Pose Estimation Errors, the Ultimate Diagnosis

Report

This report is associated with the work: Pose Estimation Errors, the Ultimate Diagnosis. C. Redondo-Cabrera, R. J. Lopez-Sastre, Y. Xiang , T. Tuytelaars, S. Savarese. ECCV, 2016.

Contents

1 Information **3**

2 Overall: Main Characteristics **4**

3 Results per Object Class **12**

3.1 aeroplane 12
3.2 bicycle 22
3.3 boat 32
3.4 bus 42
3.5 car 52
3.6 chair 62
3.7 diningtable 72
3.8 motorbike 82
3.9 sofa 92
3.10 train 102
3.11 tvmonitor 112

Pose Estimation Errors, the Ultimate Diagnosis: Report

September 12, 2016

1 Information

The **vdpm-gt** detector is analyzed. This is an automatically generated report.

Our diagnostic tool analyzes the frequency and impact of different types of false positives, and the infuence on the performance of the main object characteristics. Analyzing the different types of false pose estimations of the methods, we can gather very interesting information to improve them. Since it is difficult to characterize the error modes for generic rotations, we restrict our analysis to only the predicted azimuth. We discretize the azimuth angle into K bins, such that the bin centers have an equidistant spacing of $\frac{2\pi}{K}$. For our evaluations we set $K = 24$. Thus, we define the following types of error modes. *Opposite viewpoint error*, which measures the efect of flipped estimates (*e.g.* confusion between frontal and rear views of a car). *Nearby viewpoint errors*. Nearby pose bins are confused due to they are very correlated in terms of appearance. Finally, the *Other rotation errors*, which include the rest of false positives.

We also provide the success rate for each pose estimator considering 8 viewpoints: F: frontal. F-L: frontal-left. L: Left. L-RE: left-rear. RE: rear. RE-R: rear-right. R: right. R-F: right-frontal.

With respect to the impact of the main object characteristic, the following characteristic are considered in our study: occlusion/truncation, which indicates whether the object is occluded/truncated or not; object size and aspect ratio, which organizes the objects in different sets, depending on their size or aspect ratio; visible sides, which indicates if the object is in frontal, rear or side view position; and part visibility, which marks whether a 3D part is visible or not. For the object size, we measure the pixel area of the bounding box. We assign each object to a size category, depending on the object percentile size within its object category: extra-small (XS: bottom 10%); small (S: next 20%); large (L: next 80%); extra-large (XL: next 100%). Likewise, for the aspect ratio, objects are categorized into extra-tall (XT), tall (T), wide (W), and extra-wide (XW), using the same percentiles.

2 Overall: Main Characteristics

Note that the results shown in this section are obtained by computing the average over **the selected object classes** (*i.e*: aeroplane, bicycle, boat, bus, car, chair, diningtable, motorbike, sofa, train, tvmonitor) included in the PASCAL3D+ dataset.

Tables 1 and 2 summarize the detection and pose estimation results. Table 1 shows the model performance working on continuous pose estimation. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$. Table 2 summarizes the results achieved working on discrete pose estimation. The AVP and PEAP metrics are obtained considering different viewpoint discretizations: 4, 8, 16 and 24 views.

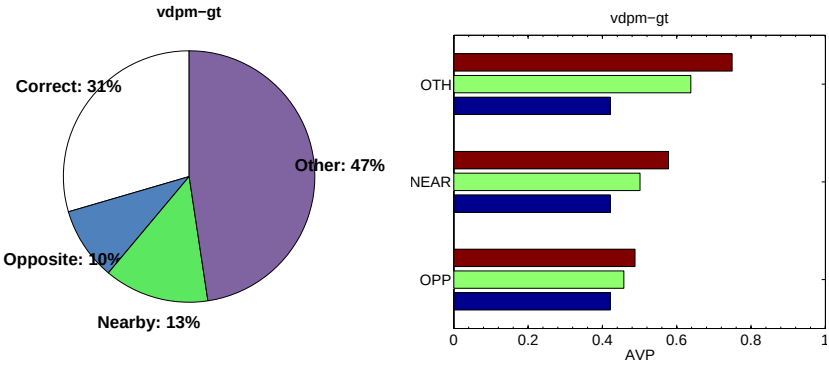
Table 1: **vdpm-gt: Detection and Continuous Pose Estimation Results on PASCAL3D+ dataset. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$.**

Metric	aeroplane	bicycle	boat	bus	car	chair	diningtable	motorbike	sofa	train	tvmonitor	AVG
AP	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
AOS	81.3	87.6	55.9	77.0	70.8	70.5	74.2	85.8	74.2	81.5	90.9	77.2
AVP ($\frac{\pi}{12}$)	36.1	44.5	10.5	71.6	40.7	25.9	31.6	43.1	40.8	68.1	50.4	42.1
PEAP ($\frac{\pi}{12}$)	10.1	15.5	1.0	36.9	12.3	5.3	8.3	15.4	12.9	35.0	20.0	15.7

Table 2: **vdpm-gt: Detection and Discrete Pose Estimation Results on PASCAL3D+ dataset. Results obtained considering: 4, 8, 16 and 24 Views.**

Metric	aeroplane	bicycle	boat	bus	car	chair	diningtable	motorbike	sofa	train	tvmonitor	AVG
AP	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
AVP (4 views)	62.9	77.2	31.1	73.7	62.0	46.3	42.8	75.3	48.0	75.1	85.3	61.8
PEAP (4 views)	31.3	47.4	9.0	40.3	30.4	18.5	16.5	46.6	20.5	44.7	59.7	33.2
AVP (8 views)	50.5	64.9	15.9	60.1	44.9	35.4	35.1	58.6	47.4	64.9	63.3	49.2
PEAP (8 views)	19.2	31.8	2.3	26.8	15.6	9.7	10.6	27.7	17.5	32.5	31.7	20.5
AVP (16 views)	30.9	33.0	9.5	62.3	29.7	18.0	23.1	30.6	31.0	52.2	44.1	33.1
PEAP (16 views)	7.3	9.1	0.8	27.6	6.5	2.7	4.3	7.8	7.3	20.4	15.1	9.9
AVP (24 views)	17.8	24.6	5.9	54.4	24.6	16.3	21.7	24.9	24.2	53.9	26.0	26.8
PEAP (24 views)	2.5	5.1	0.3	21.3	4.4	2.1	3.8	5.3	4.4	21.4	5.3	6.9

Figure 1 summarizes the impact on the pose performance of each type of error. Figure 1a reports, for the correct detections obtained by the vdpm-gt model, the frequency and impact on the pose performance of each type of false positive. For this figure a pose estimation is considered as: a) correct if its error is $< 15°$; b) opposite if its pose error is $> 165°$; c) nearby if its pose error is $\in [15°; 30°]$; d) other for the rest of situations. Figure 1b shows the impact of different type of pose errors. Figure 2 shows the success rate considering the 8 viewpoints described in Section 1. Hence, considering only the correct detections, Figure 1b also summarizes the percentage of success and error on pose estimation.



(a) (b)

Figure 1: **Analysis and Impact of Pose Estimation Errors.** (a) Pie chart shows the fraction of pose errors that are due to Opposite viewpoints (Opposite), confusion with Nearby viewpoints (Nearby), confusion with Other rotations (Other). It also reports the percentage of correct pose estimations (Correct). (b) Impact of Pose Errors. Blue Bars display the AVP performance obtained when all estimations are considered. Green Bars display the AVP improvement by **removing** all estimations of one type: **OTH** removes confusion with other rotation viewpoints; **NEAR** removes confusion with nearby viewpoints; **OPP** removes confusion with opposite viewpoints. Brick Red Bars show the AVP improvement by **correcting** all estimations of one type: **OTH**, **NEAR** and **OPP**.

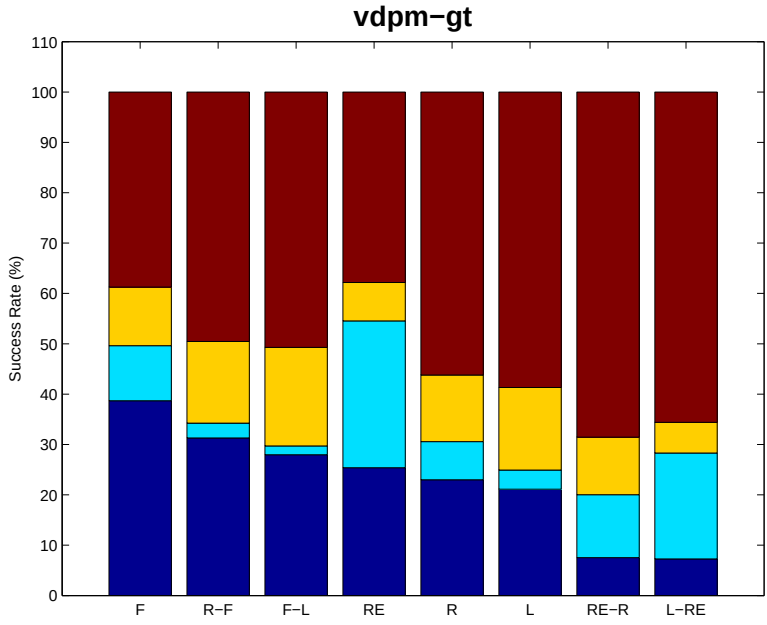
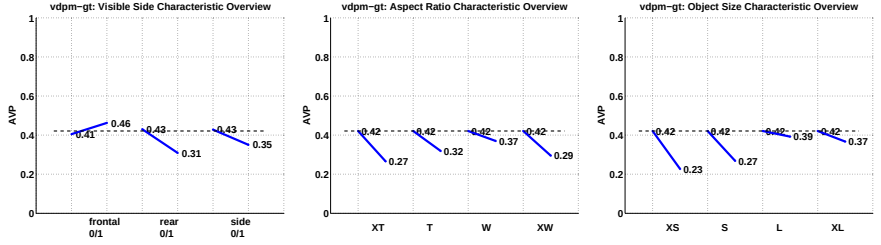


Figure 2: **Success Rate considering 8 viewpoints.** For each of the 8 viewpoints, we report: In Blue, the percentage of the correct pose estimations. In Cyan, the percentage of confusion with the opposite viewpoints. In Yellow, the percentage of confusion with nearby viewpoints. In Brick Red, the percentage of the confusions with other rotations.

Figure 3 summarizes the main object characteristic influences. Figure 3a shows the influence of the side visibility. Figures 3b and 3c report the aspect ratio and object size effects, respectively. Figure 4 provides a summary of the sensitivity to each characteristic and the potential impact on improving pose estimation robustness. The worst-performing and best-performing combinations for each object characteristic are averaged over the selected object categories. The difference between the best and the worst performance indicates sensitivity; the difference between the best and the overall indicates the potential impact.



(a) (b) (c)

Figure 3: **Object Characteristic Influences in terms of AVP**. (a) Side Visibility Influence. Visible side: **1** = performance obtained when the corresponding side is visible; **0** = accuracy achieved when the side is not visible. (b) Aspect Ratio Effect. We show the overall accuracy by the black dashed line, and the performance achieved when we only consider: XT , T, W and XW objects. (c) Object Size Effect. We show the overall accuracy by the black dashed line, and the performance achieved when we only consider: XS, S, L and XL objects.

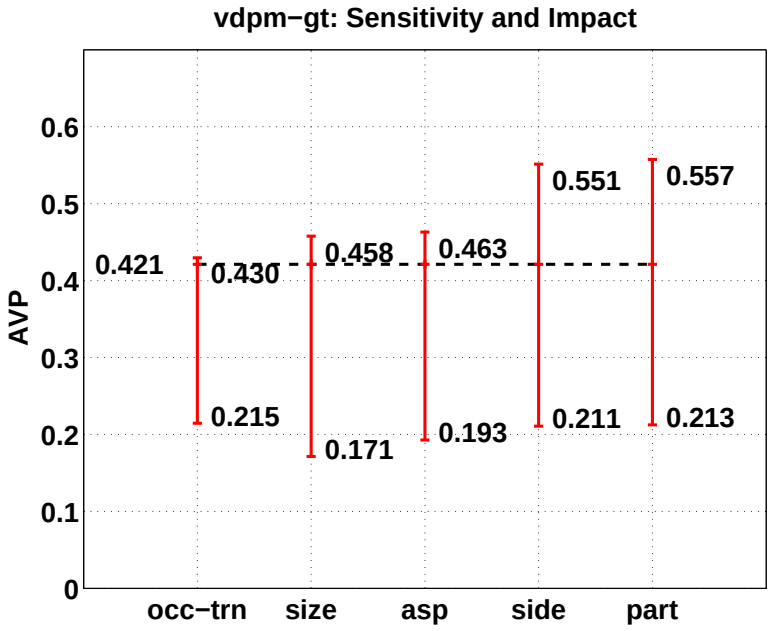
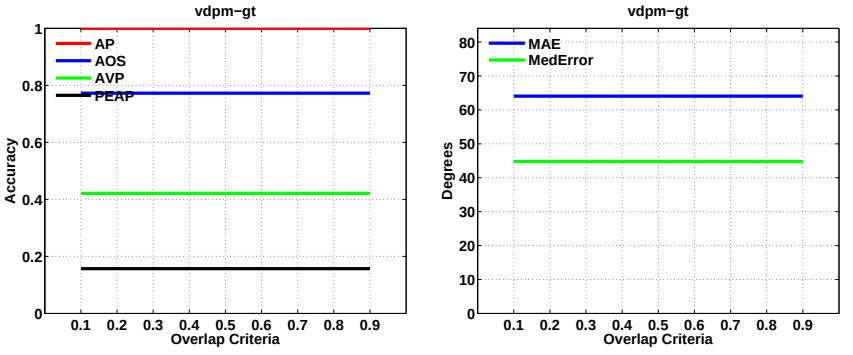


Figure 4: **Summary of Sensitivity and Impact of Object Characteristics.** We show the AVP performance of the highest performing and lowest performing subsets within each characteristic (occlusion/truncation (occ-trn), bounding box area or object size (size), aspect ratio (asp), visible sides (side) and part visibility (part)). Overall accuracy is indicated by the black dashed line. The difference between max and min indicates sensitivity; the difference between max and overall indicates the impact.

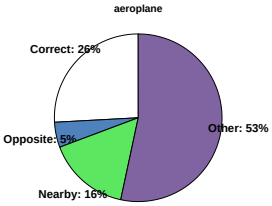
Figure 5 shows an analysis of the influence of the overlap criterion considering all the metrics. Figure 5a shows AOS, AVP and PEAP, and Figure 5b reports MAE and MedError. For this overlap criterion analysis we follow the PASCAL VOC formulation: to be considered a true positive, the area of overlap between the predicted BB and GT BB must exceed a threshold.



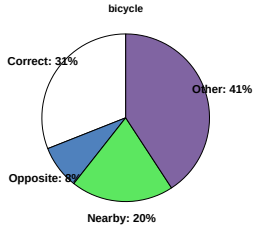
(a) (b)

Figure 5: **Simultaneous Object Detection and Pose Estimation.** Analysis of pose estimation performance considering different metrics and different overlap criteria. (a) AP, AOS, AVP and PEAP. (b) MAE and MedError.

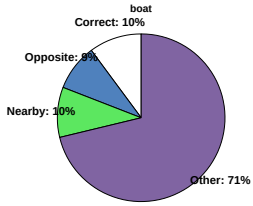
Finally, we summarize the main pose estimation errors (Figure 6) and the success rate considering 8 viewpoints (Figure 7) for each of the selected object classes: aeroplane, bicycle, boat, bus, car, chair, diningtable, motorbike, sofa, train, tvmonitor from PASCAL3D+ dataset.



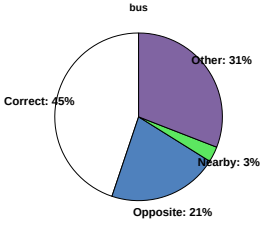
(a) aeroplane



(b) bicycle

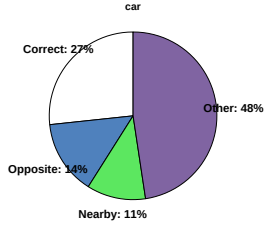


(c) boat

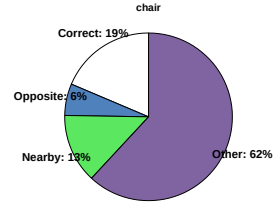


n

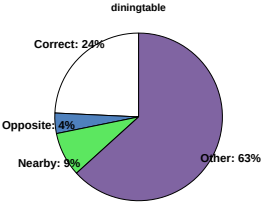
(d) bus



(e) car

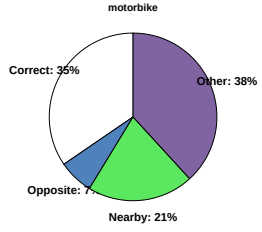


(f) chair

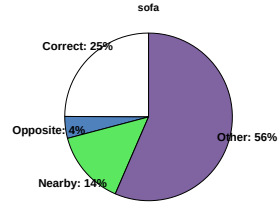


n

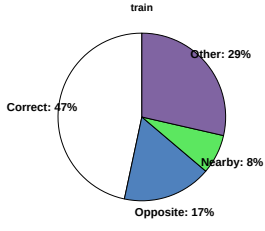
(g) diningtable



(h) motorbike

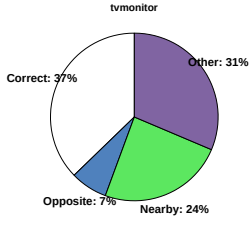


(i) sofa



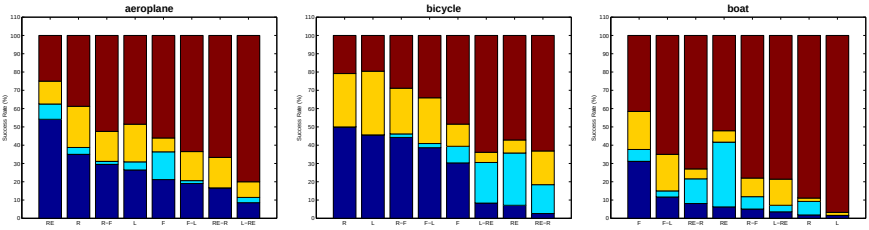
n

(j) train

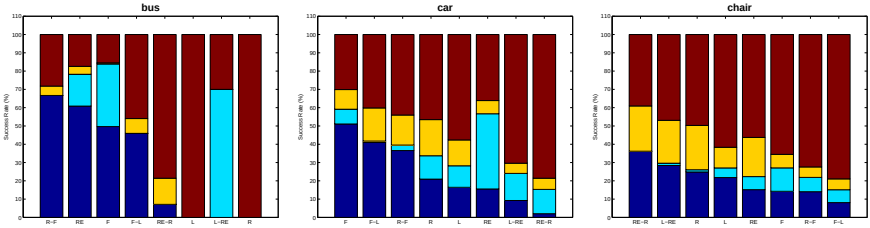


(k) tvmonitor

Figure 6: **Main Pose Estimation Errors for all selected Object Classes from PASCAL3D+ dataset.**

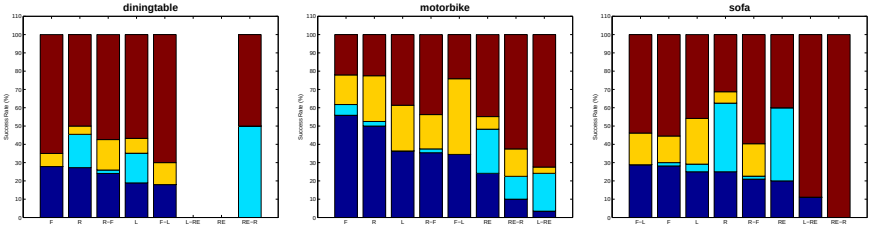


(a) aeroplane (b) bicycle (c) boat



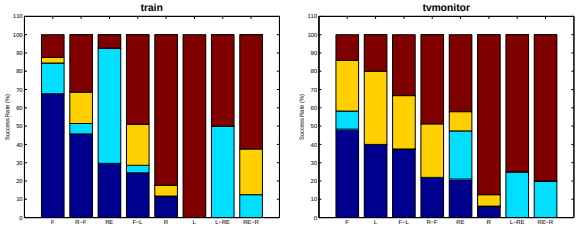
n

(d) bus (e) car (f) chair



n

(g) diningtable (h) motorbike (i) sofa



n

(j) train (k) tvmonitor

Figure 7: **Success Rate considering 8 viewpoints for all selected Object Classes from PASCAL3D+ dataset.**

3 Results per Object Class

3.1 aeroplane

Statistics for aeroplane class:

Number of objects = 479, No difficult objects = 433
Truncated: None = 265, Truncated = 168
Occluded: None = 399, Occluded = 80
Occluded and Truncated: 57
side visible:

```
bottom: Yes = 0  No = 479
front: Yes = 36  No = 443
rear: Yes = 31  No = 448
side: Yes = 260  No = 219
top: Yes = 0  No = 479
```

part visible:

```
part 1 = left_elevator: Yes = 268  No = 211
part 2 = left_wing: Yes = 277  No = 202
part 3 = noselanding: Yes = 367  No = 112
part 4 = right_elevator: Yes = 253  No = 226
part 5 = right_wing: Yes = 265  No = 214
part 6 = rudder_lower: Yes = 290  No = 189
part 7 = rudder_upper: Yes = 358  No = 121
part 8 = tail: Yes = 261  No = 218
```

Figure 8 summarizes the viewpoint distribution on the test set for aeroplane class considering the 8 views described in Section 1. Figure 9 shows the analysis and the impact on the pose performance of each type of pose error. We also report in Figure 10 the success rate for aeroplane class considering each of the 8 viewpoints described in Section 1.

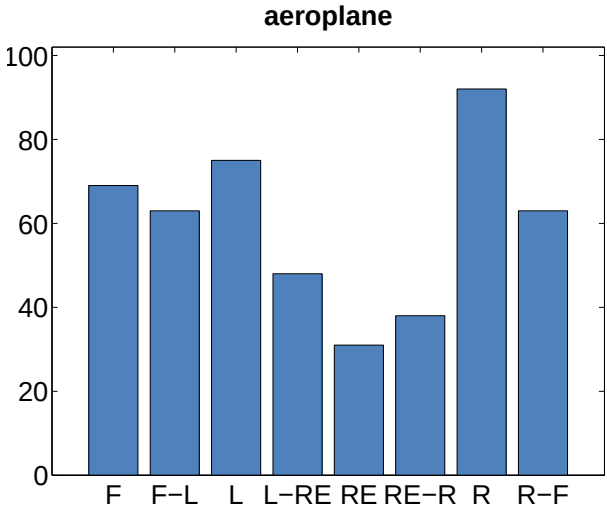
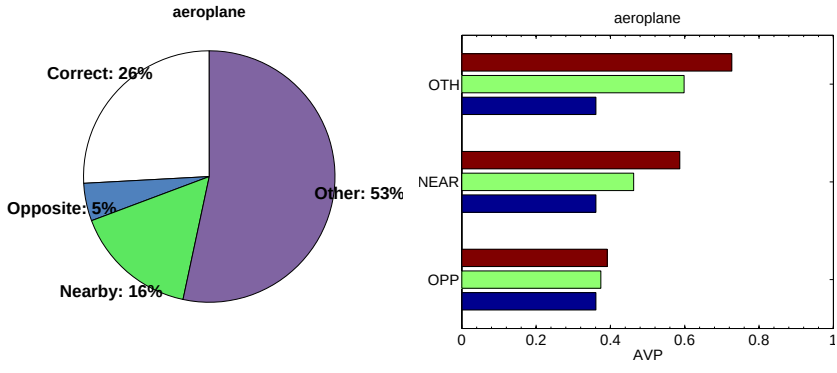


Figure 8: **Analysis of Pose Distribution.** Histogram represents viewpoint distribution on test set for aeroplane class from PASCAL3D+ dataset.



(a) (b)

Figure 9: **Analysis of the Main Pose Estimation Errors.** (a) Given the correct detections, pie chart shows the fraction of pose error that are due to Opposite viewpoints (Opposite), confusion with Nearby viewpoints (Nearby), confusion with Other rotations (Other), and correct pose estimations (Correct) for aeroplane class. (b) Impact of pose errors in terms of AVP. Blue Bars show the AVP performance obtained when all estimations are considered. Green Bars display the AVP improvement by removing all estimations of one type: **OTH** removes confusion with other rotation viewpoints; **NEAR** removes confusion with nearby viewpoints; **OPP** removes confusion with opposite viewpoints. Brick Red Bars display the AVP improvement by correcting all estimations of one type: **OTH** corrects confusion with other rotation viewpoints; **NEAR** corrects confusion with nearby viewpoints; **OPP** corrects confusion with opposite viewpoints.

