**Annals of Science and Inquiry**
Vol. 2 No. 11 (2026)

# When the Visible Part Shifts the Centre of a Stationary Component

Judith Nelson · Heather Snyder · Simon L Mitchell

Published: November 26, 2026 | Article 1156

## Abstract

For a partly hidden component in an image, should the annotated location be the centre of its visible region or of its full projection? We compared the midpoint of its visible-part box with the midpoint of its full projected box in 1,000 recorded T-LESS views. With less than 30% of a component visible, the separation reached a tenth of the full-box diagonal in an average of 62.0% of observations per scene. No observation reached that distance at 90% visibility or higher. Among 68 components observed at both low and high visibility, the lower-visibility views added an average of 8.45 percentage points of normalized separation. The observed deviation ties the choice of annotation convention to how much of the component is visible.

# When the Visible Part Shifts the Centre of a Stationary Component

## Abstract

For a partly hidden component in an image, should the annotated location be the centre of its visible region or of its full projection? We compared the midpoint of its visible-part box with the midpoint of its full projected box in 1,000 recorded T-LESS views. With less than 30% of a component visible, the separation reached a tenth of the full-box diagonal in an average of 62.0% of observations per scene. No observation reached that distance at 90% visibility or higher. Among 68 components observed at both low and high visibility, the lower-visibility views added an average of 8.45 percentage points of normalized separation. The observed deviation ties the choice of annotation convention to how much of the component is visible.

## A centre attached to the visible surface

In a multi-perspective inspection setup, one camera may see a component almost in full while another sees it through gaps between neighbouring parts. The target is still the same component. Its rectangular label might cover the portion that the camera can see, or it might represent the full object projected into the image. Both choices can be useful for visual recognition, but their centres can fall on different pixels in the same photograph. Jin et al. (2024) study visual recognition of control panels in multi-perspective mixed-reality monitoring. That setting motivates a question about location evidence before any interaction with a panel is considered. **How much does the apparent position of a stationary component change when the visible-part box midpoint is used in place of the full-silhouette midpoint under different levels of recorded visibility?**

A rectangle changes when part of a horizontal or vertical extent disappears. If only the object's left edge is obscured, its observed left limit can move inward while its right limit stays fixed. The midpoint then moves right. Obscuring a central patch can remove many pixels while preserving all four extrema. In this case the visible fraction decreases and the midpoint stays where it was. We expect the gap between centre conventions to grow on average as fewer pixels remain visible, while allowing individual views to behave differently. If the association is simply due to some component types being concentrated in cluttered scenes, it may diminish when the same physical components are observed in different views. This motivates a second comparison within individual scene-component pairs.

For recognition in clutter, the idea of combining evidence from visible parts goes back to work on component-based recognition (Biederman, 1987). In computer vision, Felzenszwalb et al. (2010) give an example of detection using separate part responses. Hoiem et al. (2012) use error analysis to distinguish localization from occlusion and other properties of detection. Those distinctions matter here because an apparent localization error can first appear at the level of the annotation convention. We hold the object, view and source annotation fixed in each comparison. A detector's decision or accuracy is outside the measured outcome.

## Detecting a region and recovering an extent

Standard detection tasks formalize image-region matching through labels and overlap, as exemplified by the PASCAL VOC Challenge (Everingham et al., 2010). There are several useful ways to encode locations. The boxes in CornerNet are defined through paired corners (Law & Deng, 2018), Objects as Points starts with an object centre (Zhou et al., 2019), and Mask R-CNN adds instance masks to detected boxes (He et al., 2017). These approaches make clear that centre, mask and corners are linked, while leaving the target convention dependent on the annotations. Generalized intersection over union quantifies the spatial agreement of boxes and extends ordinary overlap to disjoint pairs (Rezatofighi et al., 2019). Zheng et al. (2020) add an explicit distance between box centres to an overlap measure. The present analysis isolates midpoint displacement and reports it in a unit with a direct geometric interpretation, so changes in box extent do not become indistinguishable from translation.

