Objects on tables etc.	If a number of small objects are occluding an object e.g. cutlery/silverware on a dining table, they can be considered part of that object. The exception is if they are sticking out of the object (e.g. candles) where they should be truncated at the object boundary.
Difficult images	Images which are overly difficult to segment to the required accuracy can be left unlabelled e.g. a nest of bicycles.