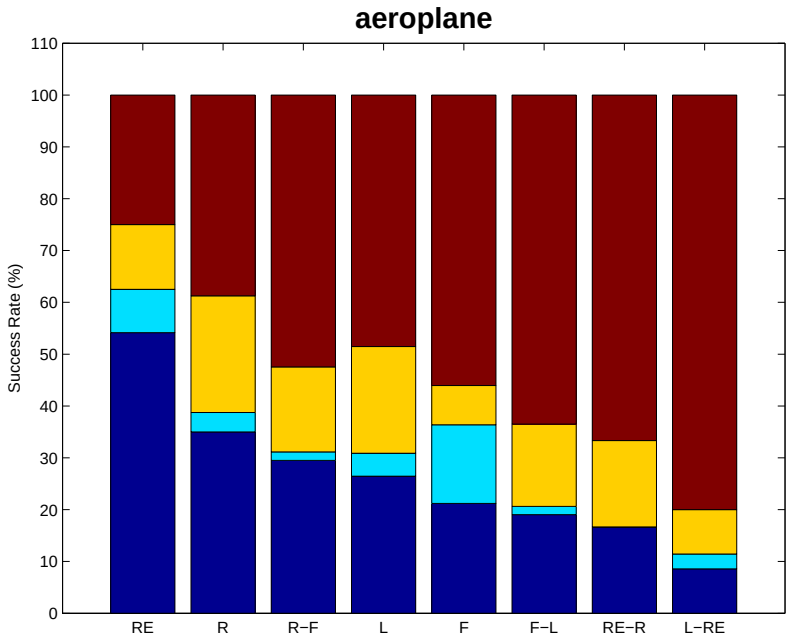
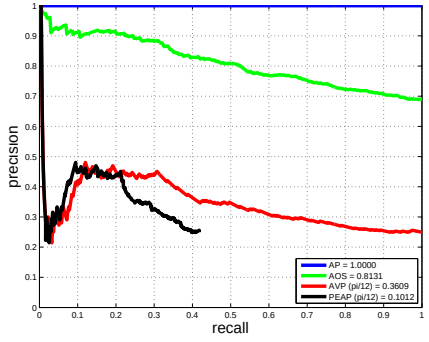
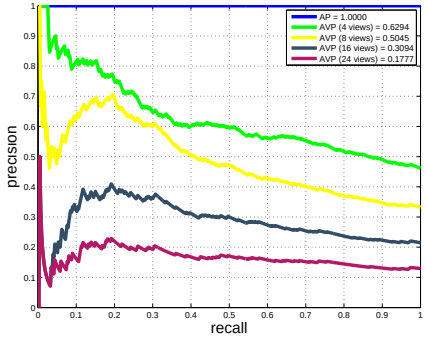


Figure 10: **Success rate for aeroplane class considering 8 views.** For each of the 8 viewpoints, we report: In Blue, the percentage of the correct pose estimations. In Cyan, the percentage of confusion with the opposite viewpoints. In Yellow, the percentage of confusion with nearby viewpoints. In Brick Red, the percentage of the confusions with other rotations.

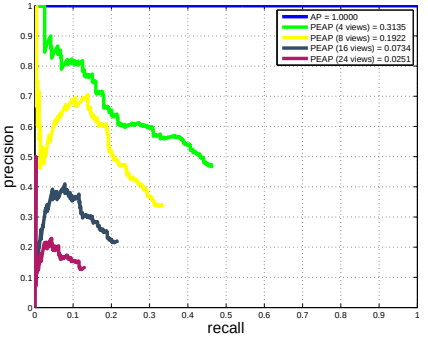
Figure 11 summarizes the detection and pose estimation performances for aeroplane class. Figure 11a shows the model performance working on continuous pose estimation. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$. Figures 11b and 11c report the AVP and PEAP performances achieve working on discrete pose estimation. The AVP and PEAP metrics are obtained considering different number of the views: 4, 8, 16 and 24.



(a)



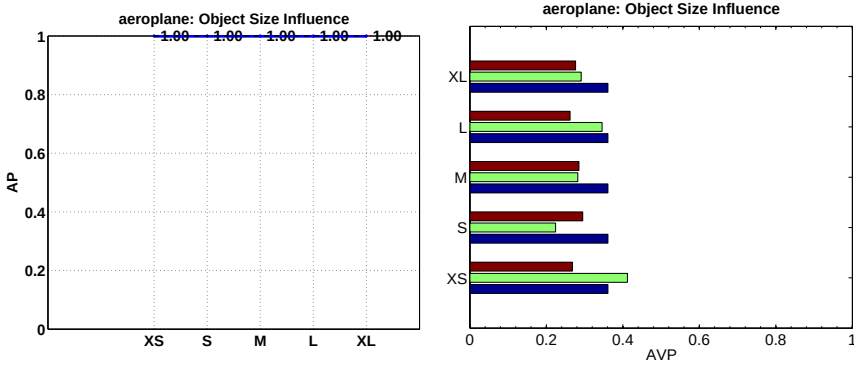
(b)



(c)

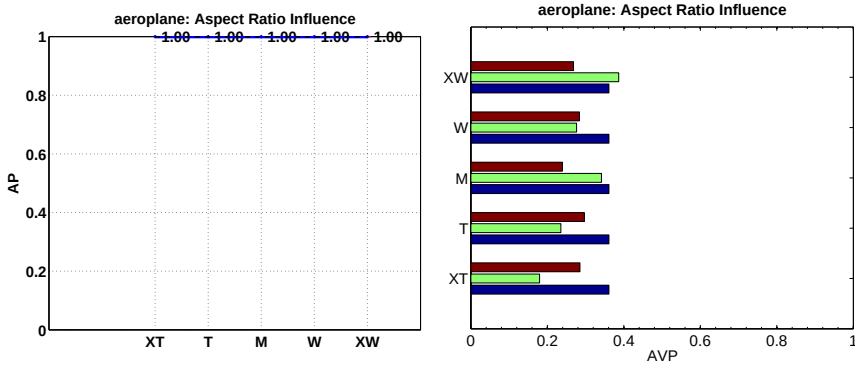
Figure 11: **Precision/Recall Curves.** (a) Performance working on continuous pose estimation. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$. (b) AVP and (c) PEAP performances working on discrete pose estimation. Results obtained by considering 4, 8, 16, and 24 views.

Figures 12, 13, 14, 15 and 16 summarize the main object characteristic influences on detection and pose estimation performances for aeroplane class. Figure 12 reports the effect of object size characteristic on object detection and pose estimation performances. Figure 13 shows the influence of aspect ratio characteristic on object detection and pose estimation performances. Figure 14 represents the influence of occluded and truncated objects on detection and pose estimation performances. Figure 15 shows the visible side influence. Figure 16 reports the part visibily effect.



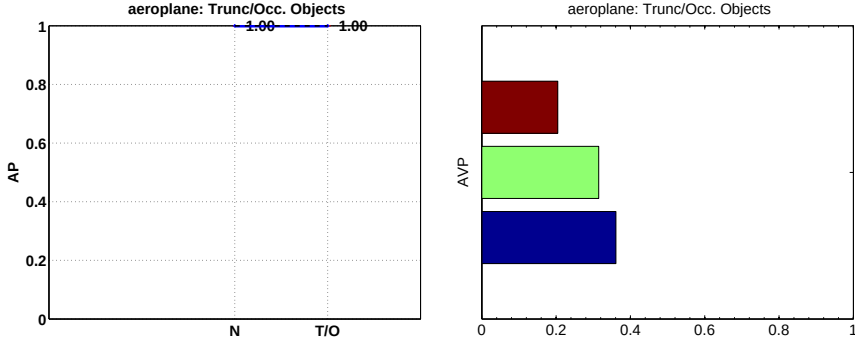
(a) (b)

Figure 12: **Effect of Object Size Characteristic on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP variance. XS and S are considered EXTRA SMALL and SMALL objects, M are MEDIUM objects, L and XL are LARGE and EXTRA LARGE objects. (b) Pose Estimation Results in terms of AVP. Blue Bars show the results when all pose estimations are considered, Green Bars show the results when only the pose estimations corresponding this object size are considered and Brick Red Bars display the results when the pose estimations corresponding this object size are removed.



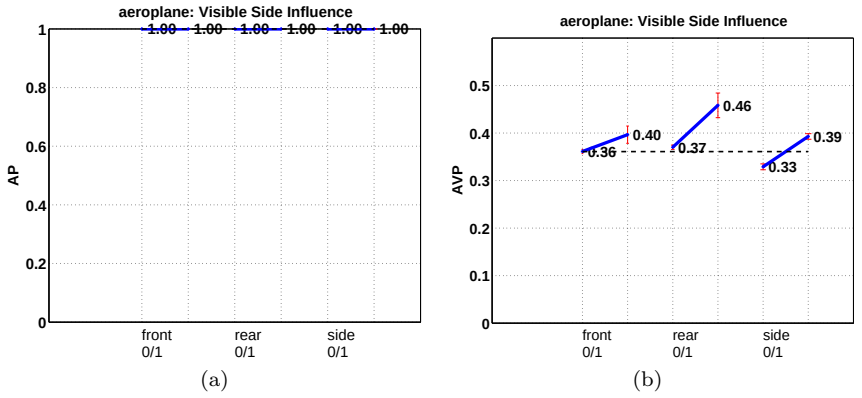
(a) (b)

Figure 13: **Effect of Object Size Characteristic on Object Detection and Pose Estimation.** (a) Detection analysis: AP variance. XS and S are considered EXTRA SMALL and SMALL objects, M are MEDIUM objects, L and XL are LARGE and EXTRA LARGE objects. (b) Pose Estimation Results in terms of AVP. Blue Bars show the results when all pose estimations are considered, Green Bars show the results when only the pose estimations corresponding this object size are considered and Brick Red Bars display the results when the pose estimations corresponding this object size are removed.



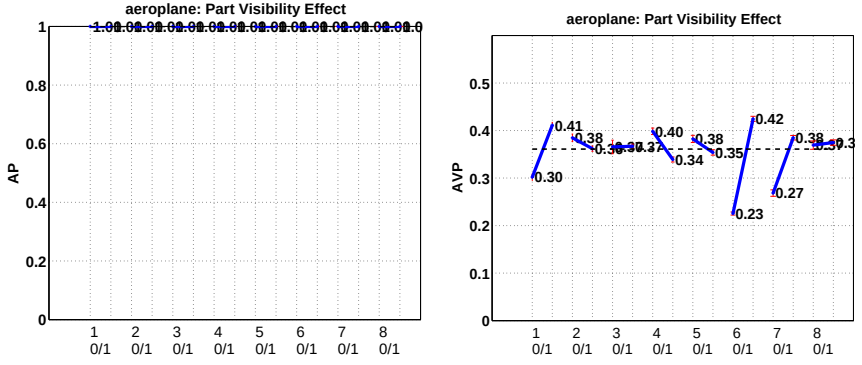
(a) (b)

Figure 14: **Effect of Occluded and Truncated Objects on Detection and Pose Estimation Performances.** (a) Detection analysis: AP variance. N non occluded or truncated object, T/O truncated or occluded object. Results on Pose Estimation in terms of the AVP metric. Blue Bars show the results when all pose estimations are considered. Green Bars display the results when the occluded and truncated objects are removed. Brick Red Bars show the performance when we are only considered the occluded and truncated objects.



(a) (b)

Figure 15: **Effect of Visible Sides on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP performance. Pose Estimation analysis: (b) Pose Estimation analysis: AVP performance. Black dashed lines indicate overall. Visible side: **1** = performance when the side is visible; **0** = performance when the side is not visible.



(a) (b)

Figure 16: **Part Visibility Influence on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP performance. Pose Estimation analysis: (b) Pose Estimation analysis: AVP performance. Black dashed lines indicate overall performance. Visible part: **1** = performance when the part is visible; **0** = performance when the part is not visible. The correspondence between the part number and the part name is indicated at the beginning of this Section (see Statistics for aeroplane class).

Figure 17 provides a summary of the sensitivity to each characteristic and the potential impact on improving pose estimation robustness. The worst-performing and best-performing combinations for each object characteristic are averaged over the selected object categories. The difference between the best and the worst performance indicates sensitivity; the difference between the best and the overall indicates the potential impact.

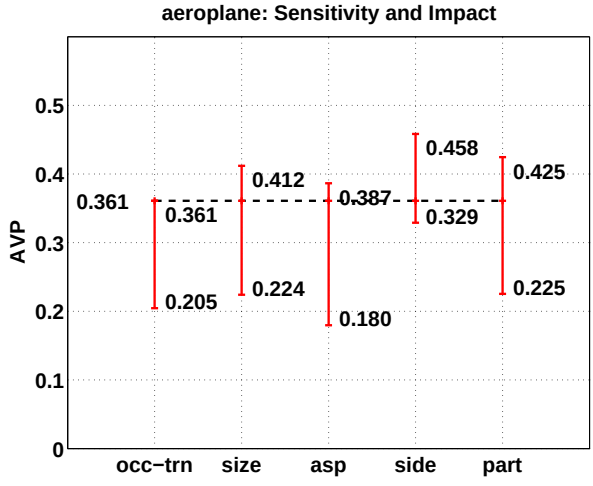


Figure 17: **Summary of Sensitivity and Impact of Object Characteristics.** We show the performance of the highest performing and lowest performing subsets within each characteristic (occlusion/truncation (occ-trn), bounding box area or object size (size), aspect ratio (asp), visible sides (side) and part visibility (part)). Overall accuracy is indicated by the black dashed line. The difference between max and min indicates sensitivity; the difference between max and overall indicates the impact.

Figure 18 shows an analysis of the influence of the overlap criterion considering the AVP metric. For this overlap criterion analysis we follow the PASCAL VOC formulation: to be considered a true positive, the area of overlap between the predicted BB and GT BB must exceed a threshold.

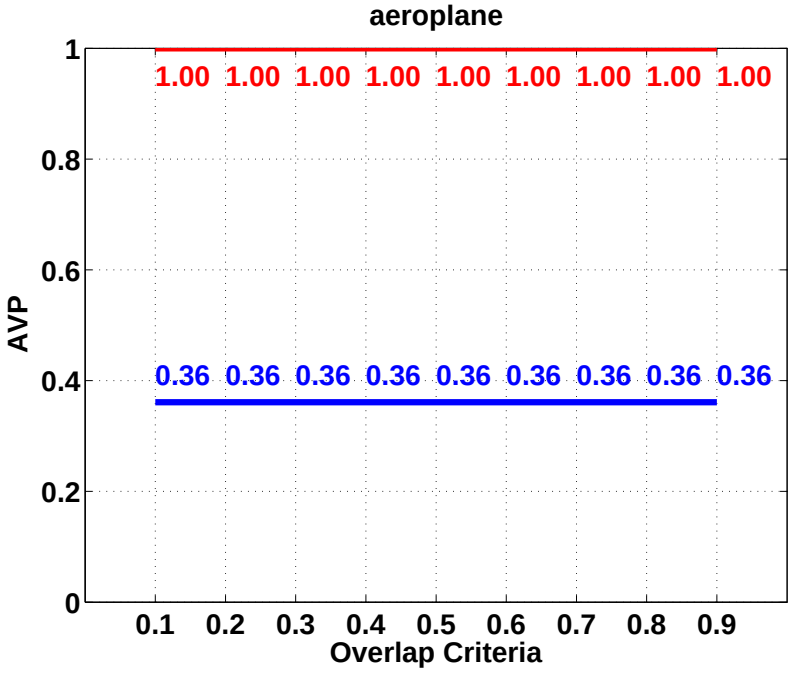


Figure 18: **Simultaneous Object Detection and Pose Estimation.** The detection performance (AP) is represented in red and the pose estimation performance (AVP) in blue.

3.2 bicycle

Statistics for bicycle class:

Number of objects = 378, No difficult objects = 358
Truncated: None = 218, Truncated = 140
Occluded: None = 194, Occluded = 184
Occluded and Truncated: 111
side visible:

```
bottom: Yes = 0  No = 378
front: Yes = 54  No = 324
rear: Yes = 24  No = 354
side: Yes = 189  No = 189
top: Yes = 0  No = 378
```

part visible:

```
part 1 = head_center: Yes = 276  No = 102
part 2 = left_back_wheel: Yes = 161  No = 217
part 3 = left_front_wheel: Yes = 212  No = 166
part 4 = left_handle: Yes = 261  No = 117
part 5 = left_pedal_center: Yes = 132  No = 246
part 6 = right_back_wheel: Yes = 157  No = 221
part 7 = right_front_wheel: Yes = 202  No = 176
part 8 = right_handle: Yes = 254  No = 124
part 9 = right_pedal_center: Yes = 109  No = 269
part 10 = seat_back: Yes = 216  No = 162
part 11 = seat_front: Yes = 199  No = 179
```

Figure 19 summarizes the viewpoint distribution on the test set for bicycle class considering the 8 views described in Section 1. Figure 20 shows the analysis and the impact on the pose performance of each type of pose error. We also report in Figure 21 the success rate for bicycle class considering each of the 8 viewpoints described in Section 1.

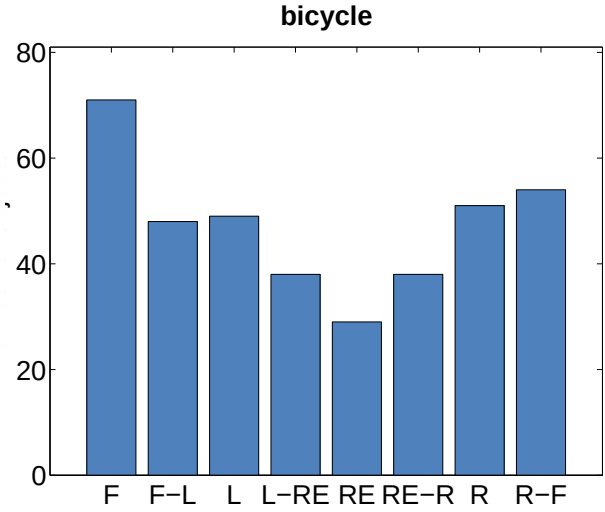
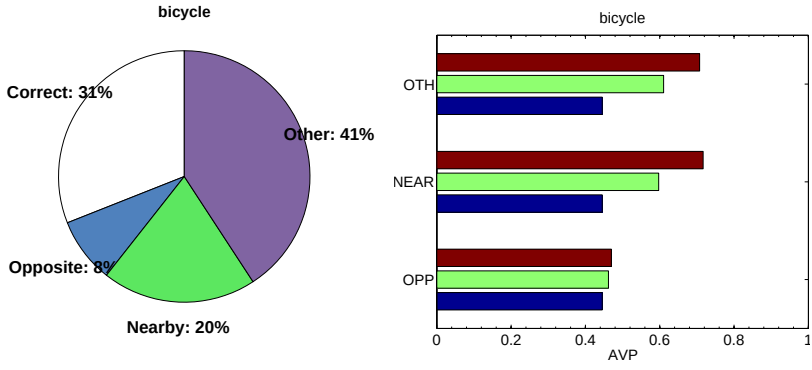


Figure 19: **Analysis of Pose Distribution.** Histogram represents viewpoint distribution on test set for bicycle class from PASCAL3D+ dataset.



(a) (b)

Figure 20: **Analysis of the Main Pose Estimation Errors.** (a) Given the correct detections, pie chart shows the fraction of pose error that are due to Opposite viewpoints (Opposite), confusion with Nearby viewpoints (Nearby), confusion with Other rotations (Other), and correct pose estimations (Correct) for bicycle class. (b) Impact of pose errors in terms of AVP. Blue Bars show the AVP performance obtained when all estimations are considered. Green Bars display the AVP improvement by removing all estimations of one type: **OTH** removes confusion with other rotation viewpoints; **NEAR** removes confusion with nearby viewpoints; **OPP** removes confusion with opposite viewpoints. Brick Red Bars display the AVP improvement by correcting all estimations of one type: **OTH** corrects confusion with other rotation viewpoints; **NEAR** corrects confusion with nearby viewpoints; **OPP** corrects confusion with opposite viewpoints.

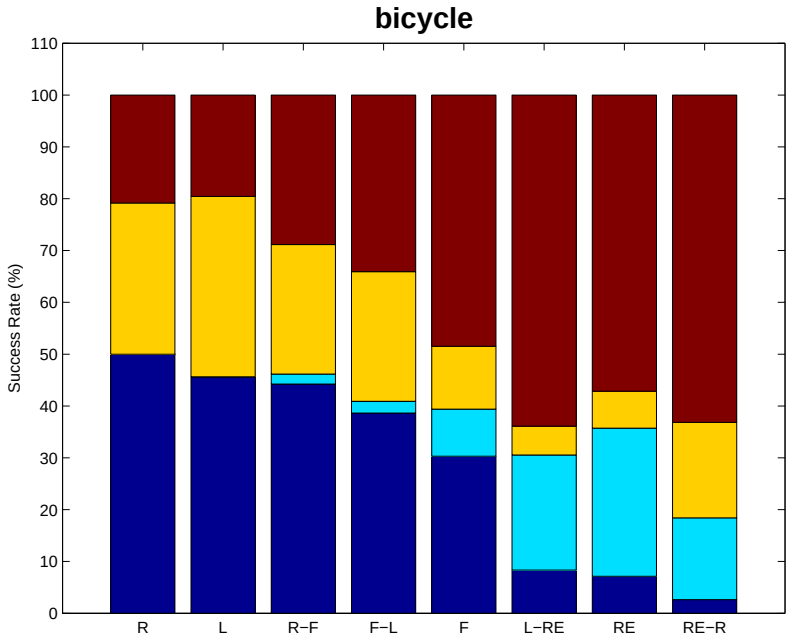
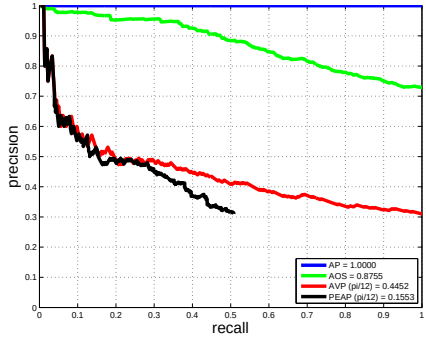
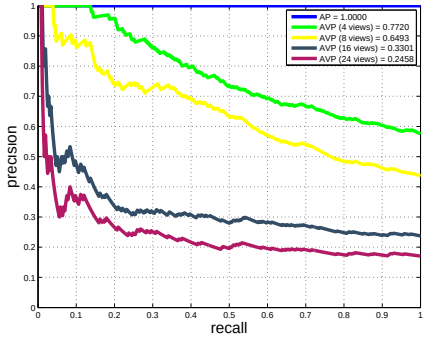


Figure 21: **Success rate for bicycle class considering 8 views.** For each of the 8 viewpoints, we report: In Blue, the percentage of the correct pose estimations. In Cyan, the percentage of confusion with the opposite viewpoints. In Yellow, the percentage of confusion with nearby viewpoints. In Brick Red, the percentage of the confusions with other rotations.

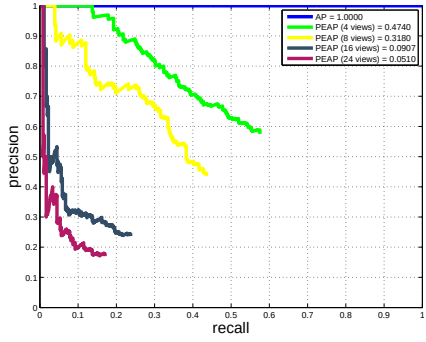
Figure 22 summarizes the detection and pose estimation performances for bicycle class. Figure 22a shows the model performance working on continuous pose estimation. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$. Figures 22b and 22c report the AVP and PEAP performances achieve working on discrete pose estimation. The AVP and PEAP metrics are obtained considering different number of the views: 4, 8, 16 and 24.



(a)



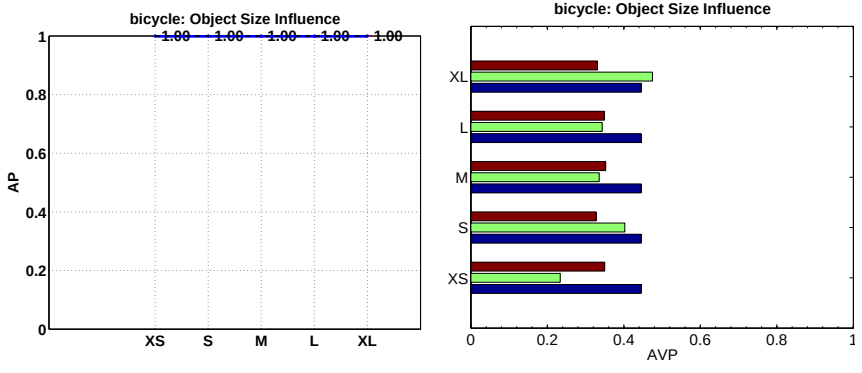
(b)



(c)

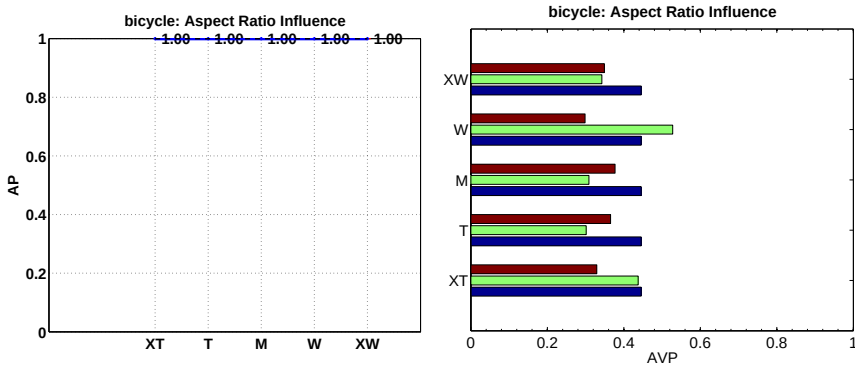
Figure 22: **Precision/Recall Curves.** (a) Performance working on continuous pose estimation. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$. (b) AVP and (c) PEAP performances working on discrete pose estimation. Results obtained by considering 4, 8, 16, and 24 views.

Figures 23, 24, 25, 26 and 27 summarize the main object characteristic influences on detection and pose estimation performances for bicycle class. Figure 23 reports the effect of object size characteristic on object detection and pose estimation performances. Figure 24 shows the influence of aspect ratio characteristic on object detection and pose estimation performances. Figure 25 represents the influence of occluded and truncated objects on detection and pose estimation performances. Figure 26 shows the visible side influence. Figure 27 reports the part visibily effect.



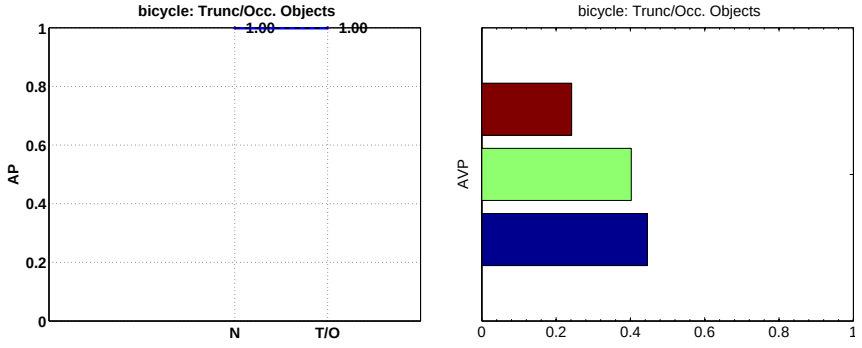
(a) (b)

Figure 23: **Effect of Object Size Characteristic on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP variance. XS and S are considered EXTRA SMALL and SMALL objects, M are MEDIUM objects, L and XL are LARGE and EXTRA LARGE objects. (b) Pose Estimation Results in terms of AVP. Blue Bars show the results when all pose estimations are considered, Green Bars show the results when only the pose estimations corresponding this object size are considered and Brick Red Bars display the results when the pose estimations corresponding this object size are removed.