In amodal segmentation, the target includes the extent that would be seen if occluding objects were absent. Zhu et al. (2017) use this distinction to annotate visible and occluded regions, and Qi et al. (2019) develop an instance-level dataset in which models must reason about invisible parts. Follmann et al. (2019) and Ehsani et al. (2018) each test ways to recover full regions from incomplete views. Li et al. (2023) study whether multiple views can help recover these regions and make both forms of annotation available. Our outcome is the discrepancy implied by switching conventions on source annotations. It estimates how consequential the choice might be for a centre-location task even before the uncertainty of an amodal prediction is introduced.

The camera archives used to benchmark rigid-pose estimation provide relevant physical observations. Hinterstoisser et al. (2013) address textureless objects in clutter, while Xiang et al. (2018) provide a pose-recognition system and an annotated camera benchmark for rigid objects. Drost et al. (2017) construct another dataset expressly around cameras and industrial parts. Multi-view work by Labbé et al. (2020) illustrates the use of object correspondences to recover a consistent scene from distinct observations. T-LESS provides recorded views of the same arrangements from multiple camera poses and convention-specific boxes in its BOP distribution. The dataset therefore lets us ask how an image-space centre changes while the physical arrangement stays fixed.

## Camera records and the two locations assigned to each object

We used the Primesense camera partition of T-LESS from the official BOP19 package. Hodaň et al. (2017) describe the original archive’s capture of rigid, low-texture components. Hodaň et al. (2018) introduce the BOP evaluation and common format. The chosen package has 50 camera views for each of 20 stationary scene configurations. From the published archive we retrieved each scene’s camera parameters, object poses and visibility metadata. We joined the three files at the scene and frame level and matched pose and visibility rows by the common object annotation position within a frame. The analysis uses the prepared 720 by 540 image coordinate system. It excludes the original collection’s other camera sensors and every training partition.

The BOP visibility record includes the full projected bounding box, the box around its visible part and the fraction of the full silhouette counted as visible. The full region comes from the object model projected at the labelled pose. The visible subdivision uses the relationship between the model projection and the recorded scene depth, according to the benchmark’s rules. The two boxes are thus linked to the same geometry and annotation process. The visibility ratio and midpoint separation are also constructed from the same silhouette information, so the analysis estimates the strength of a geometric association in the archive. It would take independently annotated observations to quantify how well the underlying geometry represents a component’s real boundaries.

We computed each midpoint by adding half the annotated width and height to its top-left corner. The distance between the visible and full midpoints is the length of their two-dimensional difference vector. The full-box diagonal provides the scale for normalization. A ten-pixel shift can be substantial for a small component but comparatively minor for one that occupies much of the view. The midpoint of a rectangle can also fall into empty space between parts of the silhouette. The analysis measures the centre of the box and makes no inference about an object's physical centre of mass.

From 6,900 object-view annotations, we excluded 21 with no visible pixels or an empty silhouette and six with a nonpositive side length on a required box. We checked that the full box stayed inside the image border and that the visible box lay within one pixel of its bounds under the same coordinate convention. Neither of those checks removed any additional records. The finished sample comprises 6,873 paired annotations. A fully hidden object has no visible rectangle to use and falls outside the estimand. Estimating the availability of a usable visible anchor for all known objects would require counting those missing anchors directly.

## Comparing different views of the same assembly

The first analysis groups annotations by visible fraction using 30%, 60% and 90% cutoffs. We measure two outcomes within each band. The first is the mean distance between midpoints as a percentage of the full-box diagonal. The second is how often that distance equals or exceeds 10% of the diagonal. This threshold gives the comparison a numerical scale, with checks at 5% and 15%. A pooled annotation-level average could disproportionately reflect dense scenes. Instead, we first calculate the mean and threshold rate for each scene within a visibility band. A second average gives each of the 20 scenes equal weight. These are comparisons of the distance and the frequency of a meaningful discrepancy within the archive, in the spirit of measuring agreement directly instead of relying on association alone (Haber & Barnhart, 2006; Lin, 1989).

Repeated photographs from a single scene share objects and surrounding geometry. Roberts et al. (2017) explain why dependence needs to inform a design based on repeated or structured records, and Cameron and Miller (2015) review the implications of forming groups for statistical inference. We use 5,000 percentile bootstrap replicates by resampling whole scenes, following the general resampling approach of Efron (1979) and the clustered framework discussed by Field and Welsh (2007). The intervals show sensitivity to which sampled scenes contribute to the summary. The scenes are benchmark configurations and should not be regarded as a random sample of all industrial installations. An interval restricted to zero can arise when every observed scene reports a zero frequency. It expresses the limits of the observed resampling distribution.

The second comparison follows the same component across photographs, restricting the analysis to object identifiers with at most one instance per frame within a given scene. Because one model identifier could refer to two visually identical instances, we also transform every candidate’s published camera-frame translation into its scene world frame and check that the positions remain within 2 mm of the coordinate-wise median world position for that candidate. The inverse transform follows the standard geometric relation between the world and camera systems (Hartley & Zisserman, 2004). The check verifies internal annotation identity and does not provide independent location measurements. A component must have at least two eligible views below 60% visibility and two at or above 90%. We compare its mean midpoint separation between those bands, average component-specific differences in each scene and report the equally weighted scene mean.

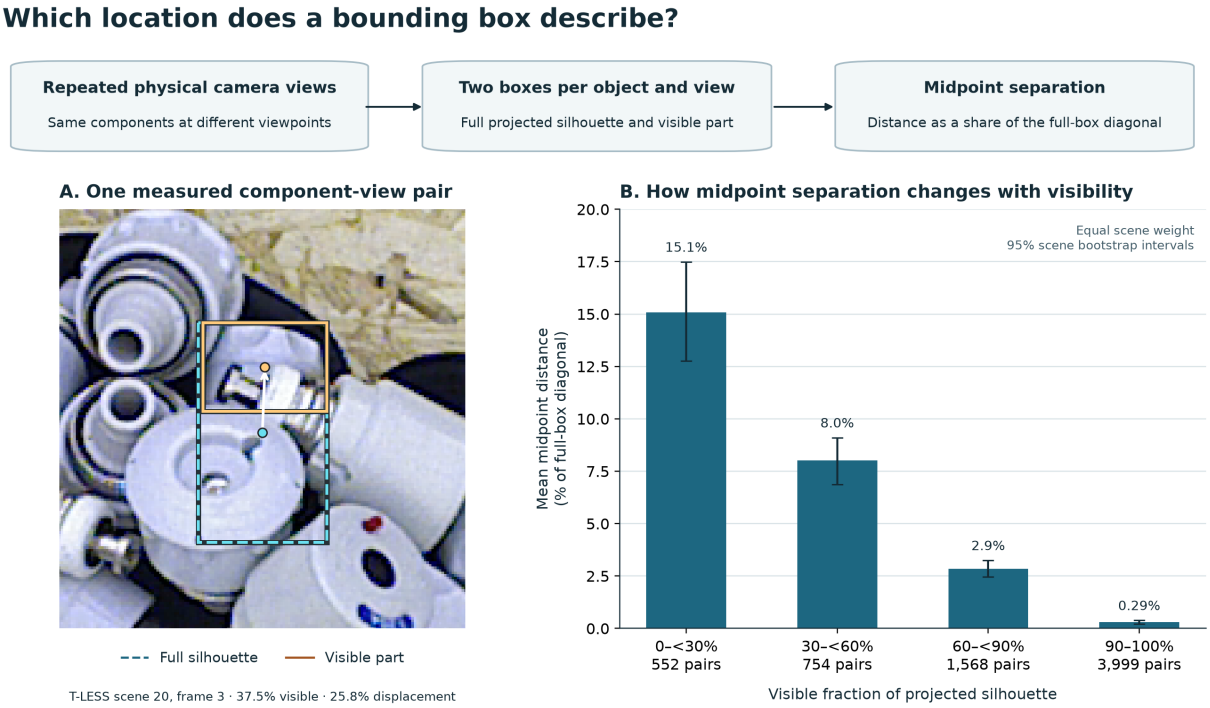


*Figure 1. Both bounds refer to the same annotated component in the same view. Panel A shows a crop of T-LESS Primesense scene 20, frame 3, with bounding boxes, midpoints and displacement arrow added. Source photograph by Hodaň et al. (2017), T-LESS, CC BY 4.0. Panel B gives the mean midpoint separation under equal scene weights in each of four visibility bands, with scene-resampling intervals.*