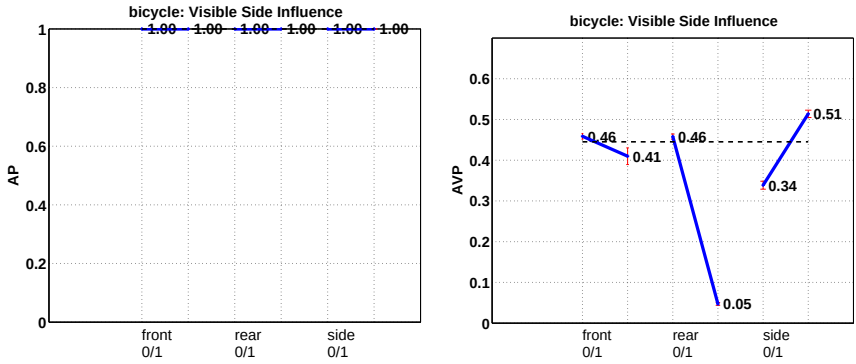
(a) (b)

Figure 24: **Effect of Object Size Characteristic on Object Detection and Pose Estimation.** (a) Detection analysis: AP variance. XS and S are considered EXTRA SMALL and SMALL objects, M are MEDIUM objects, L and XL are LARGE and EXTRA LARGE objects. (b) Pose Estimation Results in terms of AVP. Blue Bars show the results when all pose estimations are considered, Green Bars show the results when only the pose estimations corresponding this object size are considered and Brick Red Bars display the results when the pose estimations corresponding this object size are removed.



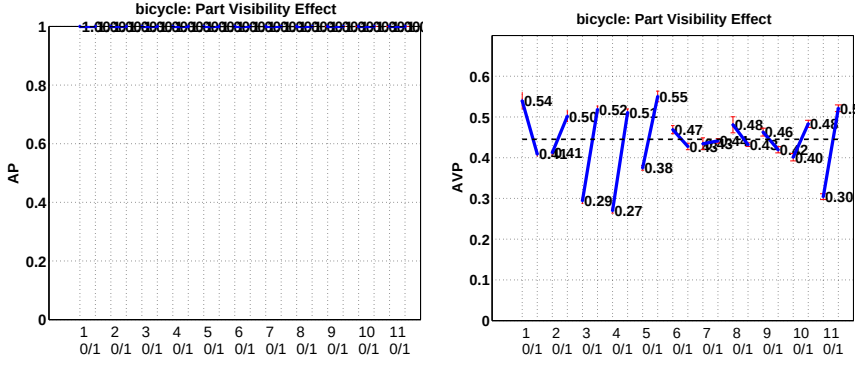
(a) (b)

Figure 25: **Effect of Occluded and Truncated Objects on Detection and Pose Estimation Performances.** (a) Detection analysis: AP variance. N non occluded or truncated object, T/O truncated or occluded object. Results on Pose Estimation in terms of the AVP metric. Blue Bars show the results when all pose estimations are considered. Green Bars display the results when the occluded and truncated objects are removed. Brick Red Bars show the performance when we are only considered the occluded and truncated objects.



(a) (b)

Figure 26: **Effect of Visible Sides on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP performance. Pose Estimation analysis: (b) Pose Estimation analysis: AVP performance. Black dashed lines indicate overall. Visible side: **1** = performance when the side is visible; **0** = performance when the side is not visible.



(a) (b)

Figure 27: **Part Visibility Influence on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP performance. Pose Estimation analysis: (b) Pose Estimation analysis: AVP performance. Black dashed lines indicate overall performance. Visible part: **1** = performance when the part is visible; **0** = performance when the part is not visible. The correspondence between the part number and the part name is indicated at the beginning of this Section (see Statistics for bicycle class).

Figure 28 provides a summary of the sensitivity to each characteristic and the potential impact on improving pose estimation robustness. The worst-performing and best-performing combinations for each object characteristic are averaged over the selected object categories. The difference between the best and the worst performance indicates sensitivity; the difference between the best and the overall indicates the potential impact.

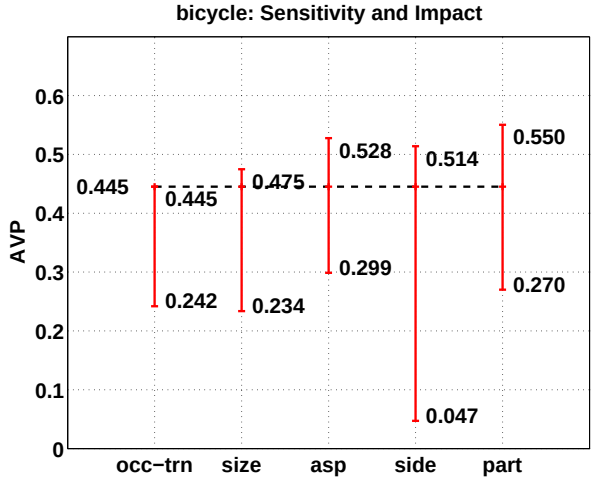


Figure 28: **Summary of Sensitivity and Impact of Object Characteristics.** We show the performance of the highest performing and lowest performing subsets within each characteristic (occlusion/truncation (occ-trn), bounding box area or object size (size), aspect ratio (asp), visible sides (side) and part visibility (part)). Overall accuracy is indicated by the black dashed line. The difference between max and min indicates sensitivity; the difference between max and overall indicates the impact.

Figure 29 shows an analysis of the influence of the overlap criterion considering the AVP metric. For this overlap criterion analysis we follow the PASCAL VOC formulation: to be considered a true positive, the area of overlap between the predicted BB and GT BB must exceed a threshold.

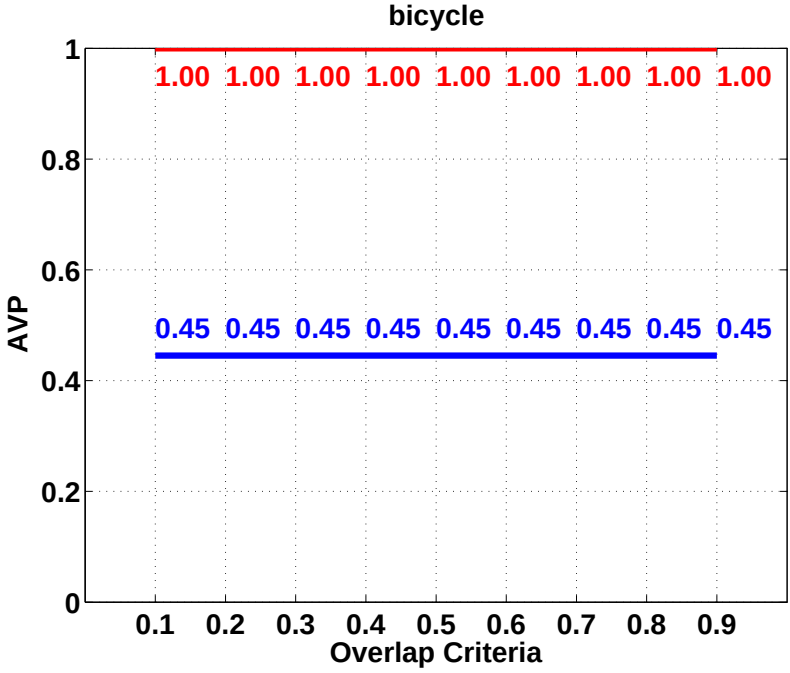


Figure 29: **Simultaneous Object Detection and Pose Estimation.** The detection performance (AP) is represented in red and the pose estimation performance (AVP) in blue.

3.3 boat

Statistics for boat class:

Number of objects = 486, No difficult objects = 424
Truncated: None = 273, Truncated = 151
Occluded: None = 353, Occluded = 133
Occluded and Truncated: 75
side visible:

```
bottom: Yes = 0  No = 486
front: Yes = 45  No = 441
rear: Yes = 28  No = 458
side: Yes = 198  No = 288
top: Yes = 0  No = 486
```

part visible:

```
part 1 = head: Yes = 319  No = 167
part 2 = head_down: Yes = 257  No = 229
part 3 = head_left: Yes = 200  No = 286
part 4 = head_right: Yes = 189  No = 297
part 5 = tail_left: Yes = 201  No = 285
part 6 = tail_right: Yes = 174  No = 312
part 7 = tail: Yes = 15  No = 471
```

Figure 30 summarizes the viewpoint distribution on the test set for boat class considering the 8 views described in Section 1. Figure 31 shows the analysis and the impact on the pose performance of each type of pose error. We also report in Figure 32 the success rate for boat class considering each of the 8 viewpoints described in Section 1.

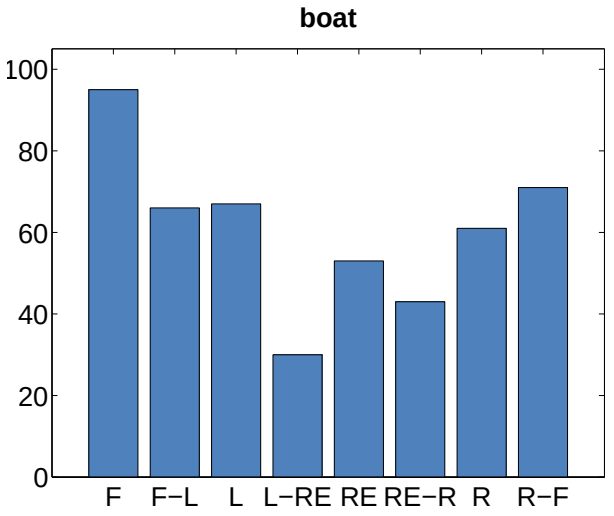
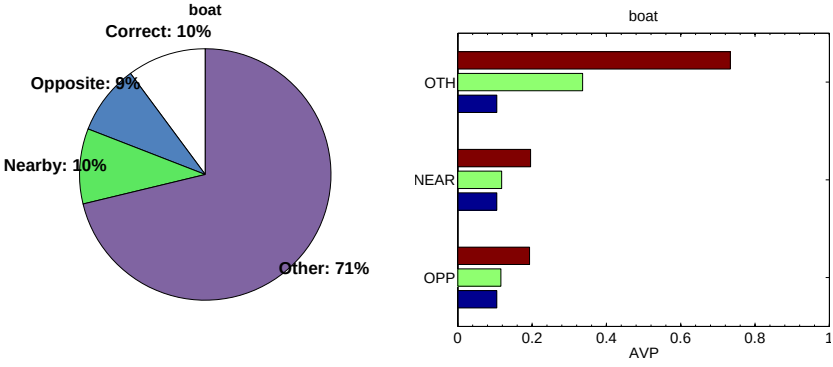


Figure 30: **Analysis of Pose Distribution.** Histogram represents viewpoint distribution on test set for boat class from PASCAL3D+ dataset.



(a) (b)

Figure 31: **Analysis of the Main Pose Estimation Errors.** (a) Given the correct detections, pie chart shows the fraction of pose error that are due to Opposite viewpoints (Opposite), confusion with Nearby viewpoints (Nearby), confusion with Other rotations (Other), and correct pose estimations (Correct) for boat class. (b) Impact of pose errors in terms of AVP. Blue Bars show the AVP performance obtained when all estimations are considered. Green Bars display the AVP improvement by removing all estimations of one type: **OTH** removes confusion with other rotation viewpoints; **NEAR** removes confusion with nearby viewpoints; **OPP** removes confusion with opposite viewpoints. Brick Red Bars display the AVP improvement by correcting all estimations of one type: **OTH** corrects confusion with other rotation viewpoints; **NEAR** corrects confusion with nearby viewpoints; **OPP** corrects confusion with opposite viewpoints.

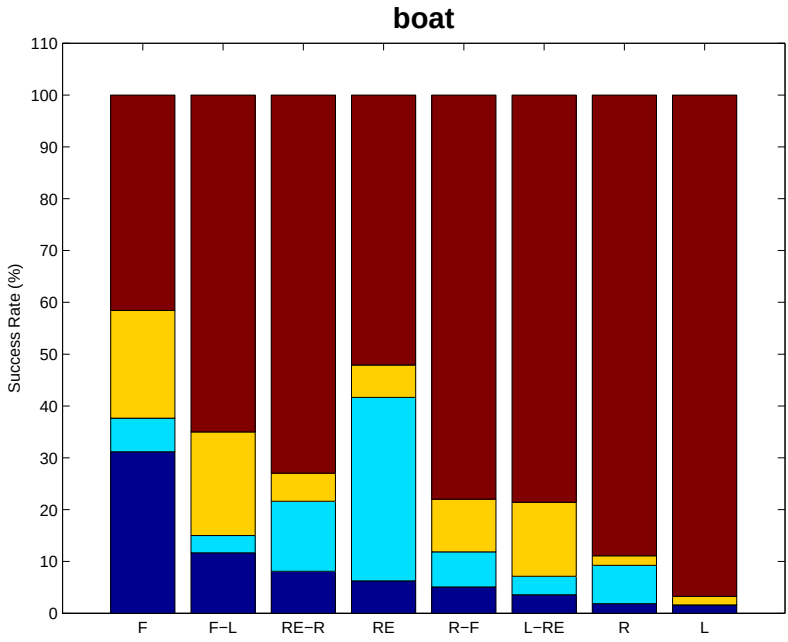
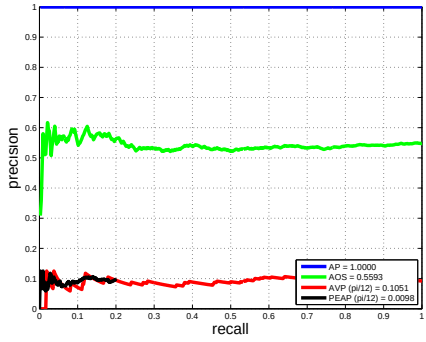
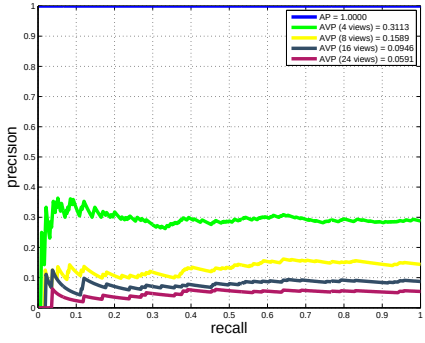


Figure 32: **Success rate for boat class considering 8 views.** For each of the 8 viewpoints, we report: In Blue, the percentage of the correct pose estimations. In Cyan, the percentage of confusion with the opposite viewpoints. In Yellow, the percentage of confusion with nearby viewpoints. In Brick Red, the percentage of the confusions with other rotations.

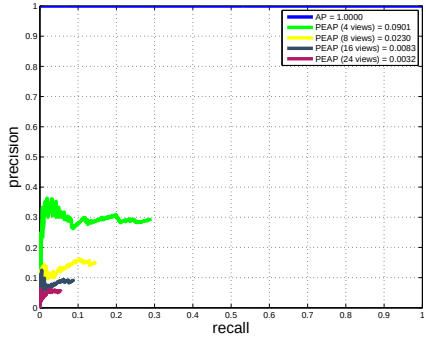
Figure 33 summarizes the detection and pose estimation performances for boat class. Figure 33a shows the model performance working on continuous pose estimation. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$. Figures 33b and 33c report the AVP and PEAP performances achieve working on discrete pose estimation. The AVP and PEAP metrics are obtained considering different number of the views: 4, 8, 16 and 24.



(a)



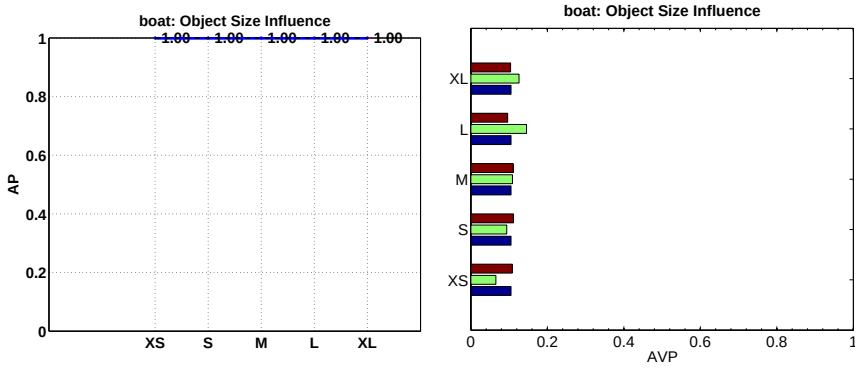
(b)



(c)

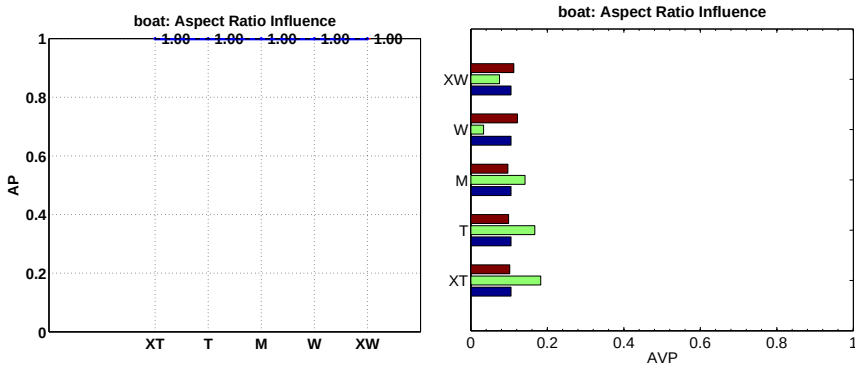
Figure 33: **Precision/Recall Curves.** (a) Performance working on continuous pose estimation. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$. (b) AVP and (c) PEAP performances working on discrete pose estimation. Results obtained by considering 4, 8, 16, and 24 views.

Figures 34, 35, 36, 37 and 38 summarize the main object characteristic influences on detection and pose estimation performances for boat class. Figure 34 reports the effect of object size characteristic on object detection and pose estimation performances. Figure 35 shows the influence of aspect ratio characteristic on object detection and pose estimation performances. Figure 36 represents the influence of occluded and truncated objects on detection and pose estimation performances. Figure 37 shows the visible side influence. Figure 38 reports the part visibily effect.



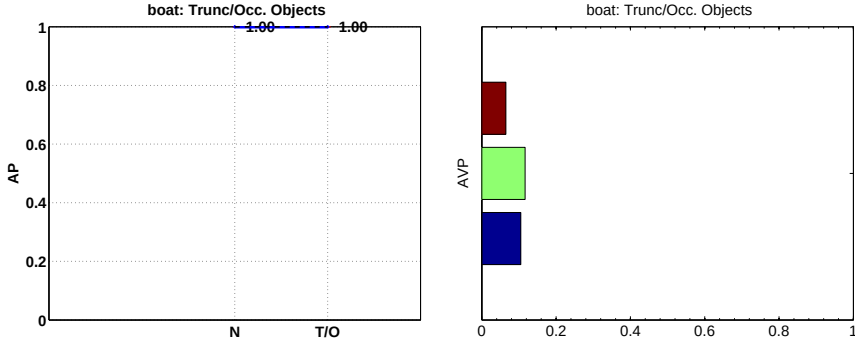
(a) (b)

Figure 34: **Effect of Object Size Characteristic on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP variance. XS and S are considered EXTRA SMALL and SMALL objects, M are MEDIUM objects, L and XL are LARGE and EXTRA LARGE objects. (b) Pose Estimation Results in terms of AVP. Blue Bars show the results when all pose estimations are considered, Green Bars show the results when only the pose estimations corresponding this object size are considered and Brick Red Bars display the results when the pose estimations corresponding this object size are removed.



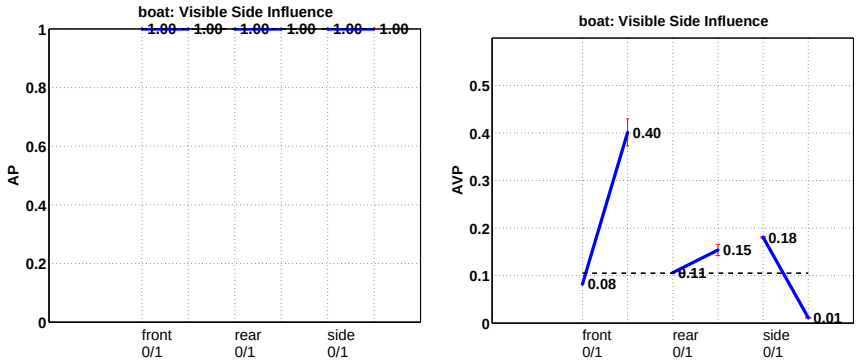
(a) (b)

Figure 35: **Effect of Object Size Characteristic on Object Detection and Pose Estimation.** (a) Detection analysis: AP variance. XS and S are considered EXTRA SMALL and SMALL objects, M are MEDIUM objects, L and XL are LARGE and EXTRA LARGE objects. (b) Pose Estimation Results in terms of AVP. Blue Bars show the results when all pose estimations are considered, Green Bars show the results when only the pose estimations corresponding this object size are considered and Brick Red Bars display the results when the pose estimations corresponding this object size are removed.



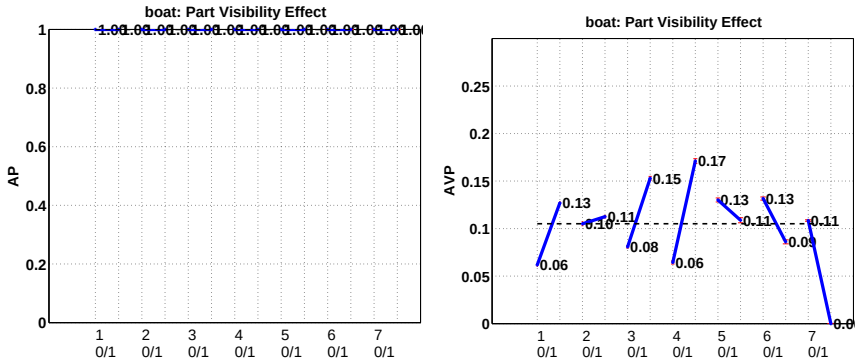
(a) (b)

Figure 36: **Effect of Occluded and Truncated Objects on Detection and Pose Estimation Performances.** (a) Detection analysis: AP variance. N non occluded or truncated object, T/O truncated or occluded object. Results on Pose Estimation in terms of the AVP metric. Blue Bars show the results when all pose estimations are considered. Green Bars display the results when the occluded and truncated objects are removed. Brick Red Bars show the performance when we are only considered the occluded and truncated objects.



(a) (b)

Figure 37: **Effect of Visible Sides on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP performance. Pose Estimation analysis: (b) Pose Estimation analysis: AVP performance. Black dashed lines indicate overall. Visible side: **1** = performance when the side is visible; **0** = performance when the side is not visible.



(a) (b)

Figure 38: **Part Visibility Influence on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP performance. Pose Estimation analysis: (b) Pose Estimation analysis: AVP performance. Black dashed lines indicate overall performance. Visible part: **1** = performance when the part is visible; **0** = performance when the part is not visible. The correspondence between the part number and the part name is indicated at the beginning of this Section (see Statistics for boat class).

Figure 39 provides a summary of the sensitivity to each characteristic and the potential impact on improving pose estimation robustness. The worst-performing and best-performing combinations for each object characteristic are averaged over the selected object categories. The difference between the best and the worst performance indicates sensitivity; the difference between the best and the overall indicates the potential impact.

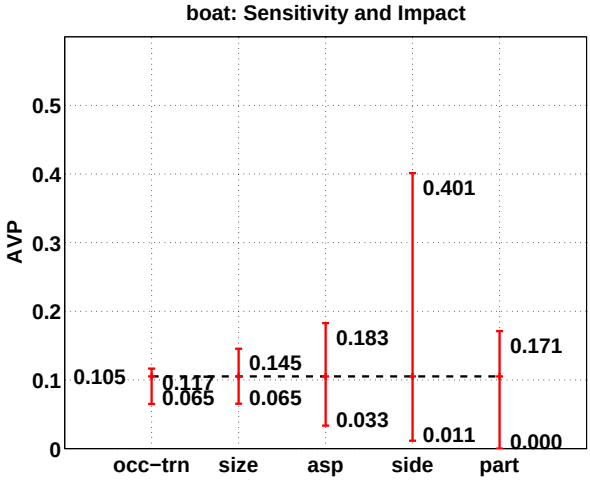


Figure 39: **Summary of Sensitivity and Impact of Object Characteristics.** We show the performance of the highest performing and lowest performing subsets within each characteristic (occlusion/truncation (occ-trn), bounding box area or object size (size), aspect ratio (asp), visible sides (side) and part visibility (part)). Overall accuracy is indicated by the black dashed line. The difference between max and min indicates sensitivity; the difference between max and overall indicates the impact.

Figure 40 shows an analysis of the influence of the overlap criterion considering the AVP metric. For this overlap criterion analysis we follow the PASCAL VOC formulation: to be considered a true positive, the area of overlap between the predicted BB and GT BB must exceed a threshold.

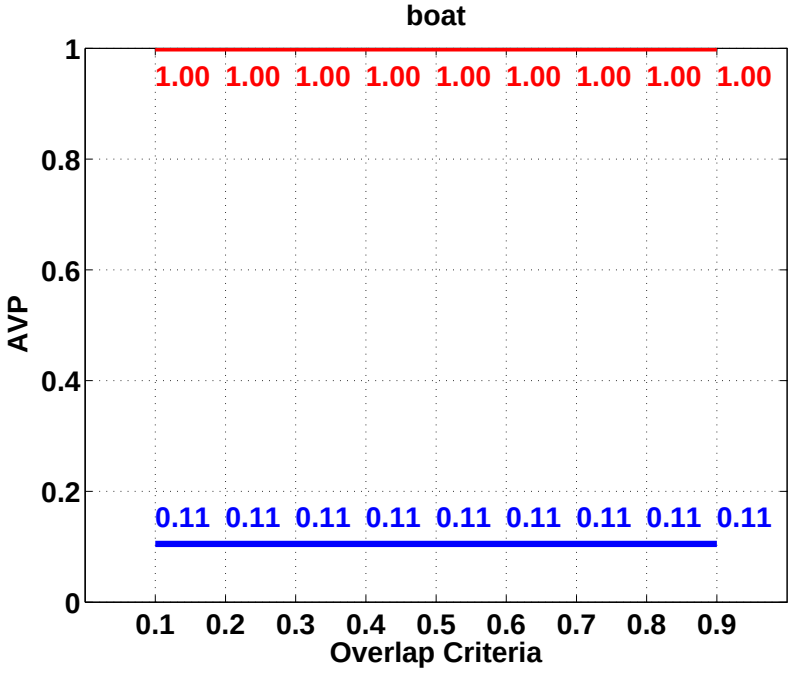


Figure 40: **Simultaneous Object Detection and Pose Estimation.** The detection performance (AP) is represented in red and the pose estimation performance (AVP) in blue.

3.4 bus

Statistics for bus class:

Number of objects = 319, No difficult objects = 301
Truncated: None = 168, Truncated = 133
Occluded: None = 229, Occluded = 90
Occluded and Truncated: 59
side visible:

```
bottom: Yes = 0  No = 319
front: Yes = 67  No = 252
rear: Yes = 18  No = 301
side: Yes = 70  No = 249
top: Yes = 0  No = 319
```

part visible:

```
part 1 = body_back_left_lower: Yes = 107  No = 212
part 2 = body_back_left_upper: Yes = 115  No = 204
part 3 = body_back_right_lower: Yes = 110  No = 209
part 4 = body_back_right_upper: Yes = 120  No = 199
part 5 = body_front_left_upper: Yes = 216  No = 103
part 6 = body_front_right_upper: Yes = 224  No = 95
part 7 = body_front_left_lower: Yes = 199  No = 120
part 8 = body_front_right_lower: Yes = 202  No = 117
part 9 = left_back_wheel: Yes = 72  No = 247
part 10 = left_front_wheel: Yes = 56  No = 263
part 11 = right_back_wheel: Yes = 83  No = 236
part 12 = right_front_wheel: Yes = 83  No = 236
```

Figure 41 summarizes the viewpoint distribution on the test set for bus class considering the 8 views described in Section 1. Figure 42 shows the analysis and the impact on the pose performance of each type of pose error. We also report in Figure 43 the success rate for bus class considering each of the 8 viewpoints described in Section 1.

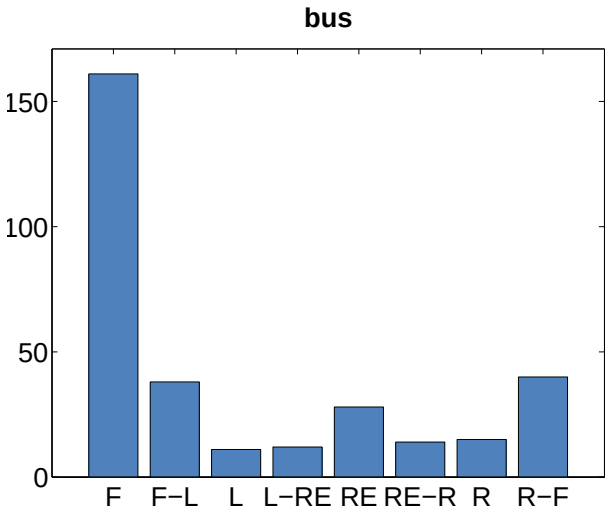
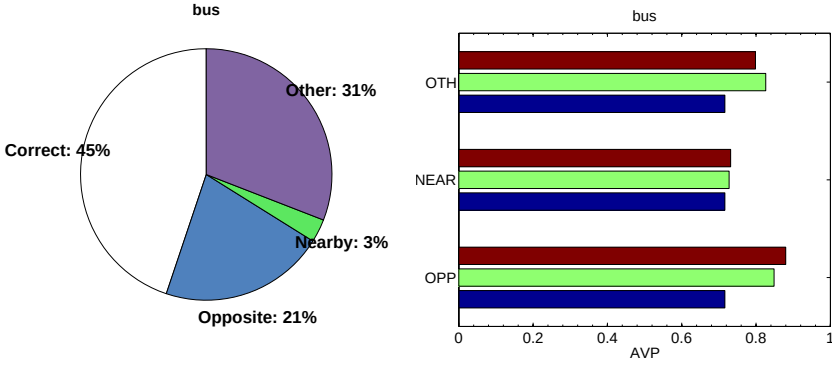


Figure 41: **Analysis of Pose Distribution.** Histogram represents viewpoint distribution on test set for bus class from PASCAL3D+ dataset.



(a) (b)

Figure 42: **Analysis of the Main Pose Estimation Errors.** (a) Given the correct detections, pie chart shows the fraction of pose error that are due to Opposite viewpoints (Opposite), confusion with Nearby viewpoints (Nearby), confusion with Other rotations (Other), and correct pose estimations (Correct) for bus class. (b) Impact of pose errors in terms of AVP. Blue Bars show the AVP performance obtained when all estimations are considered. Green Bars display the AVP improvement by removing all estimations of one type: **OTH** removes confusion with other rotation viewpoints; **NEAR** removes confusion with nearby viewpoints; **OPP** removes confusion with opposite viewpoints. Brick Red Bars display the AVP improvement by correcting all estimations of one type: **OTH** corrects confusion with other rotation viewpoints; **NEAR** corrects confusion with nearby viewpoints; **OPP** corrects confusion with opposite viewpoints.

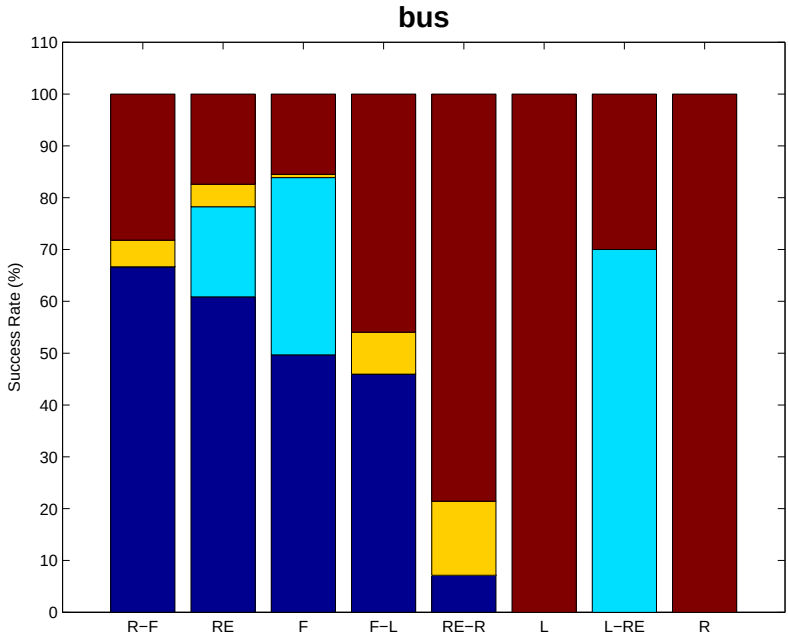
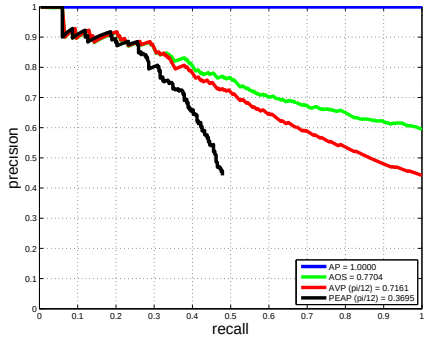
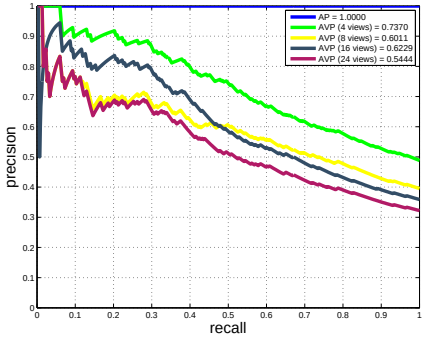


Figure 43: **Success rate for bus class considering 8 views.** For each of the 8 viewpoints, we report: In Blue, the percentage of the correct pose estimations. In Cyan, the percentage of confusion with the opposite viewpoints. In Yellow, the percentage of confusion with nearby viewpoints. In Brick Red, the percentage of the confusions with other rotations.

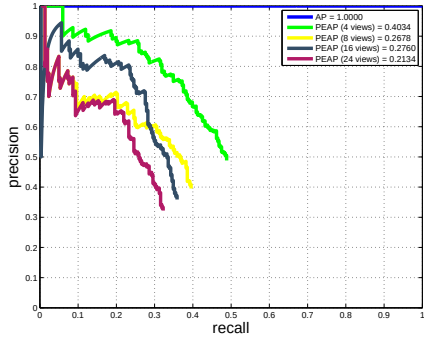
Figure 44 summarizes the detection and pose estimation performances for bus class. Figure 44a shows the model performance working on continuous pose estimation. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$. Figures 44b and 44c report the AVP and PEAP performances achieve working on discrete pose estimation. The AVP and PEAP metrics are obtained considering different number of the views: 4, 8, 16 and 24.



(a)



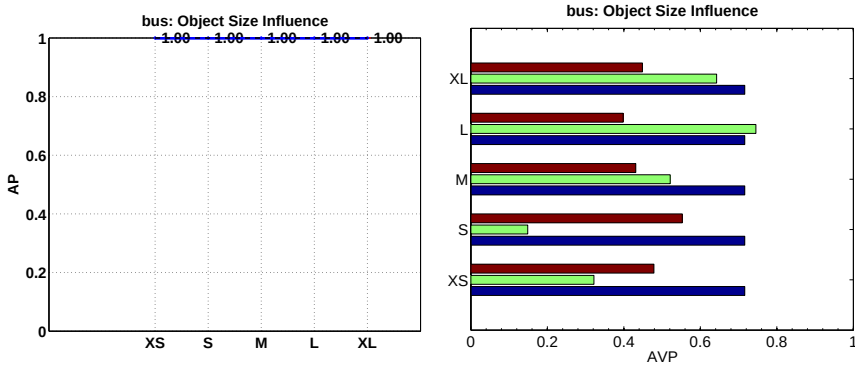
(b)



(c)

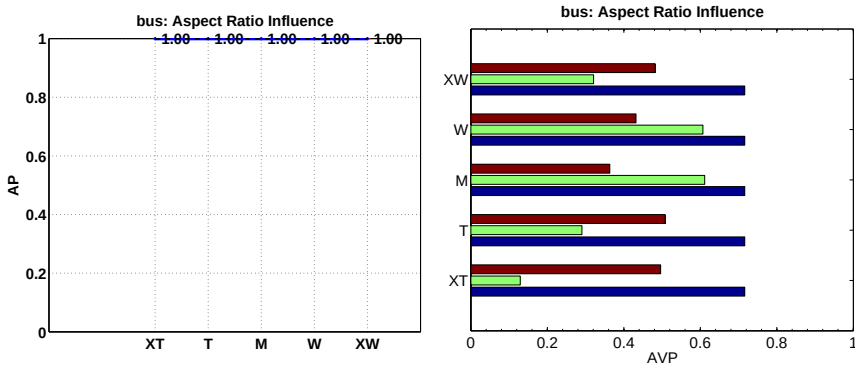
Figure 44: **Precision/Recall Curves.** (a) Performance working on continuous pose estimation. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$. (b) AVP and (c) PEAP performances working on discrete pose estimation. Results obtained by considering 4, 8, 16, and 24 views.

Figures 45, 46, 47, 48 and 49 summarize the main object characteristic influences on detection and pose estimation performances for bus class. Figure 45 reports the effect of object size characteristic on object detection and pose estimation performances. Figure 46 shows the influence of aspect ratio characteristic on object detection and pose estimation performances. Figure 47 represents the influence of occluded and truncated objects on detection and pose estimation performances. Figure 48 shows the visible side influence. Figure 49 reports the part visibily effect.



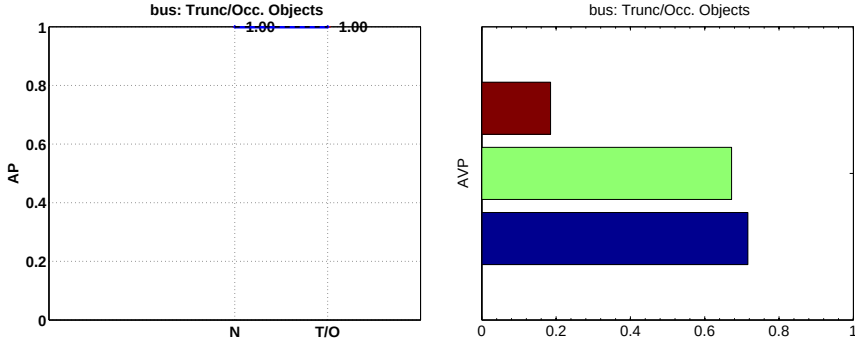
(a) (b)

Figure 45: **Effect of Object Size Characteristic on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP variance. XS and S are considered EXTRA SMALL and SMALL objects, M are MEDIUM objects, L and XL are LARGE and EXTRA LARGE objects. (b) Pose Estimation Results in terms of AVP. Blue Bars show the results when all pose estimations are considered, Green Bars show the results when only the pose estimations corresponding this object size are considered and Brick Red Bars display the results when the pose estimations corresponding this object size are removed.



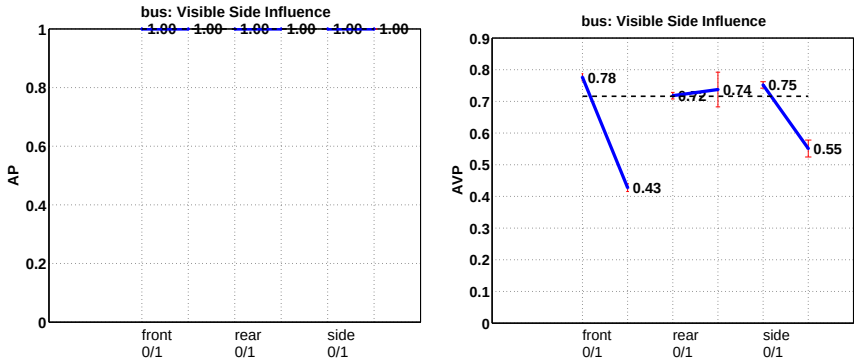
(a) (b)

Figure 46: **Effect of Object Size Characteristic on Object Detection and Pose Estimation.** (a) Detection analysis: AP variance. XS and S are considered EXTRA SMALL and SMALL objects, M are MEDIUM objects, L and XL are LARGE and EXTRA LARGE objects. (b) Pose Estimation Results in terms of AVP. Blue Bars show the results when all pose estimations are considered, Green Bars show the results when only the pose estimations corresponding this object size are considered and Brick Red Bars display the results when the pose estimations corresponding this object size are removed.



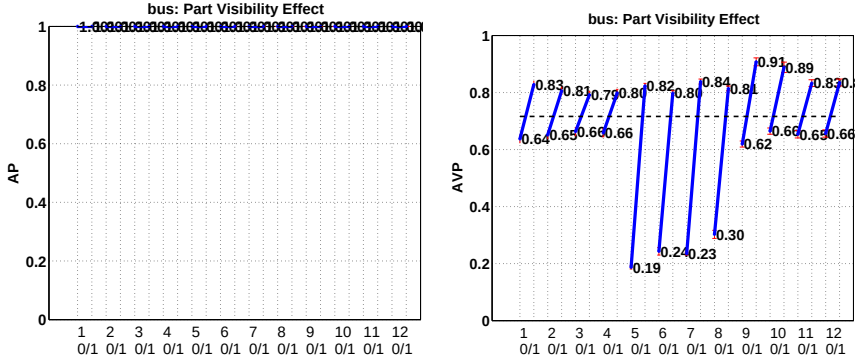
(a) (b)

Figure 47: **Effect of Occluded and Truncated Objects on Detection and Pose Estimation Performances.** (a) Detection analysis: AP variance. N non occluded or truncated object, T/O truncated or occluded object. Results on Pose Estimation in terms of the AVP metric. Blue Bars show the results when all pose estimations are considered. Green Bars display the results when the occluded and truncated objects are removed. Brick Red Bars show the performance when we are only considered the occluded and truncated objects.



(a) (b)

Figure 48: **Effect of Visible Sides on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP performance. Pose Estimation analysis: (b) Pose Estimation analysis: AVP performance. Black dashed lines indicate overall. Visible side: **1** = performance when the side is visible; **0** = performance when the side is not visible.



(a) (b)

Figure 49: **Part Visibility Influence on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP performance. Pose Estimation analysis: (b) Pose Estimation analysis: AVP performance. Black dashed lines indicate overall performance. Visible part: **1** = performance when the part is visible; **0** = performance when the part is not visible. The correspondence between the part number and the part name is indicated at the beginning of this Section (see Statistics for bus class).

Figure 50 provides a summary of the sensitivity to each characteristic and the potential impact on improving pose estimation robustness. The worst-performing and best-performing combinations for each object characteristic are averaged over the selected object categories. The difference between the best and the worst performance indicates sensitivity; the difference between the best and the overall indicates the potential impact.

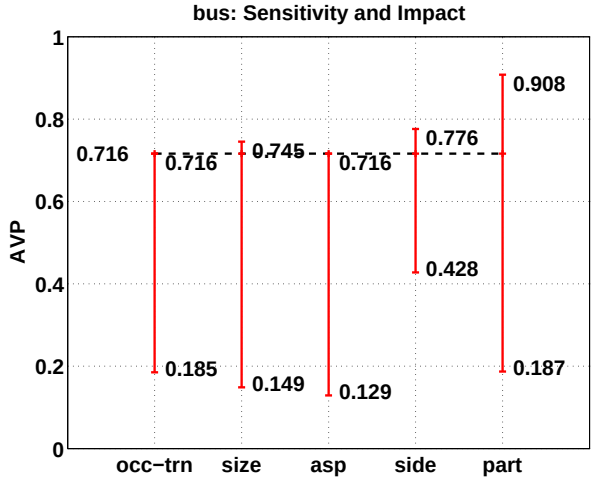


Figure 50: **Summary of Sensitivity and Impact of Object Characteristics.** We show the performance of the highest performing and lowest performing subsets within each characteristic (occlusion/truncation (occ-trn), bounding box area or object size (size), aspect ratio (asp), visible sides (side) and part visibility (part)). Overall accuracy is indicated by the black dashed line. The difference between max and min indicates sensitivity; the difference between max and overall indicates the impact.

Figure 51 shows an analysis of the influence of the overlap criterion considering the AVP metric. For this overlap criterion analysis we follow the PASCAL VOC formulation: to be considered a true positive, the area of overlap between the predicted BB and GT BB must exceed a threshold.

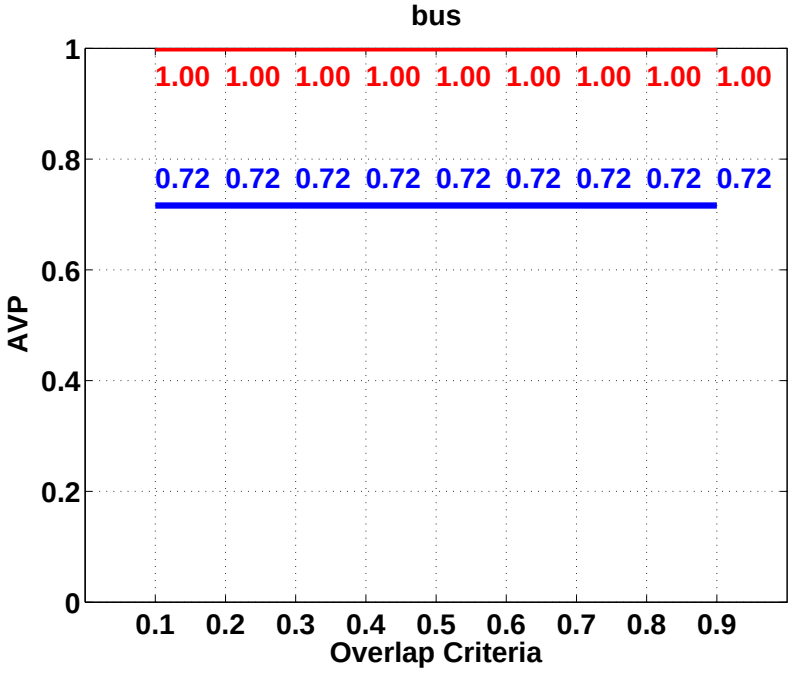


Figure 51: **Simultaneous Object Detection and Pose Estimation.** The detection performance (AP) is represented in red and the pose estimation performance (AVP) in blue.

3.5 car

Statistics for car class:

Number of objects = 1163, No difficult objects = 1004
Truncated: None = 347, Truncated = 657
Occluded: None = 645, Occluded = 518
Occluded and Truncated: 352
side visible:

```
bottom: Yes = 0  No = 1163
front: Yes = 135  No = 1028
rear: Yes = 173  No = 990
side: Yes = 441  No = 722
top: Yes = 0  No = 1163
```

part visible:

```
part 1 = left_front_wheel: Yes = 237  No = 926
part 2 = left_back_wheel: Yes = 234  No = 929
part 3 = right_front_wheel: Yes = 230  No = 933
part 4 = right_back_wheel: Yes = 236  No = 927
part 5 = upper_left_windshield: Yes = 409  No = 754
part 6 = upper_right_windshield: Yes = 400  No = 763
part 7 = upper_left_rearwindow: Yes = 376  No = 787
part 8 = upper_right_rearwindow: Yes = 374  No = 789
part 9 = left_front_light: Yes = 336  No = 827
part 10 = right_front_light: Yes = 313  No = 850
part 11 = left_back_trunk: Yes = 382  No = 781
part 12 = right_back_trunk: Yes = 377  No = 786
```

Figure 52 summarizes the viewpoint distribution on the test set for car class considering the 8 views described in Section 1. Figure 53 shows the analysis and the impact on the pose performance of each type of pose error. We also report in Figure 54 the success rate for car class considering each of the 8 viewpoints described in Section 1.

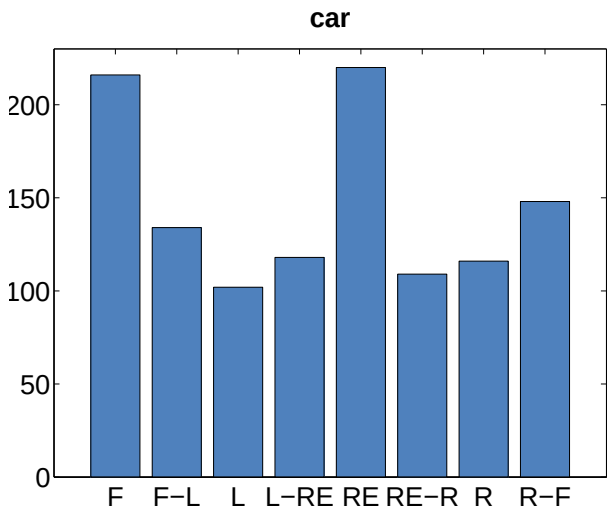
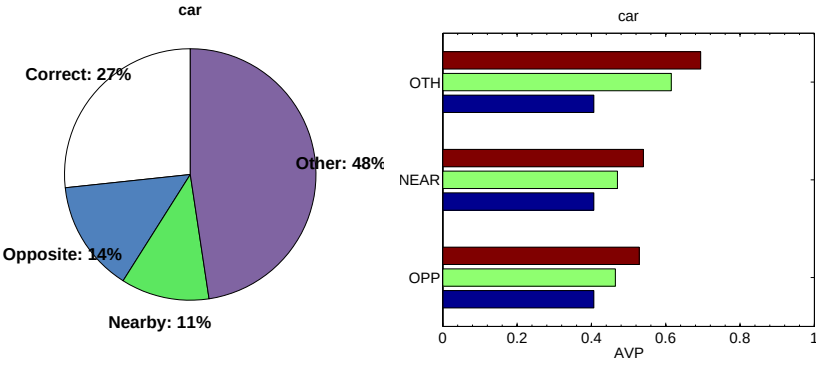


Figure 52: **Analysis of Pose Distribution.** Histogram represents viewpoint distribution on test set for car class from PASCAL3D+ dataset.



(a) (b)

Figure 53: **Analysis of the Main Pose Estimation Errors.** (a) Given the correct detections, pie chart shows the fraction of pose error that are due to Opposite viewpoints (Opposite), confusion with Nearby viewpoints (Nearby), confusion with Other rotations (Other), and correct pose estimations (Correct) for car class. (b) Impact of pose errors in terms of AVP. Blue Bars show the AVP performance obtained when all estimations are considered. Green Bars display the AVP improvement by removing all estimations of one type: **OTH** removes confusion with other rotation viewpoints; **NEAR** removes confusion with nearby viewpoints; **OPP** removes confusion with opposite viewpoints. Brick Red Bars display the AVP improvement by correcting all estimations of one type: **OTH** corrects confusion with other rotation viewpoints; **NEAR** corrects confusion with nearby viewpoints; **OPP** corrects confusion with opposite viewpoints.