## Most of the separation occurs when a large part is hidden

Table 1 shows a gradual decline in the separation across the four visibility bands. When less than 30% of the projected silhouette was visible, the scene-weighted average separation was 15.08% of the full-box diagonal. In the near-full band it was 0.29%. The median unscaled distance also decreased, from 18.33 pixels in the lowest band to zero pixels in the highest. However, some components lost substantial silhouette area without much midpoint movement. The lowest visibility band had a pooled median displacement of 16.99% of the diagonal and a ninetieth percentile of 30.94%, consistent with the geometry allowing different moves from the same general degree of visibility.

***Table 1. Separation of the visible and full box midpoints by source-labelled visibility***

| Visible share | Paired annotations | Mean distance, % of diagonal | 95% interval | Distance ≥10% of diagonal | 95% interval |
|---|---|---|---|---|---|
| Less than 30% | 552 | 15.08% | 12.77–17.49% | 61.96% | 52.57–70.97% |
| 30% to less than 60% | 754 | 8.02% | 6.87–9.11% | 36.88% | 31.06–42.16% |
| 60% to less than 90% | 1,568 | 2.85% | 2.45–3.25% | 6.10% | 3.78–8.49% |
| 90% to 100% | 3,999 | 0.29% | 0.21–0.39% | 0.00% | 0.00–0.00% |

*Estimates weight 20 scenes equally within each band. Intervals are percentiles from 5,000 whole-scene bootstrap draws. The all-zero interval for the highest band reflects zero observed events in each scene.*

The contrast persists if the separation threshold changes. At 5% of the diagonal, the equal-scene frequency was 74.39% in the lowest visibility band and 0.89% in the highest. At 15%, those frequencies were

48.40% and zero. A sensitivity check retaining 6,305 records for which the source reported a valid depth measurement for at least 90% of the full-silhouette pixel count gave a 62.00% exceedance rate under the 10% rule in the lowest band, close to the 61.96% main estimate. That comparison reduces the influence of missing measured depth in the region, but it does not remove all uncertainty about where the visibility classifier places contour pixels.

The repeated component analysis identifies 68 distinct stationary scene-component combinations in 18 scenes. Each of the 68 had a higher mean separation in its lower-visibility views. Averaging the paired differences at the scene level gave an increase of 8.45 percentage points of the diagonal, with a scene-resampling interval of 6.62 to 10.15 percentage points. This comparison holds fixed the scene and the component, and its direction agrees with the band summaries. A component may still be seen from a different angle in the two visibility groups. The comparison cannot distinguish the contribution of a new projection geometry from the contribution of changes in which parts of that projection are covered.

## What the location convention means for inspection benchmarks

Part of the practical issue is that retaining visible evidence for recognition and locating a whole component are separate objectives. A rectangle tightly surrounding a visible edge may help a system highlight the observed pixels. Its midpoint may become a poor anchor for the component as a whole when the rest of the object is behind a neighbour. A source annotation can define both positions even if only one is directly inferable from the visible part in a single image. A careful evaluation should identify which position the system is supposed to report and where its reference comes from.

Torralba and Efros (2011) show how apparently similar recognition tasks can differ when their datasets change, and Barbu et al. (2019) use controlled images to investigate recognition under altered scene conditions. The scope here is narrower. These are camera observations of staged, rigid components and published depth-assisted model annotations. Installed panels can have different fasteners, surfaces, occlusion patterns and image-scale constraints. A test of their visual anchors would need both a defined extent convention and a separately established reference. The measured separation in T-LESS quantifies a source of disagreement between location labels for repeated physical views.

## Conclusion

Across the recorded T-LESS views, the visible-part box provided nearly the same centre as the full-silhouette box when most of the silhouette remained visible. Their separation grew at lower visibility and also grew within individual components that could be identified across views. The choice of image-space centre therefore needs to be stated when repeatability across cameras is assessed. The two coordinates can be treated as interchangeable only to the extent that their observed disagreement fits the intended spatial tolerance.