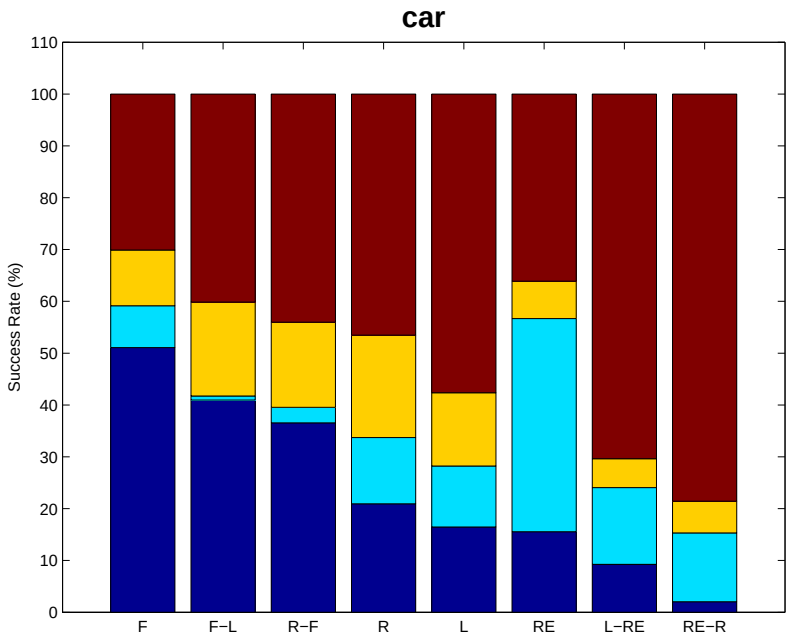
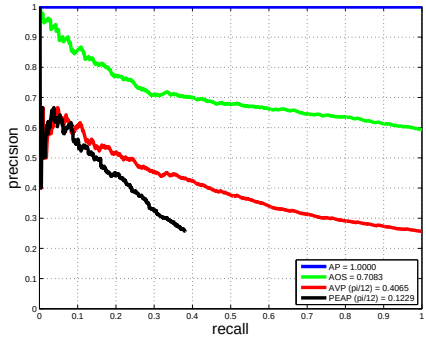
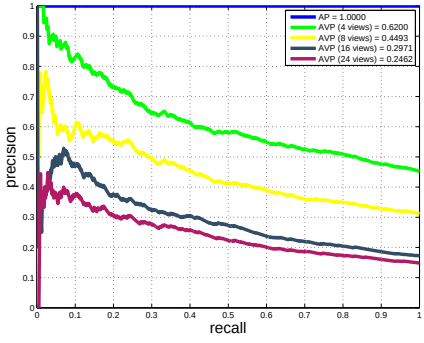


Figure 54: **Success rate for car class considering 8 views.** For each of the 8 viewpoints, we report: In Blue, the percentage of the correct pose estimations. In Cyan, the percentage of confusion with the opposite viewpoints. In Yellow, the percentage of confusion with nearby viewpoints. In Brick Red, the percentage of the confusions with other rotations.

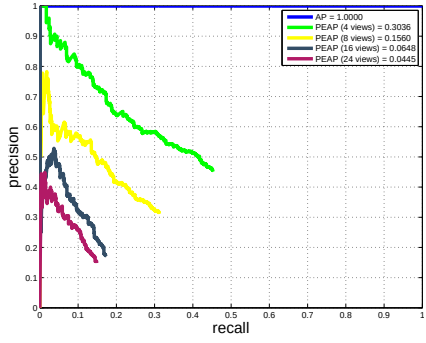
Figure 55 summarizes the detection and pose estimation performances for car class. Figure 55a shows the model performance working on continuous pose estimation. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$. Figures 55b and 55c report the AVP and PEAP performances achieve working on discrete pose estimation. The AVP and PEAP metrics are obtained considering different number of the views: 4, 8, 16 and 24.



(a)



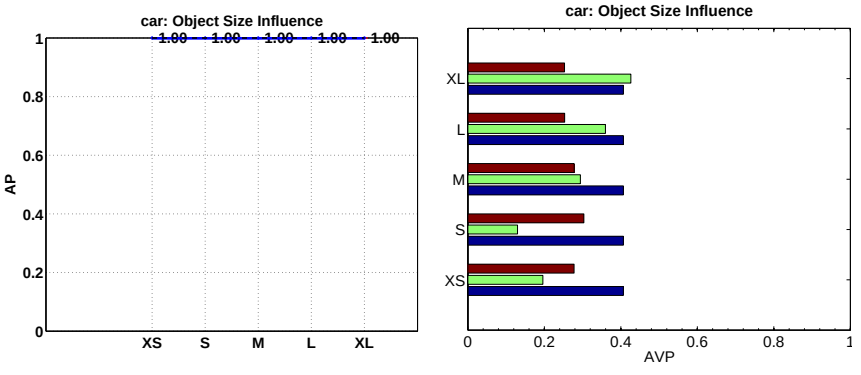
(b)



(c)

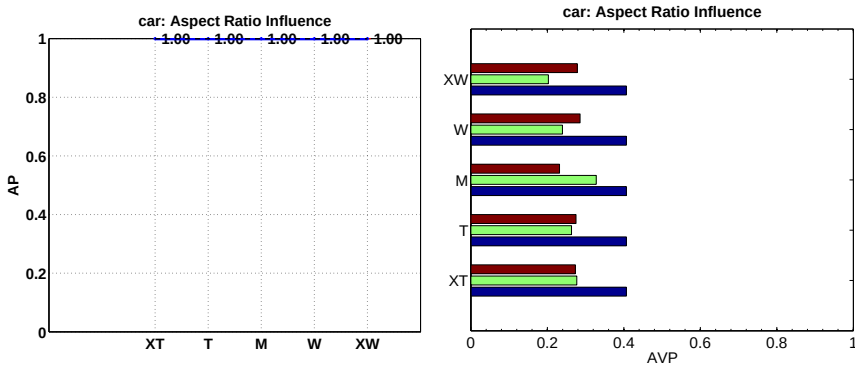
Figure 55: **Precision/Recall Curves.** (a) Performance working on continuous pose estimation. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$. (b) AVP and (c) PEAP performances working on discrete pose estimation. Results obtained by considering 4, 8, 16, and 24 views.

Figures 56, 57, 58, 59 and 60 summarize the main object characteristic influences on detection and pose estimation performances for car class. Figure 56 reports the effect of object size characteristic on object detection and pose estimation performances. Figure 57 shows the influence of aspect ratio characteristic on object detection and pose estimation performances. Figure 58 represents the influence of occluded and truncated objects on detection and pose estimation performances. Figure 59 shows the visible side influence. Figure 60 reports the part visibily effect.



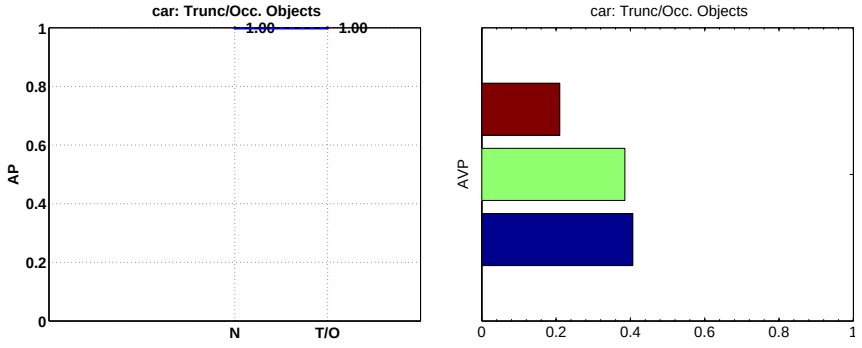
(a) (b)

Figure 56: **Effect of Object Size Characteristic on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP variance. XS and S are considered EXTRA SMALL and SMALL objects, M are MEDIUM objects, L and XL are LARGE and EXTRA LARGE objects. (b) Pose Estimation Results in terms of AVP. Blue Bars show the results when all pose estimations are considered, Green Bars show the results when only the pose estimations corresponding this object size are considered and Brick Red Bars display the results when the pose estimations corresponding this object size are removed.



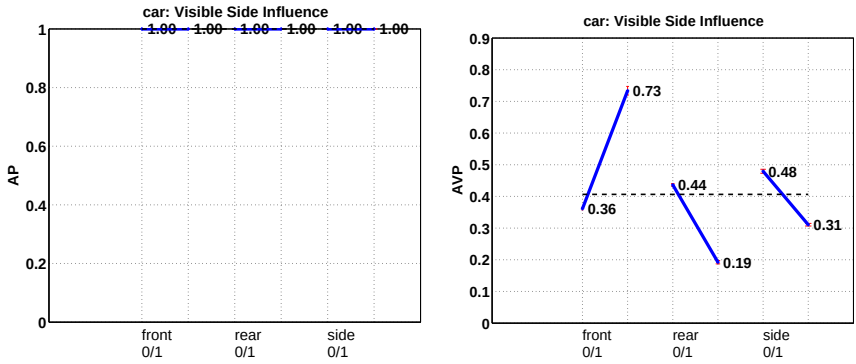
(a) (b)

Figure 57: **Effect of Object Size Characteristic on Object Detection and Pose Estimation.** (a) Detection analysis: AP variance. XS and S are considered EXTRA SMALL and SMALL objects, M are MEDIUM objects, L and XL are LARGE and EXTRA LARGE objects. (b) Pose Estimation Results in terms of AVP. Blue Bars show the results when all pose estimations are considered, Green Bars show the results when only the pose estimations corresponding this object size are considered and Brick Red Bars display the results when the pose estimations corresponding this object size are removed.



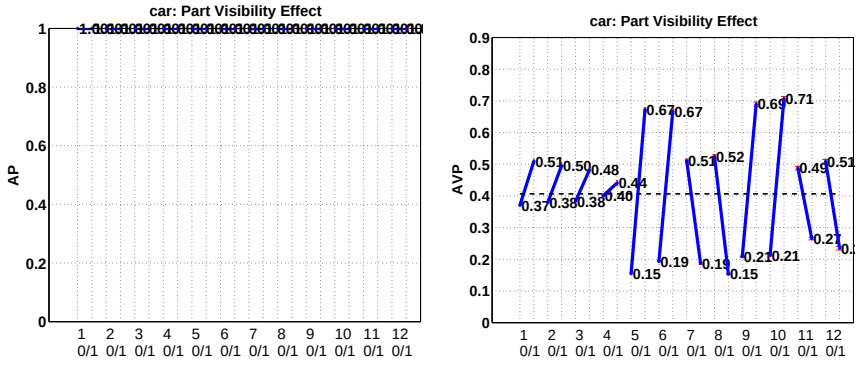
(a) (b)

Figure 58: **Effect of Occluded and Truncated Objects on Detection and Pose Estimation Performances.** (a) Detection analysis: AP variance. N non occluded or truncated object, T/O truncated or occluded object. Results on Pose Estimation in terms of the AVP metric. Blue Bars show the results when all pose estimations are considered. Green Bars display the results when the occluded and truncated objects are removed. Brick Red Bars show the performance when we are only considered the occluded and truncated objects.



(a) (b)

Figure 59: **Effect of Visible Sides on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP performance. Pose Estimation analysis: (b) Pose Estimation analysis: AVP performance. Black dashed lines indicate overall. Visible side: **1** = performance when the side is visible; **0** = performance when the side is not visible.



(a) (b)

Figure 60: **Part Visibility Influence on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP performance. Pose Estimation analysis: (b) Pose Estimation analysis: AVP performance. Black dashed lines indicate overall performance. Visible part: **1** = performance when the part is visible; **0** = performance when the part is not visible. The correspondence between the part number and the part name is indicated at the beginning of this Section (see Statistics for car class).

Figure 61 provides a summary of the sensitivity to each characteristic and the potential impact on improving pose estimation robustness. The worst-performing and best-performing combinations for each object characteristic are averaged over the selected object categories. The difference between the best and the worst performance indicates sensitivity; the difference between the best and the overall indicates the potential impact.

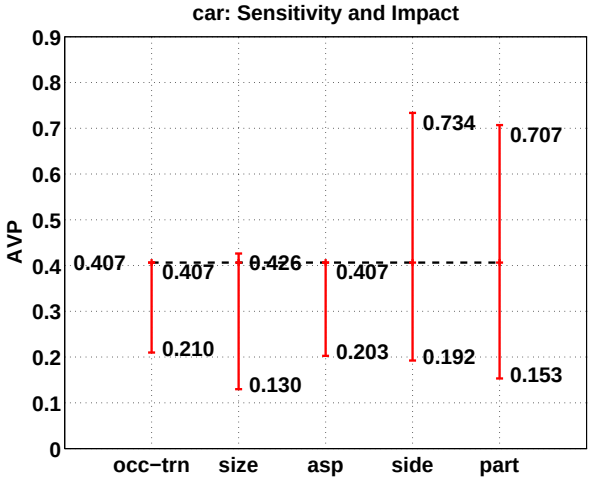


Figure 61: **Summary of Sensitivity and Impact of Object Characteristics.** We show the performance of the highest performing and lowest performing subsets within each characteristic (occlusion/truncation (occ-trn), bounding box area or object size (size), aspect ratio (asp), visible sides (side) and part visibility (part)). Overall accuracy is indicated by the black dashed line. The difference between max and min indicates sensitivity; the difference between max and overall indicates the impact.

Figure 62 shows an analysis of the influence of the overlap criterion considering the AVP metric. For this overlap criterion analysis we follow the PASCAL VOC formulation: to be considered a true positive, the area of overlap between the predicted BB and GT BB must exceed a threshold.

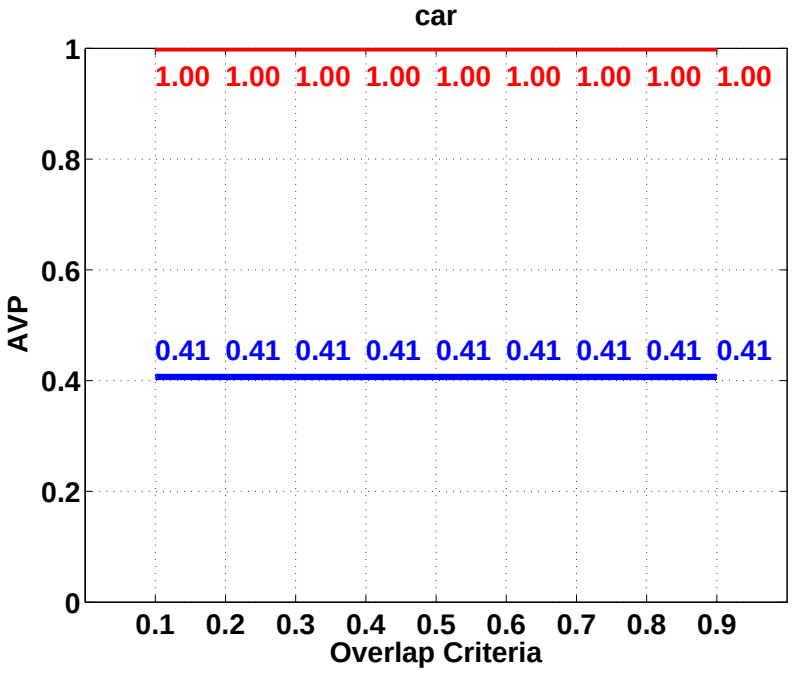


Figure 62: **Simultaneous Object Detection and Pose Estimation.** The detection performance (AP) is represented in red and the pose estimation performance (AVP) in blue.

3.6 chair

Statistics for chair class:

Number of objects = 1395, No difficult objects = 1176
Truncated: None = 271, Truncated = 905
Occluded: None = 620, Occluded = 775
Occluded and Truncated: 538
side visible:

```
bottom: Yes = 0  No = 1395
front: Yes = 287  No = 1108
rear: Yes = 172  No = 1223
side: Yes = 380  No = 1015
top: Yes = 0  No = 1395
```

part visible:

```
part 1 = back_upper_left: Yes = 1043  No = 352
part 2 = back_upper_right: Yes = 1044  No = 351
part 3 = seat_upper_left: Yes = 453  No = 942
part 4 = seat_upper_right: Yes = 446  No = 949
part 5 = seat_lower_left: Yes = 477  No = 918
part 6 = seat_lower_right: Yes = 492  No = 903
part 7 = leg_upper_left: Yes = 335  No = 1060
part 8 = leg_upper_right: Yes = 360  No = 1035
part 9 = leg_lower_left: Yes = 328  No = 1067
part 10 = leg_lower_right: Yes = 356  No = 1039
```

Figure 63 summarizes the viewpoint distribution on the test set for chair class considering the 8 views described in Section 1. Figure 64 shows the analysis and the impact on the pose performance of each type of pose error. We also report in Figure 65 the success rate for chair class considering each of the 8 viewpoints described in Section 1.

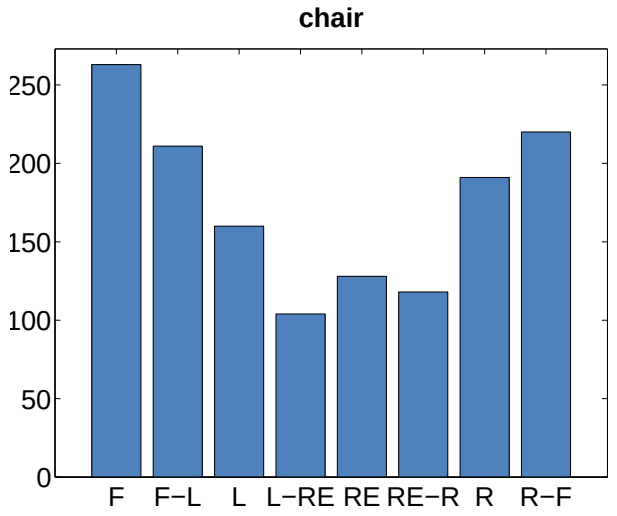
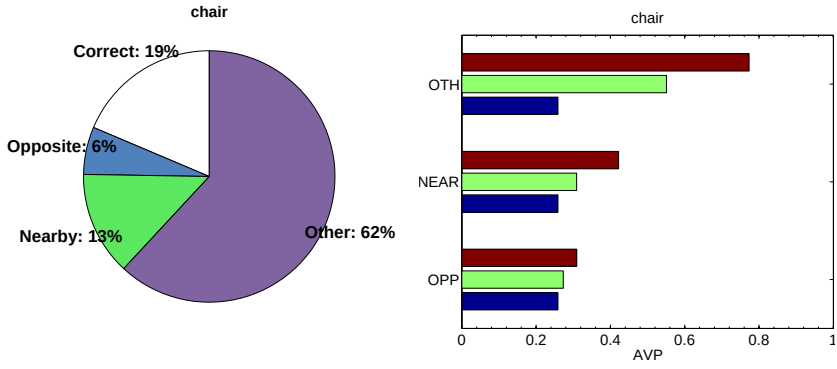


Figure 63: **Analysis of Pose Distribution.** Histogram represents viewpoint distribution on test set for chair class from PASCAL3D+ dataset.



(a) (b)

Figure 64: **Analysis of the Main Pose Estimation Errors.** (a) Given the correct detections, pie chart shows the fraction of pose error that are due to Opposite viewpoints (Opposite), confusion with Nearby viewpoints (Nearby), confusion with Other rotations (Other), and correct pose estimations (Correct) for chair class. (b) Impact of pose errors in terms of AVP. Blue Bars show the AVP performance obtained when all estimations are considered. Green Bars display the AVP improvement by removing all estimations of one type: **OTH** removes confusion with other rotation viewpoints; **NEAR** removes confusion with nearby viewpoints; **OPP** removes confusion with opposite viewpoints. Brick Red Bars display the AVP improvement by correcting all estimations of one type: **OTH** corrects confusion with other rotation viewpoints; **NEAR** corrects confusion with nearby viewpoints; **OPP** corrects confusion with opposite viewpoints.

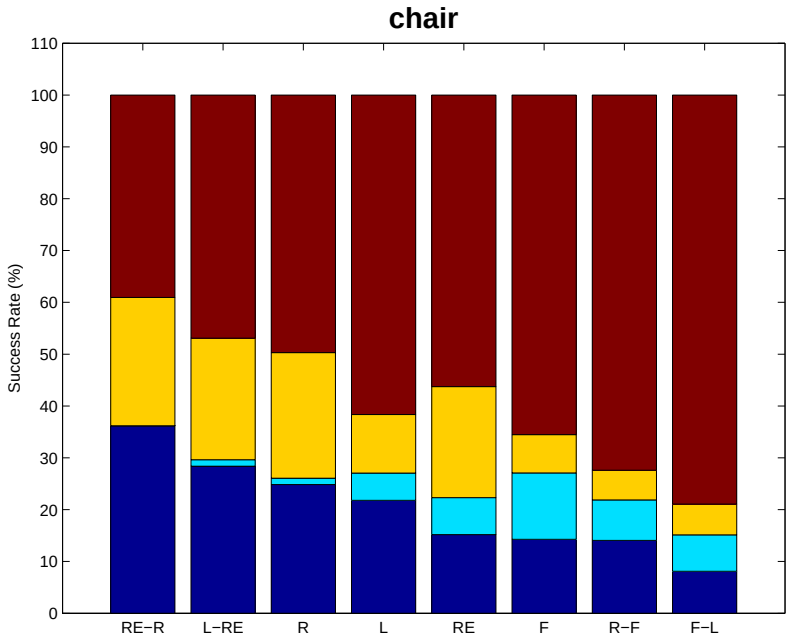
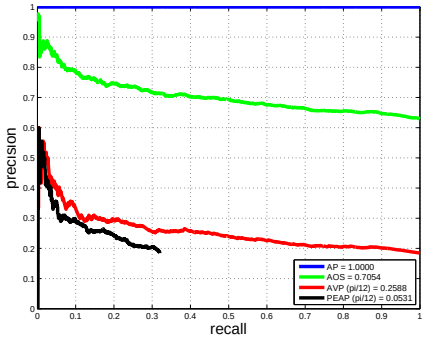
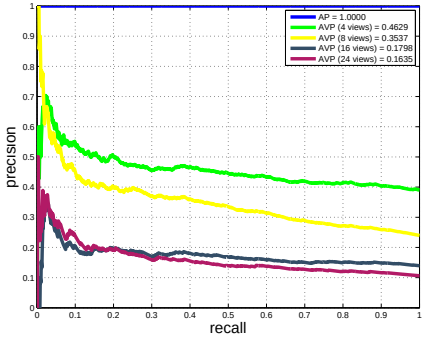


Figure 65: **Success rate for chair class considering 8 views.** For each of the 8 viewpoints, we report: In Blue, the percentage of the correct pose estimations. In Cyan, the percentage of confusion with the opposite viewpoints. In Yellow, the percentage of confusion with nearby viewpoints. In Brick Red, the percentage of the confusions with other rotations.

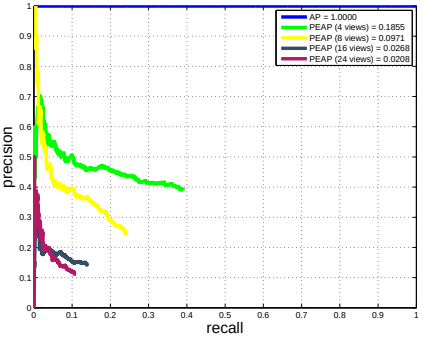
Figure 66 summarizes the detection and pose estimation performances for chair class. Figure 66a shows the model performance working on continuous pose estimation. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$. Figures 66b and 66c report the AVP and PEAP performances achieve working on discrete pose estimation. The AVP and PEAP metrics are obtained considering different number of the views: 4, 8, 16 and 24.



(a)



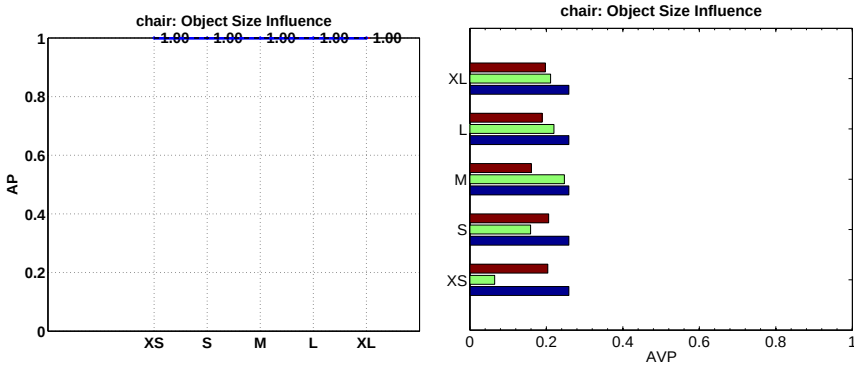
(b)



(c)

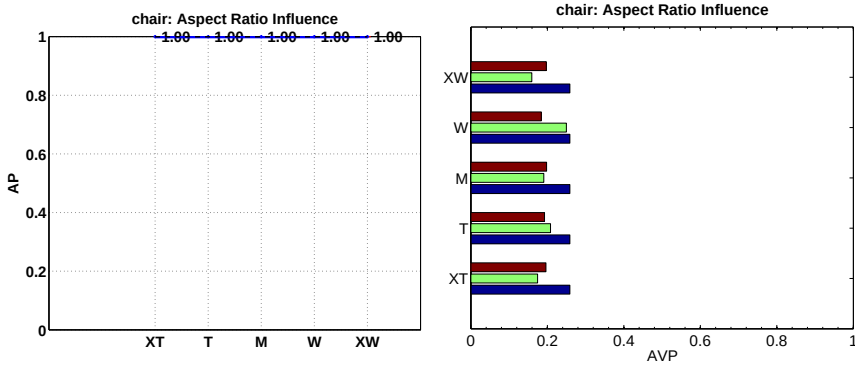
Figure 66: **Precision/Recall Curves.** (a) Performance working on continuous pose estimation. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$. (b) AVP and (c) PEAP performances working on discrete pose estimation. Results obtained by considering 4, 8, 16, and 24 views.

Figures 67, 68, 69, 70 and 71 summarize the main object characteristic influences on detection and pose estimation performances for chair class. Figure 67 reports the effect of object size characteristic on object detection and pose estimation performances. Figure 68 shows the influence of aspect ratio characteristic on object detection and pose estimation performances. Figure 69 represents the influence of occluded and truncated objects on detection and pose estimation performances. Figure 70 shows the visible side influence. Figure 71 reports the part visibily effect.



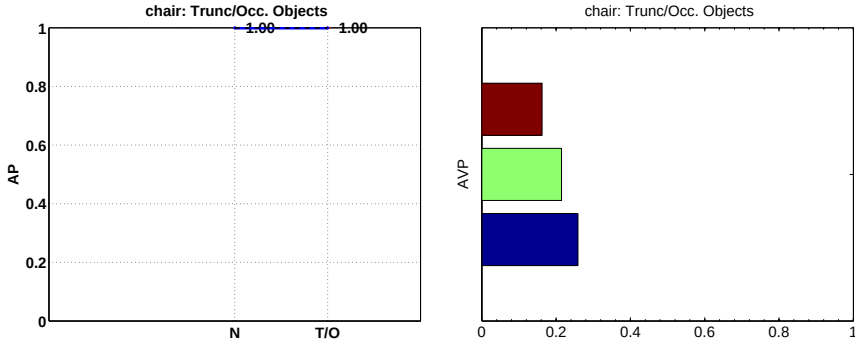
(a) (b)

Figure 67: **Effect of Object Size Characteristic on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP variance. XS and S are considered EXTRA SMALL and SMALL objects, M are MEDIUM objects, L and XL are LARGE and EXTRA LARGE objects. (b) Pose Estimation Results in terms of AVP. Blue Bars show the results when all pose estimations are considered, Green Bars show the results when only the pose estimations corresponding this object size are considered and Brick Red Bars display the results when the pose estimations corresponding this object size are removed.



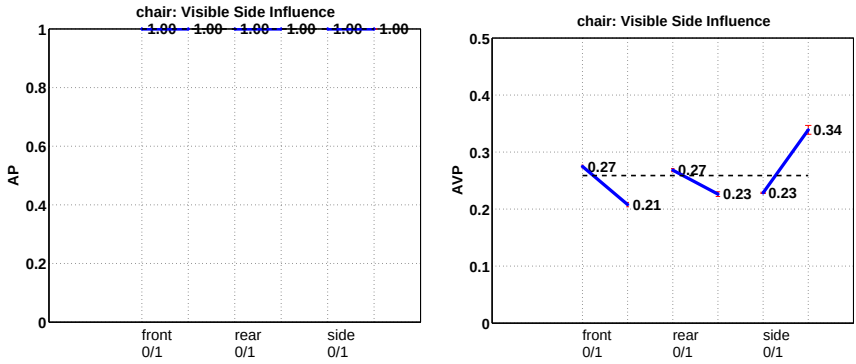
(a) (b)

Figure 68: **Effect of Object Size Characteristic on Object Detection and Pose Estimation.** (a) Detection analysis: AP variance. XS and S are considered EXTRA SMALL and SMALL objects, M are MEDIUM objects, L and XL are LARGE and EXTRA LARGE objects. (b) Pose Estimation Results in terms of AVP. Blue Bars show the results when all pose estimations are considered, Green Bars show the results when only the pose estimations corresponding this object size are considered and Brick Red Bars display the results when the pose estimations corresponding this object size are removed.



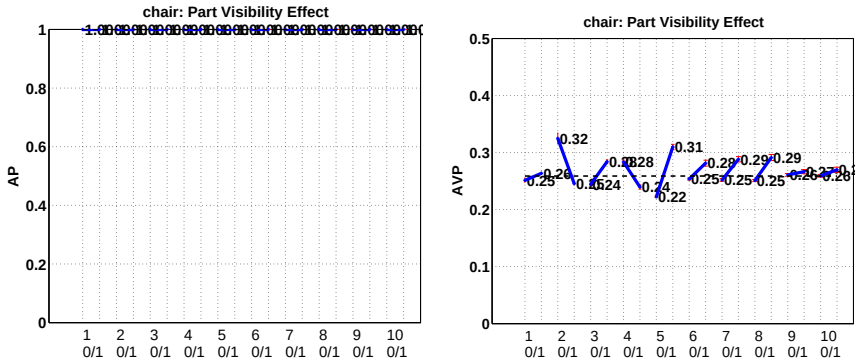
(a) (b)

Figure 69: **Effect of Occluded and Truncated Objects on Detection and Pose Estimation Performances.** (a) Detection analysis: AP variance. N non occluded or truncated object, T/O truncated or occluded object. Results on Pose Estimation in terms of the AVP metric. Blue Bars show the results when all pose estimations are considered. Green Bars display the results when the occluded and truncated objects are removed. Brick Red Bars show the performance when we are only considered the occluded and truncated objects.



(a) (b)

Figure 70: **Effect of Visible Sides on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP performance. Pose Estimation analysis: (b) Pose Estimation analysis: AVP performance. Black dashed lines indicate overall. Visible side: **1** = performance when the side is visible; **0** = performance when the side is not visible.



(a) (b)

Figure 71: **Part Visibility Influence on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP performance. Pose Estimation analysis: (b) Pose Estimation analysis: AVP performance. Black dashed lines indicate overall performance. Visible part: **1** = performance when the part is visible; **0** = performance when the part is not visible. The correspondence between the part number and the part name is indicated at the beginning of this Section (see Statistics for chair class).

Figure 72 provides a summary of the sensitivity to each characteristic and the potential impact on improving pose estimation robustness. The worst-performing and best-performing combinations for each object characteristic are averaged over the selected object categories. The difference between the best and the worst performance indicates sensitivity; the difference between the best and the overall indicates the potential impact.

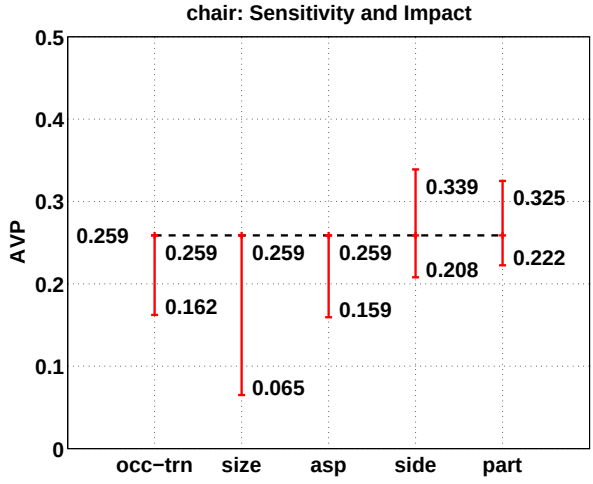


Figure 72: **Summary of Sensitivity and Impact of Object Characteristics.** We show the performance of the highest performing and lowest performing subsets within each characteristic (occlusion/truncation (occ-trn), bounding box area or object size (size), aspect ratio (asp), visible sides (side) and part visibility (part)). Overall accuracy is indicated by the black dashed line. The difference between max and min indicates sensitivity; the difference between max and overall indicates the impact.

Figure 73 shows an analysis of the influence of the overlap criterion considering the AVP metric. For this overlap criterion analysis we follow the PASCAL VOC formulation: to be considered a true positive, the area of overlap between the predicted BB and GT BB must exceed a threshold.

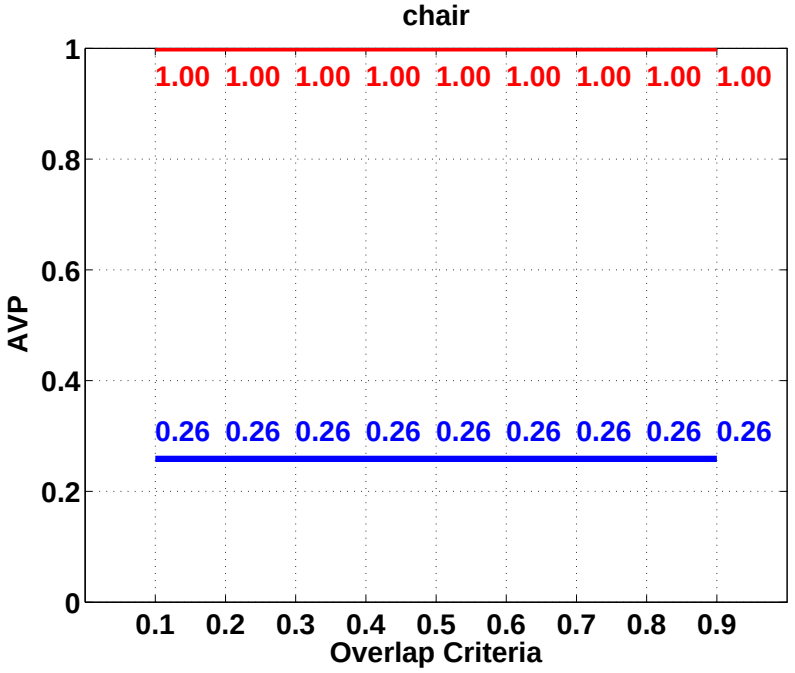


Figure 73: **Simultaneous Object Detection and Pose Estimation.** The detection performance (AP) is represented in red and the pose estimation performance (AVP) in blue.

3.7 diningtable

Statistics for diningtable class:

Number of objects = 344, No difficult objects = 305
Truncated: None = 34, Truncated = 271
Occluded: None = 112, Occluded = 232
Occluded and Truncated: 182
side visible:

```
bottom: Yes = 0  No = 344
front: Yes = 2  No = 342
rear: Yes = 0  No = 344
side: Yes = 1  No = 343
top: Yes = 0  No = 344
```

part visible:

```
part 1 = leg_upper_left: Yes = 14  No = 330
part 2 = leg_upper_right: Yes = 14  No = 330
part 3 = leg_lower_left: Yes = 27  No = 317
part 4 = leg_lower_right: Yes = 26  No = 318
part 5 = top_upper_left: Yes = 124  No = 220
part 6 = top_upper_right: Yes = 141  No = 203
part 7 = top_lower_left: Yes = 128  No = 216
part 8 = top_lower_right: Yes = 132  No = 212
part 9 = top_up: Yes = 41  No = 303
part 10 = top_down: Yes = 33  No = 311
part 11 = top_left: Yes = 42  No = 302
part 12 = top_right: Yes = 38  No = 306
```

Figure 74 summarizes the viewpoint distribution on the test set for diningtable class considering the 8 views described in Section 1. Figure 75 shows the analysis and the impact on the pose performance of each type of pose error. We also report in Figure 76 the success rate for diningtable class considering each of the 8 viewpoints described in Section 1.

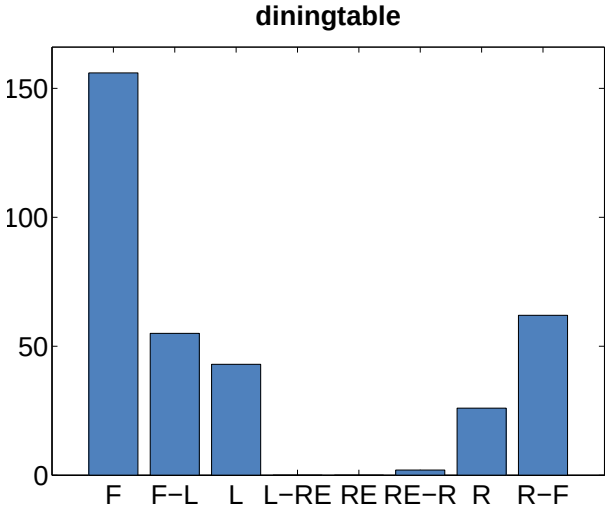
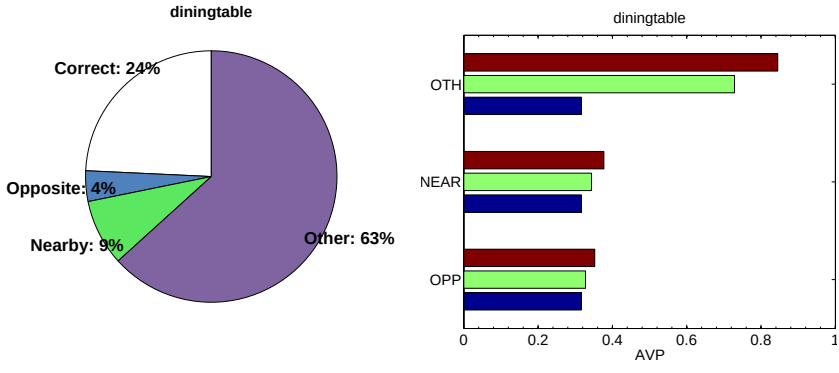


Figure 74: **Analysis of Pose Distribution.** Histogram represents viewpoint distribution on test set for diningtable class from PASCAL3D+ dataset.



(a) (b)

Figure 75: **Analysis of the Main Pose Estimation Errors.** (a) Given the correct detections, pie chart shows the fraction of pose error that are due to Opposite viewpoints (Opposite), confusion with Nearby viewpoints (Nearby), confusion with Other rotations (Other), and correct pose estimations (Correct) for diningtable class. (b) Impact of pose errors in terms of AVP. Blue Bars show the AVP performance obtained when all estimations are considered. Green Bars display the AVP improvement by removing all estimations of one type: **OTH** removes confusion with other rotation viewpoints; **NEAR** removes confusion with nearby viewpoints; **OPP** removes confusion with opposite viewpoints. Brick Red Bars display the AVP improvement by correcting all estimations of one type: **OTH** corrects confusion with other rotation viewpoints; **NEAR** corrects confusion with nearby viewpoints; **OPP** corrects confusion with opposite viewpoints.

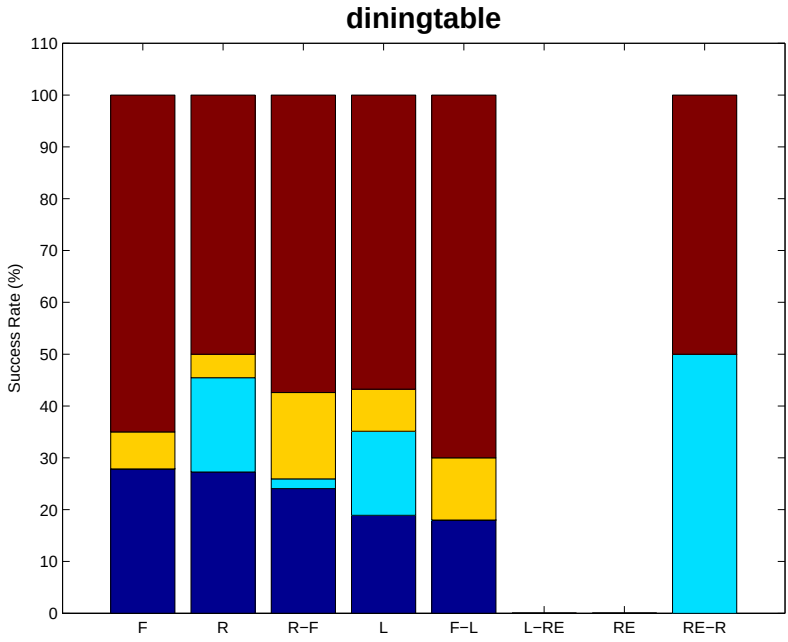
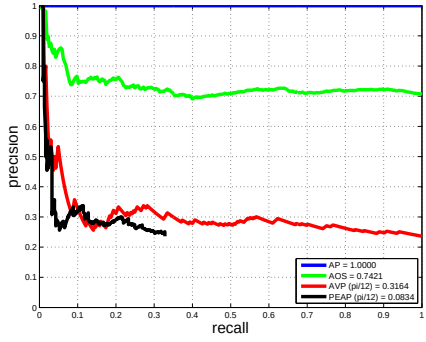
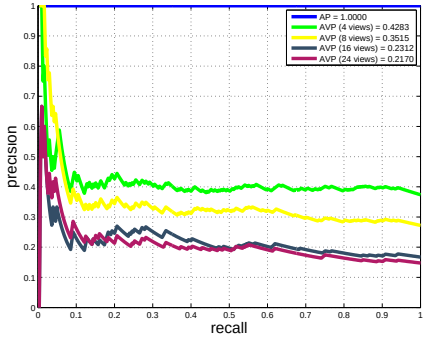


Figure 76: **Success rate for diningtable class considering 8 views.** For each of the 8 viewpoints, we report: In Blue, the percentage of the correct pose estimations. In Cyan, the percentage of confusion with the opposite viewpoints. In Yellow, the percentage of confusion with nearby viewpoints. In Brick Red, the percentage of the confusions with other rotations.

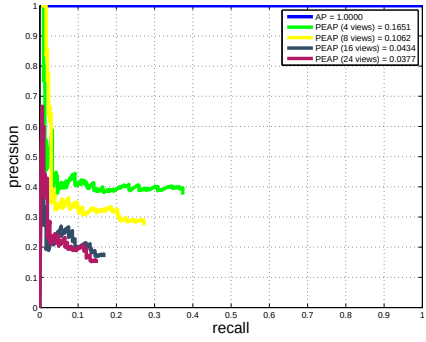
Figure 77 summarizes the detection and pose estimation performances for diningtable class. Figure 77a shows the model performance working on continuous pose estimation. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$. Figures 77b and 77c report the AVP and PEAP performances achieve working on discrete pose estimation. The AVP and PEAP metrics are obtained considering different number of the views: 4, 8, 16 and 24.



(a)



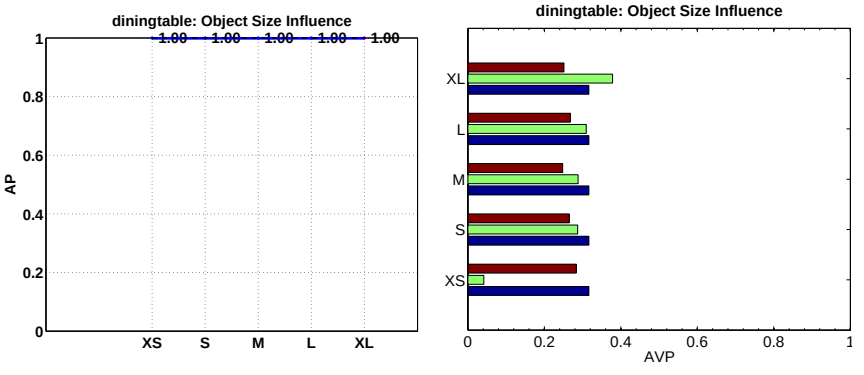
(b)



(c)

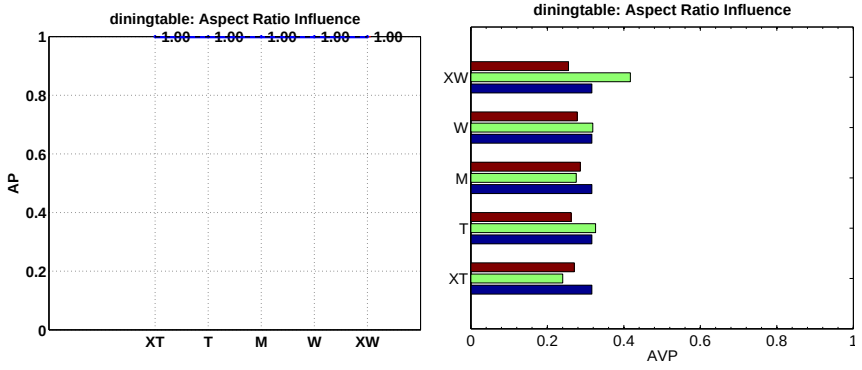
Figure 77: **Precision/Recall Curves.** (a) Performance working on continuous pose estimation. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$. (b) AVP and (c) PEAP performances working on discrete pose estimation. Results obtained by considering 4, 8, 16, and 24 views.

Figures 78, 79, 80, 81 and 82 summarize the main object characteristic influences on detection and pose estimation performances for diningtable class. Figure 78 reports the effect of object size characteristic on object detection and pose estimation performances. Figure 79 shows the influence of aspect ratio characteristic on object detection and pose estimation performances. Figure 80 represents the influence of occluded and truncated objects on detection and pose estimation performances. Figure 81 shows the visible side influence. Figure 82 reports the part visibily effect.



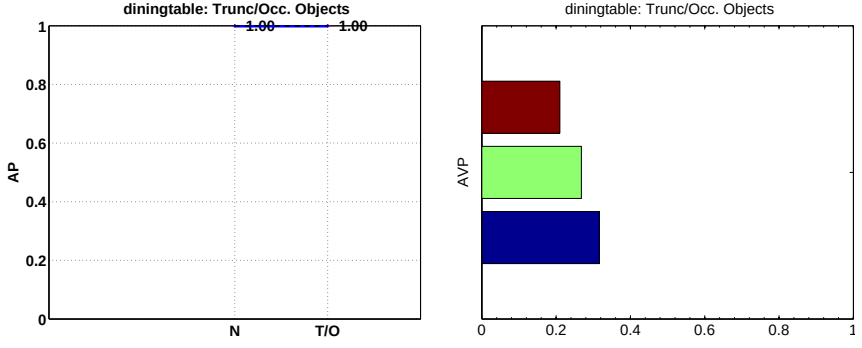
(a) (b)

Figure 78: **Effect of Object Size Characteristic on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP variance. XS and S are considered EXTRA SMALL and SMALL objects, M are MEDIUM objects, L and XL are LARGE and EXTRA LARGE objects. (b) Pose Estimation Results in terms of AVP. Blue Bars show the results when all pose estimations are considered, Green Bars show the results when only the pose estimations corresponding this object size are considered and Brick Red Bars display the results when the pose estimations corresponding this object size are removed.



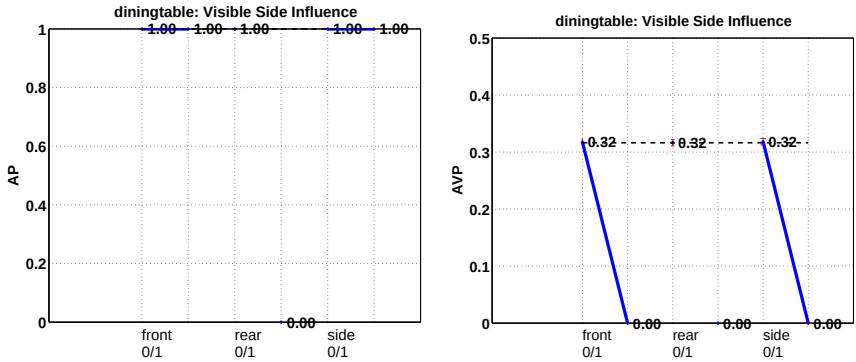
(a) (b)

Figure 79: **Effect of Object Size Characteristic on Object Detection and Pose Estimation.** (a) Detection analysis: AP variance. XS and S are considered EXTRA SMALL and SMALL objects, M are MEDIUM objects, L and XL are LARGE and EXTRA LARGE objects. (b) Pose Estimation Results in terms of AVP. Blue Bars show the results when all pose estimations are considered, Green Bars show the results when only the pose estimations corresponding this object size are considered and Brick Red Bars display the results when the pose estimations corresponding this object size are removed.



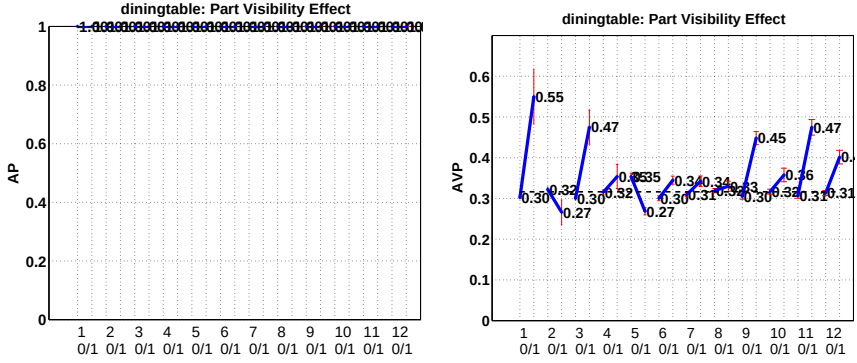
(a) (b)

Figure 80: **Effect of Occluded and Truncated Objects on Detection and Pose Estimation Performances.** (a) Detection analysis: AP variance. N non occluded or truncated object, T/O truncated or occluded object. Results on Pose Estimation in terms of the AVP metric. Blue Bars show the results when all pose estimations are considered. Green Bars display the results when the occluded and truncated objects are removed. Brick Red Bars show the performance when we are only considered the occluded and truncated objects.



(a) (b)

Figure 81: **Effect of Visible Sides on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP performance. Pose Estimation analysis: (b) Pose Estimation analysis: AVP performance. Black dashed lines indicate overall. Visible side: **1** = performance when the side is visible; **0** = performance when the side is not visible.



(a) (b)

Figure 82: **Part Visibility Influence on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP performance. Pose Estimation analysis: (b) Pose Estimation analysis: AVP performance. Black dashed lines indicate overall performance. Visible part: **1** = performance when the part is visible; **0** = performance when the part is not visible. The correspondence between the part number and the part name is indicated at the beginning of this Section (see Statistics for diningtable class).

Figure 83 provides a summary of the sensitivity to each characteristic and the potential impact on improving pose estimation robustness. The worst-performing and best-performing combinations for each object characteristic are averaged over the selected object categories. The difference between the best and the worst performance indicates sensitivity; the difference between the best and the overall indicates the potential impact.