## Data and reproducibility

The T-LESS project and the BOP archive publish the source photographs and annotations under Creative Commons Attribution 4.0. The analysis files record the joined annotation rows and scene-level summaries. A retrieval manifest pins the official archive revision and records each extracted member’s ZIP CRC and local SHA-256. The scripts regenerate the table values and figure from those members.

## References


Barbu, A.; Mayo, D.; Alverio, J.; Luo, W.; Wang, C.; Gutfreund, D.; Tenenbaum, J.; Katz, B. (2019). ObjectNet: A large-scale bias-controlled dataset for pushing the limits of object recognition models. *Advances in Neural Information Processing Systems 32*. https://papers.nips.cc/paper/2019/hash/97af07a14cacba681feacf3012730892-Abstract.html

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94(2), 115–147. https://doi.org/10.1037/0033-295x.94.2.115

Cameron, A. C.; Miller, D. L. (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, 50(2), 317–372. https://doi.org/10.3368/jhr.50.2.317

Jin, T.; Kan, D.; Yin, S.; Zhao, R. (2024). A Study on Multi-Perspective Master-Slave Monitoring and Visual Recognition of Control Panels Based on Mixed Reality. *2024 10th International Conference on Systems and Informatics (ICSAI)*, 1–5. https://doi.org/10.1109/icsai65059.2024.10893733

Drost, B.; Ulrich, M.; Bergmann, P.; Hartinger, P.; Steger, C. (2017). Introducing MVTec ITODD - A Dataset for 3D Object Recognition in Industry. *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2200–2208. https://openaccess.thecvf.com/content_ICCV_2017_workshops/w31/html/Drost_Introducing_MVTec_ITODD_ICCV_2017_paper.html

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1), 1–26. https://doi.org/10.1214/aos/1176344552

Ehsani, K.; Mottaghi, R.; Farhadi, A. (2018). SeGAN: Segmenting and Generating the Invisible. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6144–6153. https://openaccess.thecvf.com/content_cvpr_2018/html/Ehsani_SeGAN_Segmenting_and_CVPR_2018_paper.html

Everingham, M.; Van Gool, L.; Williams, C. K. I.; Winn, J.; Zisserman, A. (2010). The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, 88(2), 303–338. https://doi.org/10.1007/s11263-009-0275-4

Felzenszwalb, P. F.; Girshick, R. B.; McAllester, D.; Ramanan, D. (2010). Object Detection with Discriminatively Trained Part-Based Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9), 1627–1645. https://doi.org/10.1109/tpami.2009.167

Field, C. A.; Welsh, A. H. (2007). Bootstrapping Clustered Data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 69(3), 369–390. https://doi.org/10.1111/j.1467-9868.2007.00593.x

Follmann, P.; König, R.; Härtinger, P.; Klostermann, M.; Böttger, T. (2019). Learning to See the Invisible: End-to-End Trainable Amodal Instance Segmentation. *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1328–1336. https://doi.org/10.1109/WACV.2019.00146

Haber, M.; Barnhart, H. X. (2006). Coefficients of Agreement for Fixed Observers. *Statistical Methods in Medical Research*, 15(3), 255–271. https://journals.sagepub.com/doi/10.1191/0962280206sm441oa

Hartley, R.; Zisserman, A. (2004). Multiple View Geometry in Computer Vision. *Cambridge University Press (2nd edition)*. https://doi.org/10.1017/cbo9780511811685

He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. (2017). Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, 2961–2969. https://openaccess.thecvf.com/content_iccv_2017/html/He_Mask_R-CNN_ICCV_2017_paper.html

Hinterstoisser, S.; Lepetit, V.; Ilic, S.; Holzer, S.; Bradski, G.; Konolige, K.; Navab, N. (2013). Model Based Training, Detection and Pose Estimation of Texture-Less 3D Objects in Heavily Cluttered Scenes. *Computer Vision – ACCV 2012*, 7724, 548–562. https://doi.org/10.1007/978-3-642-37331-2_42