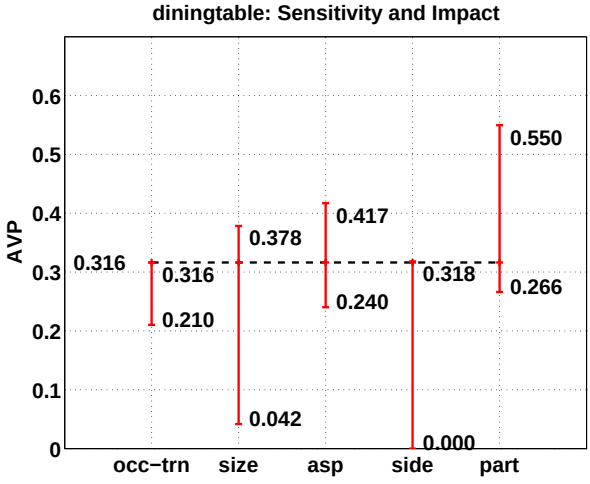


Figure 83: **Summary of Sensitivity and Impact of Object Characteristics.** We show the performance of the highest performing and lowest performing subsets within each characteristic (occlusion/truncation (occ-trn), bounding box area or object size (size), aspect ratio (asp), visible sides (side) and part visibility (part)). Overall accuracy is indicated by the black dashed line. The difference between max and min indicates sensitivity; the difference between max and overall indicates the impact.

Figure 84 shows an analysis of the influence of the overlap criterion considering the AVP metric. For this overlap criterion analysis we follow the PASCAL VOC formulation: to be considered a true positive, the area of overlap between the predicted BB and GT BB must exceed a threshold.

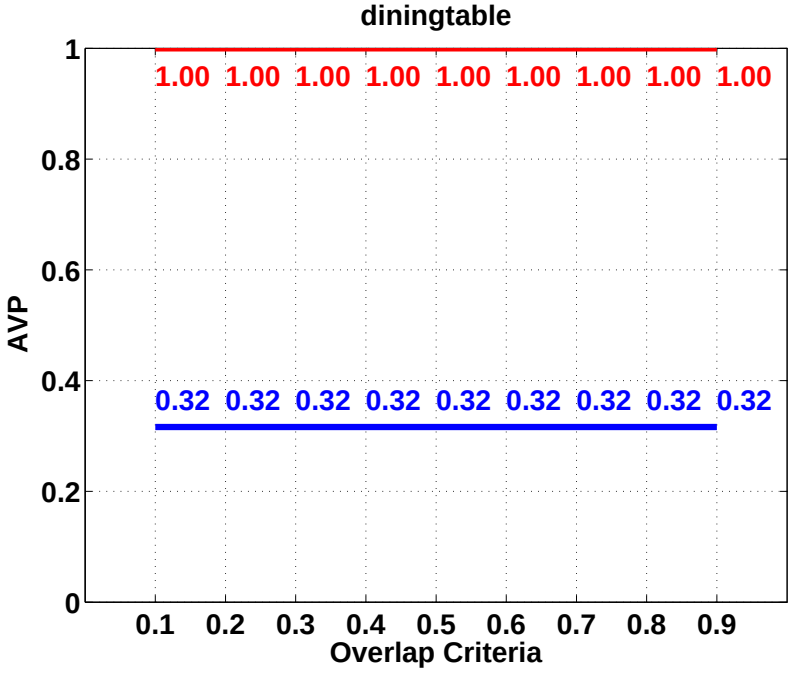


Figure 84: **Simultaneous Object Detection and Pose Estimation.** The detection performance (AP) is represented in red and the pose estimation performance (AVP) in blue.

3.8 motorbike

Statistics for motorbike class:

Number of objects = 376, No difficult objects = 356
Truncated: None = 219, Truncated = 137
Occluded: None = 211, Occluded = 165
Occluded and Truncated: 93
side visible:

```
bottom: Yes = 0  No = 376
front: Yes = 55  No = 321
rear: Yes = 34  No = 342
side: Yes = 171  No = 205
top: Yes = 0  No = 376
```

part visible:

```
part 1 = back_seat: Yes = 201  No = 175
part 2 = front_seat: Yes = 122  No = 254
part 3 = head_center: Yes = 185  No = 191
part 4 = headlight_center: Yes = 225  No = 151
part 5 = left_back_wheel: Yes = 92  No = 284
part 6 = left_front_wheel: Yes = 124  No = 252
part 7 = left_handle_center: Yes = 218  No = 158
part 8 = right_back_wheel: Yes = 102  No = 274
part 9 = right_front_wheel: Yes = 148  No = 228
part 10 = right_handle_center: Yes = 238  No = 138
```

Figure 85 summarizes the viewpoint distribution on the test set for motorbike class considering the 8 views described in Section 1. Figure 86 shows the analysis and the impact on the pose performance of each type of pose error. We also report in Figure 87 the success rate for motorbike class considering each of the 8 viewpoints described in Section 1.

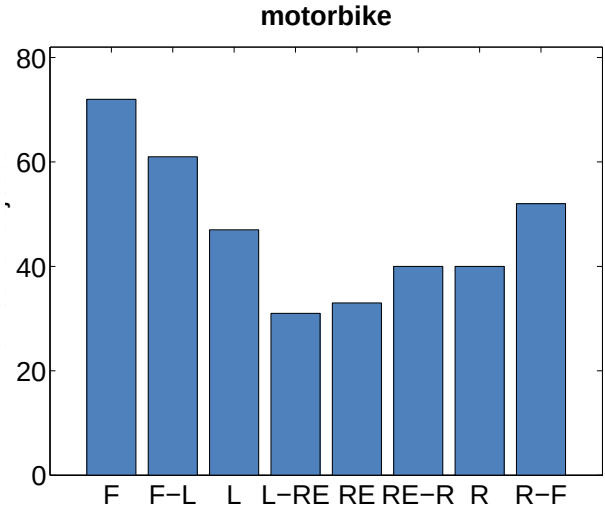
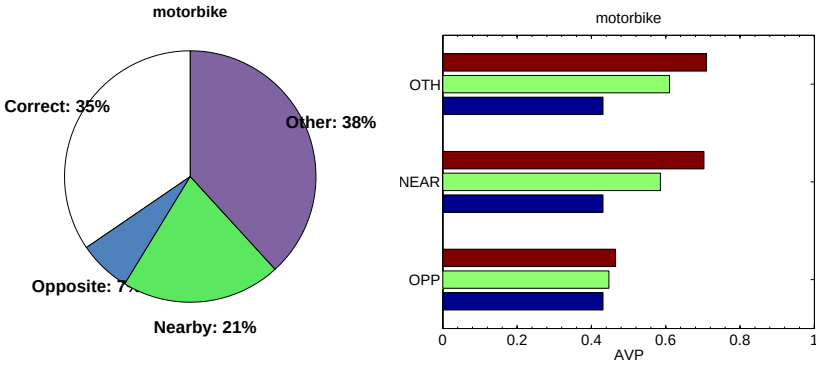


Figure 85: **Analysis of Pose Distribution.** Histogram represents viewpoint distribution on test set for motorbike class from PASCAL3D+ dataset.



(a) (b)

Figure 86: **Analysis of the Main Pose Estimation Errors.** (a) Given the correct detections, pie chart shows the fraction of pose error that are due to Opposite viewpoints (Opposite), confusion with Nearby viewpoints (Nearby), confusion with Other rotations (Other), and correct pose estimations (Correct) for motorbike class. (b) Impact of pose errors in terms of AVP. Blue Bars show the AVP performance obtained when all estimations are considered. Green Bars display the AVP improvement by removing all estimations of one type: **OTH** removes confusion with other rotation viewpoints; **NEAR** removes confusion with nearby viewpoints; **OPP** removes confusion with opposite viewpoints. Brick Red Bars display the AVP improvement by correcting all estimations of one type: **OTH** corrects confusion with other rotation viewpoints; **NEAR** corrects confusion with nearby viewpoints; **OPP** corrects confusion with opposite viewpoints.

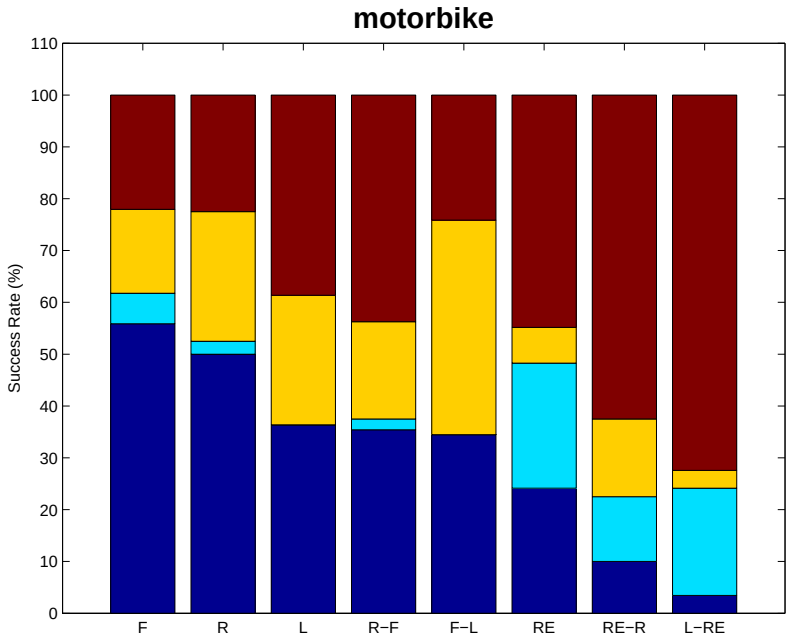
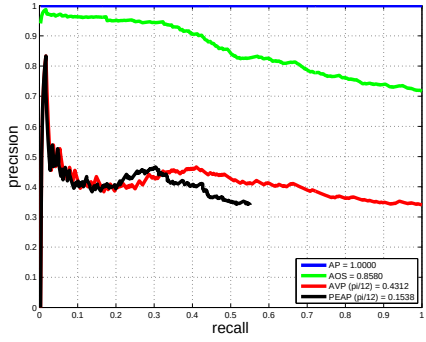
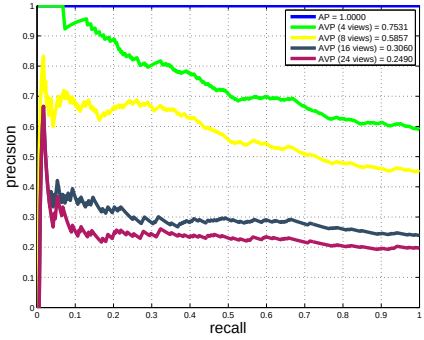


Figure 87: **Success rate for motorbike class considering 8 views.** For each of the 8 viewpoints, we report: In Blue, the percentage of the correct pose estimations. In Cyan, the percentage of confusion with the opposite viewpoints. In Yellow, the percentage of confusion with nearby viewpoints. In Brick Red, the percentage of the confusions with other rotations.

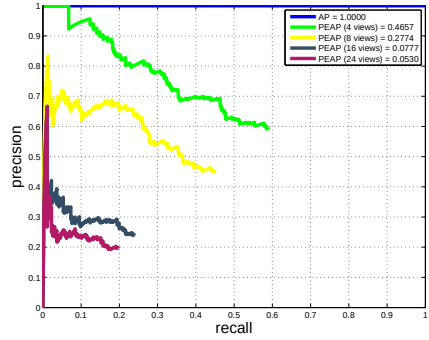
Figure 88 summarizes the detection and pose estimation performances for motorbike class. Figure 88a shows the model performance working on continuous pose estimation. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$. Figures 88b and 88c report the AVP and PEAP performances achieve working on discrete pose estimation. The AVP and PEAP metrics are obtained considering different number of the views: 4, 8, 16 and 24.



(a)



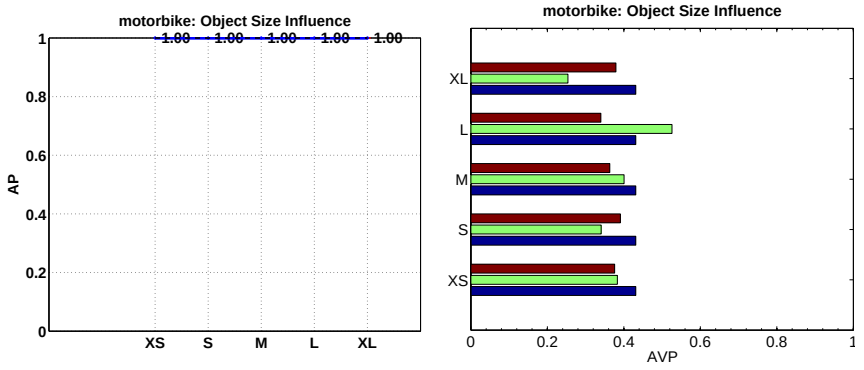
(b)



(c)

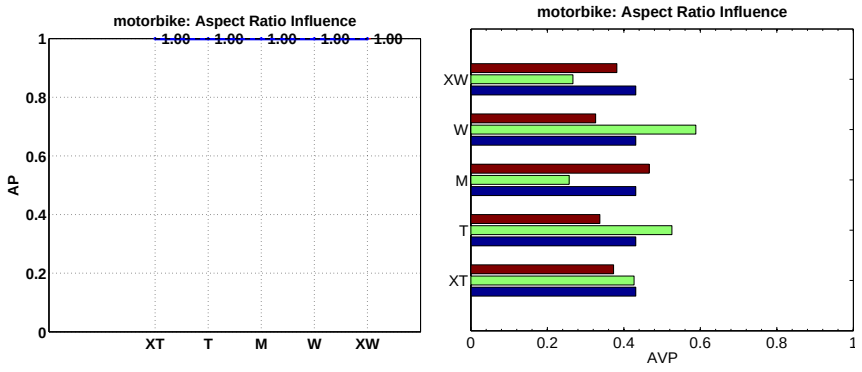
Figure 88: **Precision/Recall Curves.** (a) Performance working on continuous pose estimation. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$. (b) AVP and (c) PEAP performances working on discrete pose estimation. Results obtained by considering 4, 8, 16, and 24 views.

Figures 89, 90, 91, 92 and 93 summarize the main object characteristic influences on detection and pose estimation performances for motorbike class. Figure 89 reports the effect of object size characteristic on object detection and pose estimation performances. Figure 90 shows the influence of aspect ratio characteristic on object detection and pose estimation performances. Figure 91 represents the influence of occluded and truncated objects on detection and pose estimation performances. Figure 92 shows the visible side influence. Figure 93 reports the part visibily effect.



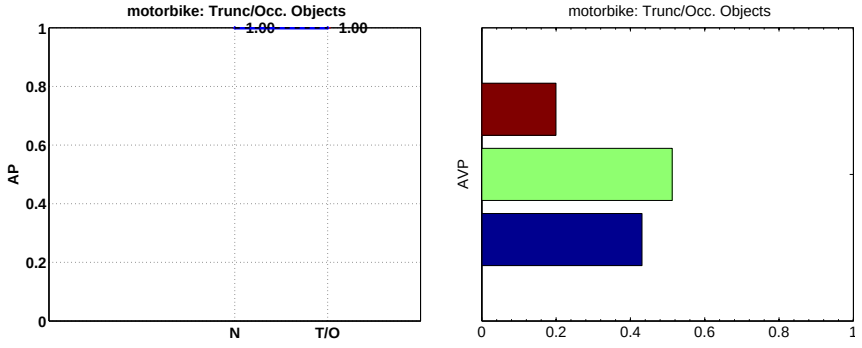
(a) (b)

Figure 89: **Effect of Object Size Characteristic on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP variance. XS and S are considered EXTRA SMALL and SMALL objects, M are MEDIUM objects, L and XL are LARGE and EXTRA LARGE objects. (b) Pose Estimation Results in terms of AVP. Blue Bars show the results when all pose estimations are considered, Green Bars show the results when only the pose estimations corresponding this object size are considered and Brick Red Bars display the results when the pose estimations corresponding this object size are removed.



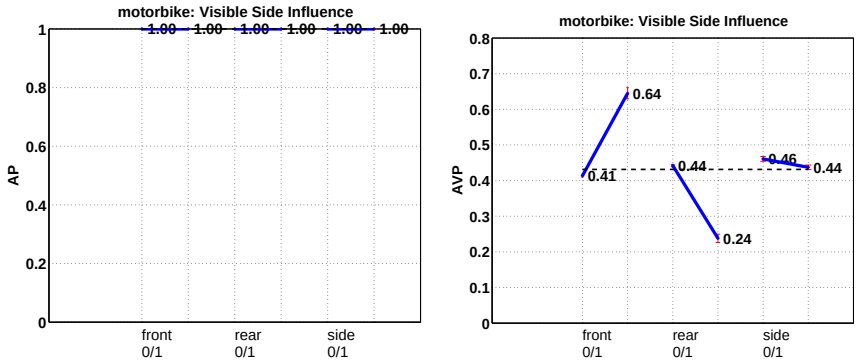
(a) (b)

Figure 90: **Effect of Object Size Characteristic on Object Detection and Pose Estimation.** (a) Detection analysis: AP variance. XS and S are considered EXTRA SMALL and SMALL objects, M are MEDIUM objects, L and XL are LARGE and EXTRA LARGE objects. (b) Pose Estimation Results in terms of AVP. Blue Bars show the results when all pose estimations are considered, Green Bars show the results when only the pose estimations corresponding this object size are considered and Brick Red Bars display the results when the pose estimations corresponding this object size are removed.



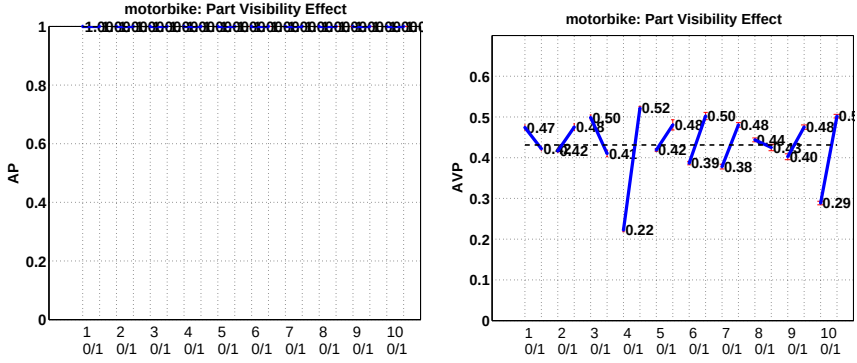
(a) (b)

Figure 91: **Effect of Occluded and Truncated Objects on Detection and Pose Estimation Performances.** (a) Detection analysis: AP variance. N non occluded or truncated object, T/O truncated or occluded object. Results on Pose Estimation in terms of the AVP metric. Blue Bars show the results when all pose estimations are considered. Green Bars display the results when the occluded and truncated objects are removed. Brick Red Bars show the performance when we are only considered the occluded and truncated objects.



(a) (b)

Figure 92: **Effect of Visible Sides on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP performance. Pose Estimation analysis: (b) Pose Estimation analysis: AVP performance. Black dashed lines indicate overall. Visible side: **1** = performance when the side is visible; **0** = performance when the side is not visible.



(a) (b)

Figure 93: **Part Visibility Influence on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP performance. Pose Estimation analysis: (b) Pose Estimation analysis: AVP performance. Black dashed lines indicate overall performance. Visible part: **1** = performance when the part is visible; **0** = performance when the part is not visible. The correspondence between the part number and the part name is indicated at the beginning of this Section (see Statistics for motorbike class).

Figure 94 provides a summary of the sensitivity to each characteristic and the potential impact on improving pose estimation robustness. The worst-performing and best-performing combinations for each object characteristic are averaged over the selected object categories. The difference between the best and the worst performance indicates sensitivity; the difference between the best and the overall indicates the potential impact.

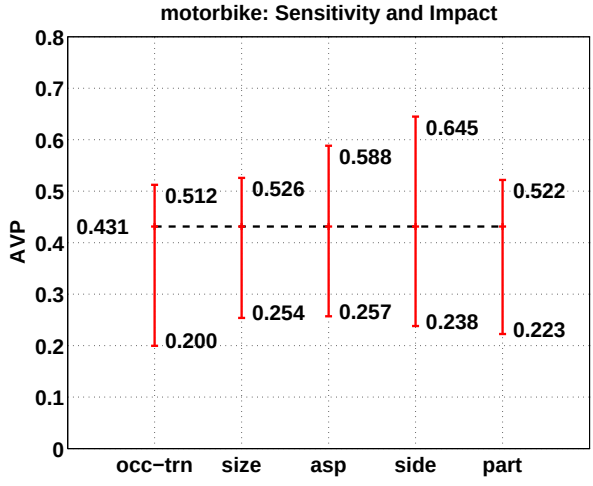


Figure 94: **Summary of Sensitivity and Impact of Object Characteristics.** We show the performance of the highest performing and lowest performing subsets within each characteristic (occlusion/truncation (occ-trn), bounding box area or object size (size), aspect ratio (asp), visible sides (side) and part visibility (part)). Overall accuracy is indicated by the black dashed line. The difference between max and min indicates sensitivity; the difference between max and overall indicates the impact.

Figure 95 shows an analysis of the influence of the overlap criterion considering the AVP metric. For this overlap criterion analysis we follow the PASCAL VOC formulation: to be considered a true positive, the area of overlap between the predicted BB and GT BB must exceed a threshold.

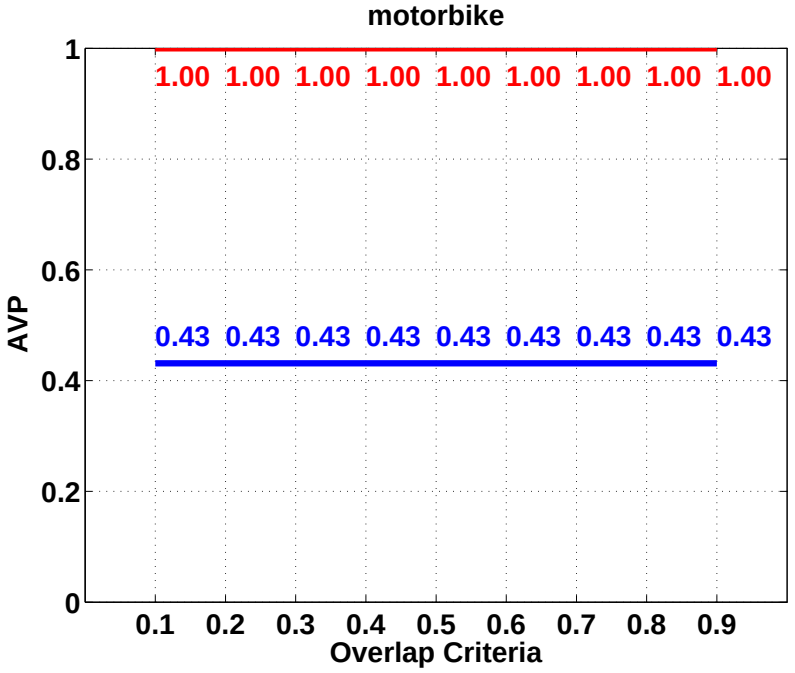


Figure 95: **Simultaneous Object Detection and Pose Estimation.** The detection performance (AP) is represented in red and the pose estimation performance (AVP) in blue.

3.9 sofa

Statistics for sofa class:

Number of objects = 328, No difficult objects = 285
Truncated: None = 59, Truncated = 226
Occluded: None = 105, Occluded = 223
Occluded and Truncated: 163
side visible:

```
bottom: Yes = 0  No = 328
front: Yes = 130  No = 198
rear: Yes = 12  No = 316
side: Yes = 50  No = 278
top: Yes = 0  No = 328
```

part visible:

```
part 1 = front_bottom_left: Yes = 112  No = 216
part 2 = front_bottom_right: Yes = 125  No = 203
part 3 = seat_bottom_left: Yes = 169  No = 159
part 4 = seat_bottom_right: Yes = 168  No = 160
part 5 = left_bottom_back: Yes = 26  No = 302
part 6 = right_bottom_back: Yes = 34  No = 294
part 7 = top_left_corner: Yes = 177  No = 151
part 8 = top_right_corner: Yes = 191  No = 137
part 9 = seat_top_left: Yes = 43  No = 285
part 10 = seat_top_right: Yes = 46  No = 282
```

Figure 96 summarizes the viewpoint distribution on the test set for sofa class considering the 8 views described in Section 1. Figure 97 shows the analysis and the impact on the pose performance of each type of pose error. We also report in Figure 98 the success rate for sofa class considering each of the 8 viewpoints described in Section 1.

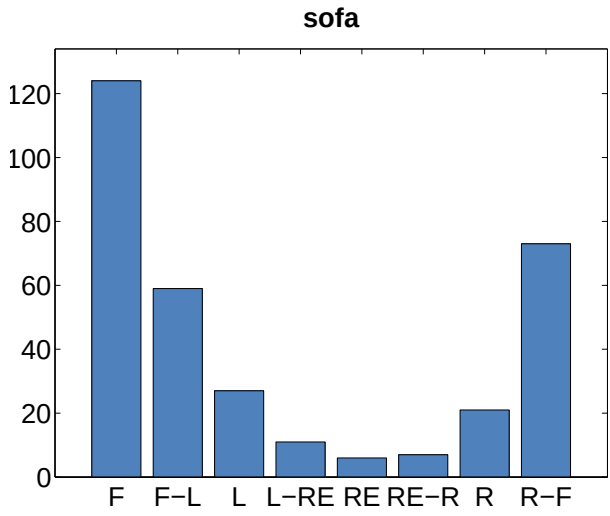
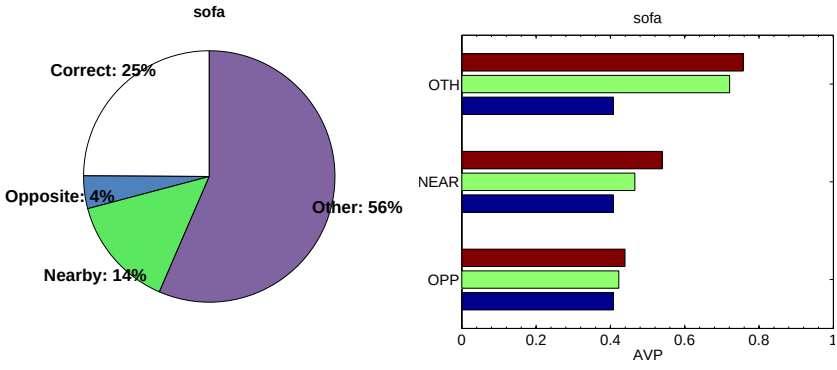


Figure 96: **Analysis of Pose Distribution.** Histogram represents viewpoint distribution on test set for sofa class from PASCAL3D+ dataset.