Hodaň, T.; Haluza, P.; Obdržálek, Š.; Matas, J.; Lourakis, M.; Zabulis, X. (2017). T-LESS: An RGB-D Dataset for 6D Pose Estimation of Texture-less Objects. *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 880–888. https://doi.org/10.1109/WACV.2017.103

Hodaň, T.; Michel, F.; Brachmann, E.; Kehl, W.; Glent Buch, A.; Kraft, D.; Drost, B.; Vidal, J.; Ihrke, S.; Zabulis, X.; et al. (2018). BOP: Benchmark for 6D Object Pose Estimation. *Proceedings of the European Conference on Computer Vision (ECCV)*, 19–34. https://openaccess.thecvf.com/content_ECCV_2018/html/Tomas_Hodan_PESTO_6D_Object_ECCV_2018_paper.html

Hoiem, D.; Chodpathumwan, Y.; Dai, Q. (2012). Diagnosing Error in Object Detectors. *Lecture Notes in Computer Science*, 340–353. https://doi.org/10.1007/978-3-642-33712-3_25

Labbé, Y.; Carpentier, J.; Aubry, M.; Sivic, J. (2020). CosyPose: Consistent multi-view multi-object 6D pose estimation. *Proceedings of the European Conference on Computer Vision (ECCV)*, 574–591. https://www.di.ens.fr/willow/research/cosypose/

Law, H.; Deng, J. (2018). CornerNet: Detecting Objects as Paired Keypoints. *Proceedings of the European Conference on Computer Vision (ECCV)*, 734–750. https://openaccess.thecvf.com/content_ECCV_2018/html/Hei_Law_CornerNet_Detecting_Objects_ECCV_2018_paper.html

Li, Z.; Ye, W.; Terven, J.; Bennett, Z.; Zheng, Y.; Jiang, T.; Huang, T. (2023). MUVA: A New Large-Scale Benchmark for Multi-View Amodal Instance Segmentation in the Shopping Scenario. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 23504–23513. https://openaccess.thecvf.com/content/ICCV2023/html/Li_MUVA_A_New_Large-Scale_Benchmark_for_Multi-View_Amodal_Instance_Segmentation_ICCV_2023_paper.html

Lin, L. I. K. (1989). A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics*, 45(1), 255–268. https://doi.org/10.2307/2532051

Qi, L.; Jiang, L.; Liu, S.; Shen, X.; Jia, J. (2019). Amodal Instance Segmentation With KINS Dataset. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3014–3023. https://openaccess.thecvf.com/content_CVPR_2019/html/Qi_Amodal_Instance_Segmentation_With_KINS_Dataset_CVPR_2019_paper.html

Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. (2019). Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 658–666. https://openaccess.thecvf.com/content_CVPR_2019/html/Rezatofighi_Generalized_Intersection_Over_Union_A_Metric_and_a_Loss_for_CVPR_2019_paper.html

Roberts, D. R.; Bahn, V.; Ciuti, S.; Boyce, M. S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J. J.; Schröder, B.; Thuiller, W.; et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8), 913–929. https://doi.org/10.1111/ecog.02881

Torralba, A.; Efros, A. A. (2011). Unbiased Look at Dataset Bias. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1521–1528. https://people.csail.mit.edu/torralba/research/bias/

Xiang, Y.; Schmidt, T.; Narayanan, V.; Fox, D. (2018). PoseCNN: A Convolutional Neural Network for 6D Object Pose Estimation in Cluttered Scenes. *Robotics: Science and Systems XIV*. https://www.roboticsproceedings.org/rss14/p19.html

Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. (2020). Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(07), 12993–13000. https://doi.org/10.1609/aaai.v34i07.6999

Zhou, X.; Wang, D.; Krähenbühl, P. (2019). Objects as Points. *arXiv:1904.07850*. https://arxiv.org/abs/1904.07850

Zhu, Y.; Tian, Y.; Metaxas, D.; Dollar, P. (2017). Semantic Amodal Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1464–1472. https://openaccess.thecvf.com/content_cvpr_2017/html/Zhu_Semantic_Amodal_Segmentation_CVPR_2017_paper.html