(a) (b)

Figure 97: **Analysis of the Main Pose Estimation Errors.** (a) Given the correct detections, pie chart shows the fraction of pose error that are due to Opposite viewpoints (Opposite), confusion with Nearby viewpoints (Nearby), confusion with Other rotations (Other), and correct pose estimations (Correct) for sofa class. (b) Impact of pose errors in terms of AVP. Blue Bars show the AVP performance obtained when all estimations are considered. Green Bars display the AVP improvement by removing all estimations of one type: **OTH** removes confusion with other rotation viewpoints; **NEAR** removes confusion with nearby viewpoints; **OPP** removes confusion with opposite viewpoints. Brick Red Bars display the AVP improvement by correcting all estimations of one type: **OTH** corrects confusion with other rotation viewpoints; **NEAR** corrects confusion with nearby viewpoints; **OPP** corrects confusion with opposite viewpoints.

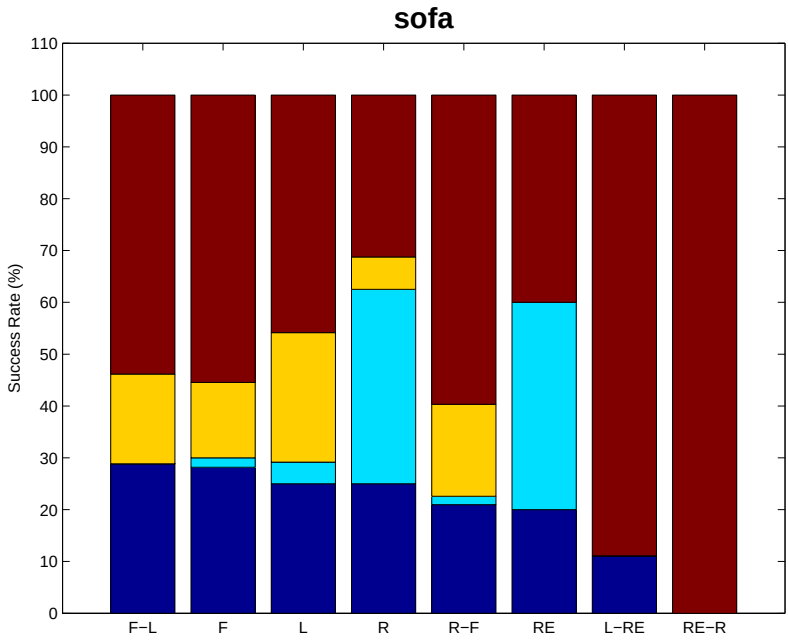
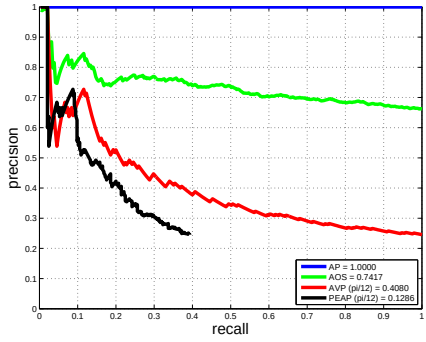
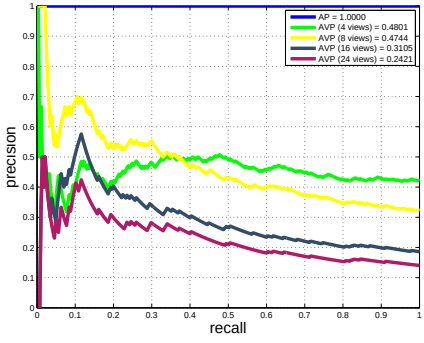


Figure 98: **Success rate for sofa class considering 8 views.** For each of the 8 viewpoints, we report: In Blue, the percentage of the correct pose estimations. In Cyan, the percentage of confusion with the opposite viewpoints. In Yellow, the percentage of confusion with nearby viewpoints. In Brick Red, the percentage of the confusions with other rotations.

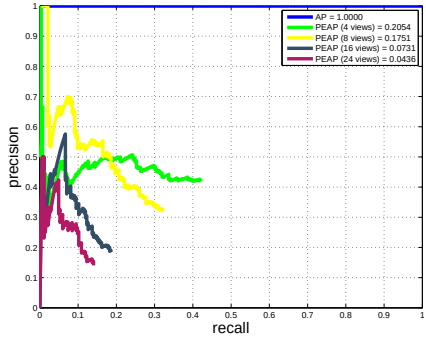
Figure 99 summarizes the detection and pose estimation performances for sofa class. Figure 99a shows the model performance working on continuous pose estimation. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$. Figures 99b and 99c report the AVP and PEAP performances achieve working on discrete pose estimation. The AVP and PEAP metrics are obtained considering different number of the views: 4, 8, 16 and 24.



(a)



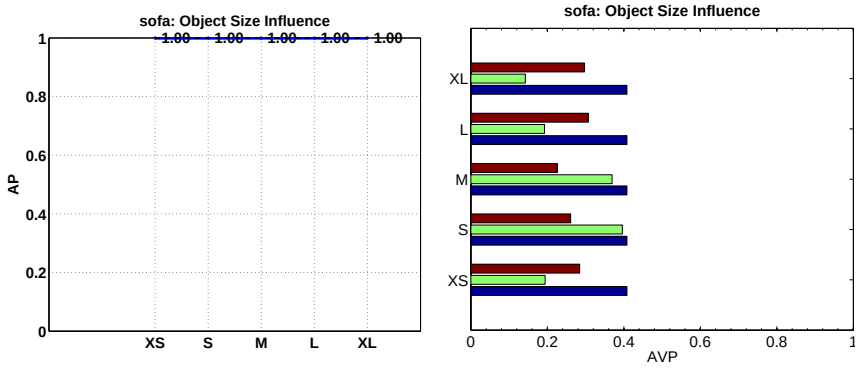
(b)



(c)

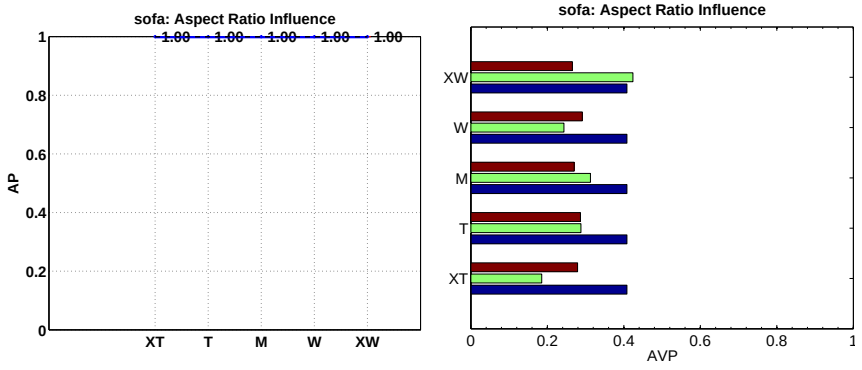
Figure 99: **Precision/Recall Curves.** (a) Performance working on continuous pose estimation. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$. (b) AVP and (c) PEAP performances working on discrete pose estimation. Results obtained by considering 4, 8, 16, and 24 views.

Figures 100, 101, 102, 103 and 104 summarize the main object characteristic influences on detection and pose estimation performances for sofa class. Figure 100 reports the effect of object size characteristic on object detection and pose estimation performances. Figure 101 shows the influence of aspect ratio characteristic on object detection and pose estimation performances. Figure 102 represents the influence of occluded and truncated objects on detection and pose estimation performances. Figure 103 shows the visible side influence. Figure 104 reports the part visibily effect.



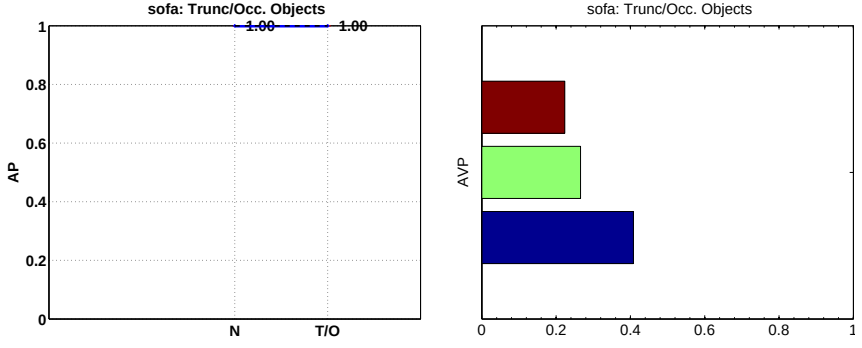
(a) (b)

Figure 100: **Effect of Object Size Characteristic on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP variance. XS and S are considered EXTRA SMALL and SMALL objects, M are MEDIUM objects, L and XL are LARGE and EXTRA LARGE objects. (b) Pose Estimation Results in terms of AVP. Blue Bars show the results when all pose estimations are considered, Green Bars show the results when only the pose estimations corresponding this object size are considered and Brick Red Bars display the results when the pose estimations corresponding this object size are removed.



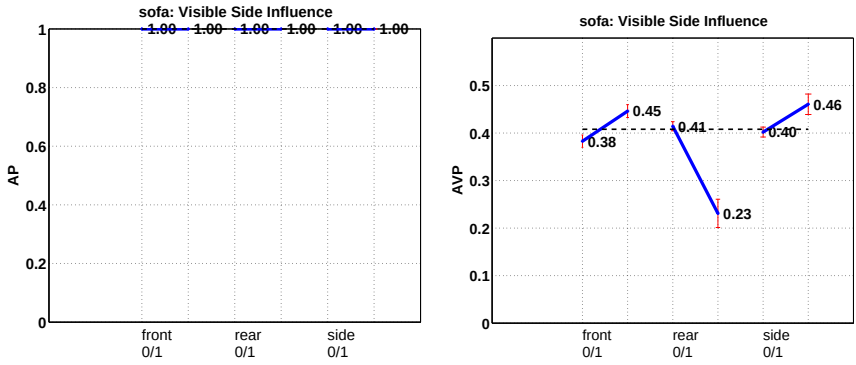
(a) (b)

Figure 101: **Effect of Object Size Characteristic on Object Detection and Pose Estimation.** (a) Detection analysis: AP variance. XS and S are considered EXTRA SMALL and SMALL objects, M are MEDIUM objects, L and XL are LARGE and EXTRA LARGE objects. (b) Pose Estimation Results in terms of AVP. Blue Bars show the results when all pose estimations are considered, Green Bars show the results when only the pose estimations corresponding this object size are considered and Brick Red Bars display the results when the pose estimations corresponding this object size are removed.



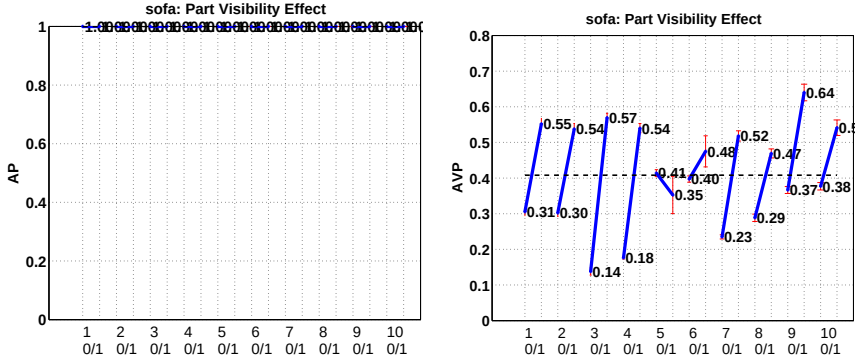
(a) (b)

Figure 102: **Effect of Occluded and Truncated Objects on Detection and Pose Estimation Performances.** (a) Detection analysis: AP variance. N non occluded or truncated object, T/O truncated or occluded object. Results on Pose Estimation in terms of the AVP metric. Blue Bars show the results when all pose estimations are considered. Green Bars display the results when the occluded and truncated objects are removed. Brick Red Bars show the performance when we are only considered the occluded and truncated objects.



(a) (b)

Figure 103: **Effect of Visible Sides on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP performance. Pose Estimation analysis: (b) Pose Estimation analysis: AVP performance. Black dashed lines indicate overall. Visible side: **1** = performance when the side is visible; **0** = performance when the side is not visible.



(a) (b)

Figure 104: **Part Visibility Influence on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP performance. Pose Estimation analysis: (b) Pose Estimation analysis: AVP performance. Black dashed lines indicate overall performance. Visible part: **1** = performance when the part is visible; **0** = performance when the part is not visible. The correspondence between the part number and the part name is indicated at the beginning of this Section (see Statistics for sofa class).

Figure 105 provides a summary of the sensitivity to each characteristic and the potential impact on improving pose estimation robustness. The worst-performing and best-performing combinations for each object characteristic are averaged over the selected object categories. The difference between the best and the worst performance indicates sensitivity; the difference between the best and the overall indicates the potential impact.

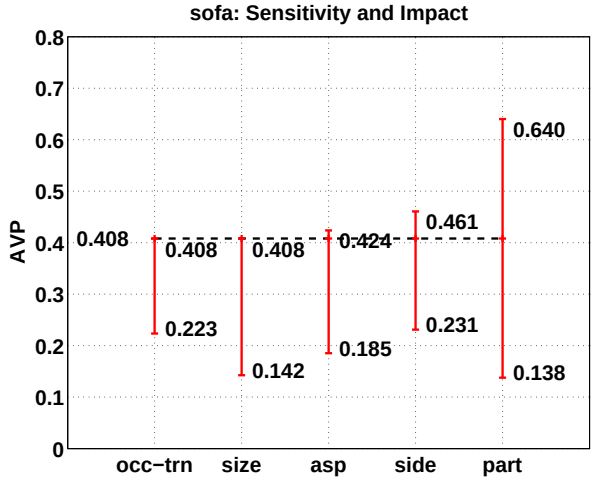


Figure 105: **Summary of Sensitivity and Impact of Object Characteristics.** We show the performance of the highest performing and lowest performing subsets within each characteristic (occlusion/truncation (occ-trn), bounding box area or object size (size), aspect ratio (asp), visible sides (side) and part visibility (part)). Overall accuracy is indicated by the black dashed line. The difference between max and min indicates sensitivity; the difference between max and overall indicates the impact.

Figure 106 shows an analysis of the influence of the overlap criterion considering the AVP metric. For this overlap criterion analysis we follow the PASCAL VOC formulation: to be considered a true positive, the area of overlap between the predicted BB and GT BB must exceed a threshold.

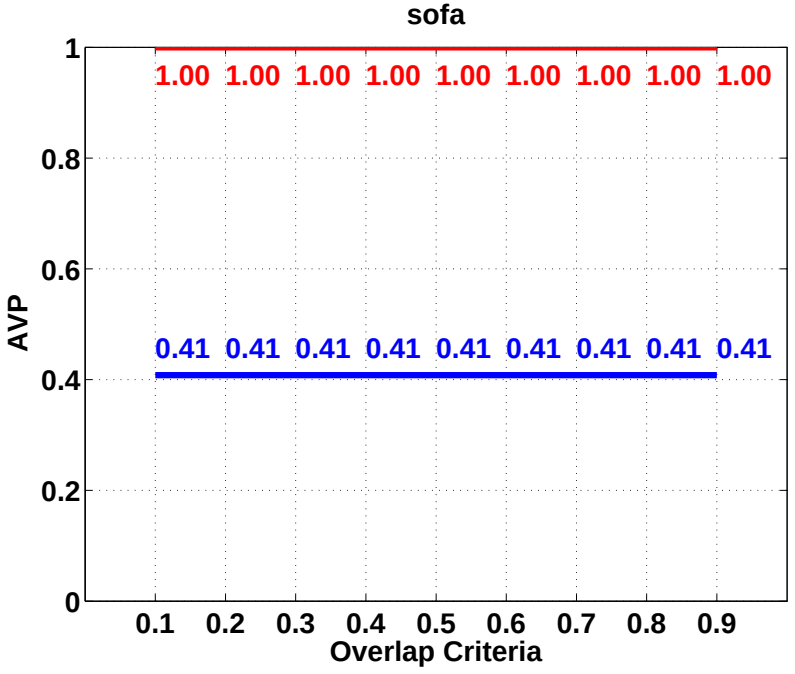


Figure 106: **Simultaneous Object Detection and Pose Estimation.** The detection performance (AP) is represented in red and the pose estimation performance (AVP) in blue.

3.10 train

Statistics for train class:

Number of objects = 325, No difficult objects = 315
Truncated: None = 130, Truncated = 185
Occluded: None = 231, Occluded = 94
Occluded and Truncated: 68
side visible:

```
bottom: Yes = 0  No = 325
front: Yes = 54  No = 271
rear: Yes = 16  No = 309
side: Yes = 55  No = 270
top: Yes = 0  No = 325
```

part visible:

```
part 1 = head_left_bottom: Yes = 176  No = 149
part 2 = head_left_top: Yes = 143  No = 182
part 3 = head_right_bottom: Yes = 174  No = 151
part 4 = head_right_top: Yes = 141  No = 184
part 5 = head_top: Yes = 49  No = 276
part 6 = mid1_left_bottom: Yes = 96  No = 229
part 7 = mid1_left_top: Yes = 111  No = 214
part 8 = mid1_right_bottom: Yes = 120  No = 205
part 9 = mid1_right_top: Yes = 131  No = 194
part 10 = mid2_left_bottom: Yes = 61  No = 264
part 11 = mid2_left_top: Yes = 65  No = 260
part 12 = mid2_right_bottom: Yes = 70  No = 255
part 13 = mid2_right_top: Yes = 74  No = 251
part 14 = tail_left_bottom: Yes = 38  No = 287
part 15 = tail_left_top: Yes = 40  No = 285
part 16 = tail_right_bottom: Yes = 32  No = 293
part 17 = tail_right_top: Yes = 33  No = 292
```

Figure 107 summarizes the viewpoint distribution on the test set for train class considering the 8 views described in Section 1. Figure 108 shows the analysis and the impact on the pose performance of each type of pose error. We also report in Figure 109 the success rate for train class considering each of the 8 viewpoints described in Section 1.

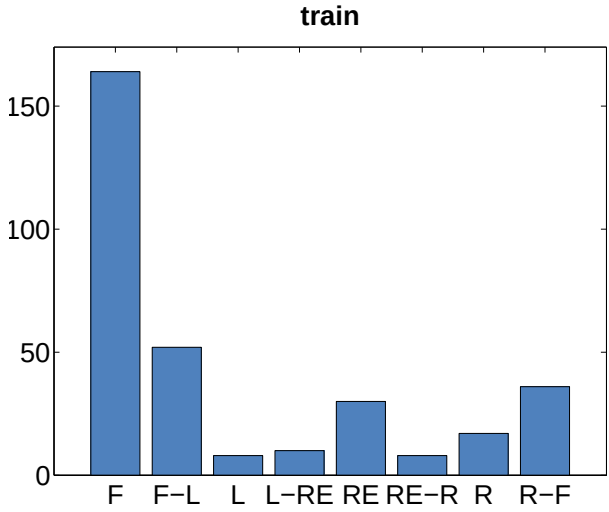
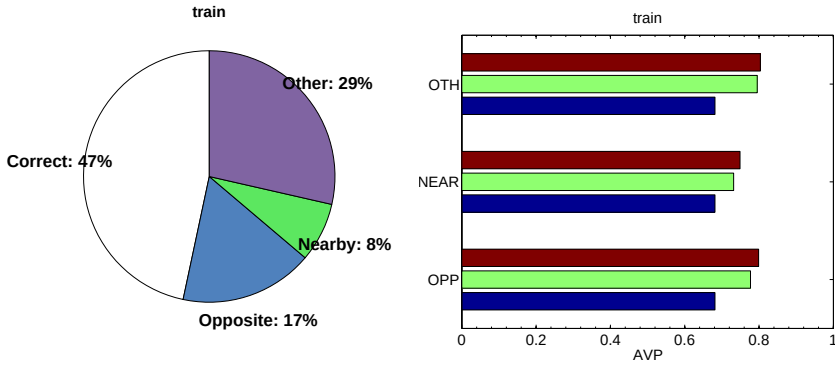


Figure 107: **Analysis of Pose Distribution.** Histogram represents viewpoint distribution on test set for train class from PASCAL3D+ dataset.



(a) (b)

Figure 108: **Analysis of the Main Pose Estimation Errors.** (a) Given the correct detections, pie chart shows the fraction of pose error that are due to Opposite viewpoints (Opposite), confusion with Nearby viewpoints (Nearby), confusion with Other rotations (Other), and correct pose estimations (Correct) for train class. (b) Impact of pose errors in terms of AVP. Blue Bars show the AVP performance obtained when all estimations are considered. Green Bars display the AVP improvement by removing all estimations of one type: **OTH** removes confusion with other rotation viewpoints; **NEAR** removes confusion with nearby viewpoints; **OPP** removes confusion with opposite viewpoints. Brick Red Bars display the AVP improvement by correcting all estimations of one type: **OTH** corrects confusion with other rotation viewpoints; **NEAR** corrects confusion with nearby viewpoints; **OPP** corrects confusion with opposite viewpoints.

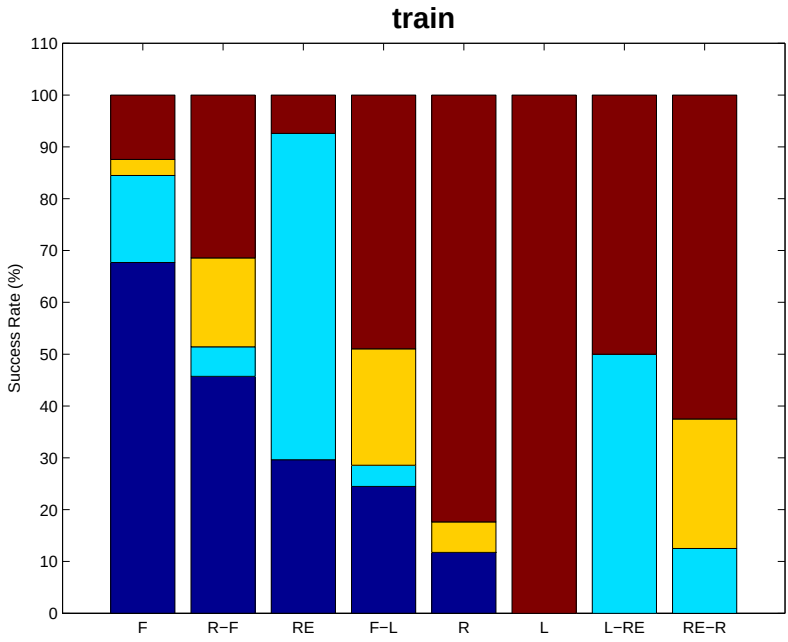
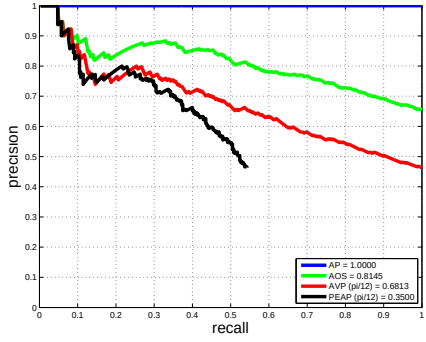
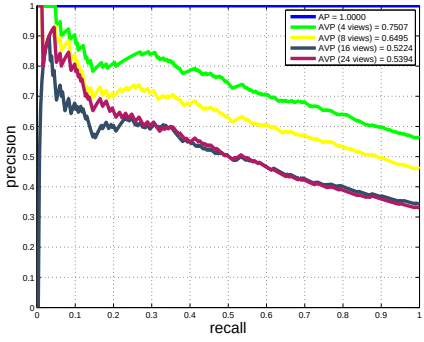


Figure 109: **Success rate for train class considering 8 views.** For each of the 8 viewpoints, we report: In Blue, the percentage of the correct pose estimations. In Cyan, the percentage of confusion with the opposite viewpoints. In Yellow, the percentage of confusion with nearby viewpoints. In Brick Red, the percentage of the confusions with other rotations.

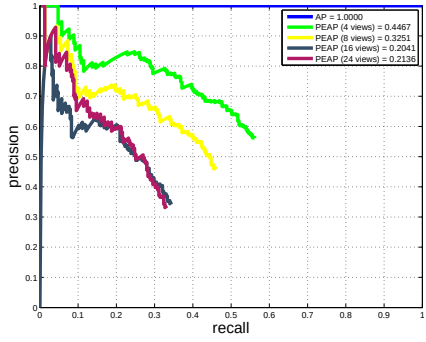
Figure 110 summarizes the detection and pose estimation performances for train class. Figure 110a shows the model performance working on continuous pose estimation. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$. Figures 110b and 110c report the AVP and PEAP performances achieve working on discrete pose estimation. The AVP and PEAP metrics are obtained considering different number of the views: 4, 8, 16 and 24.



(a)



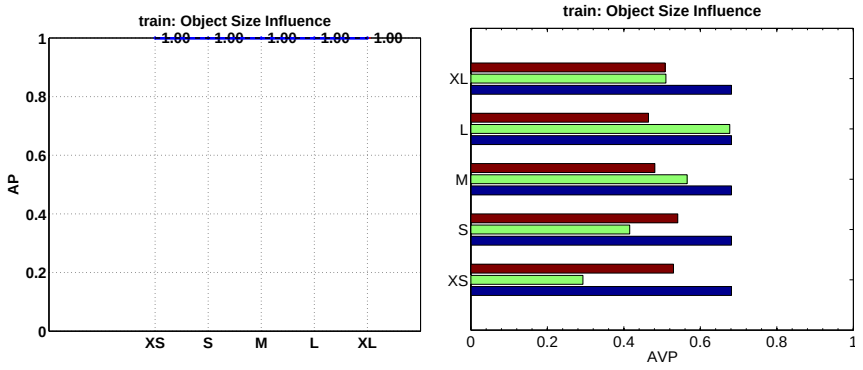
(b)



(c)

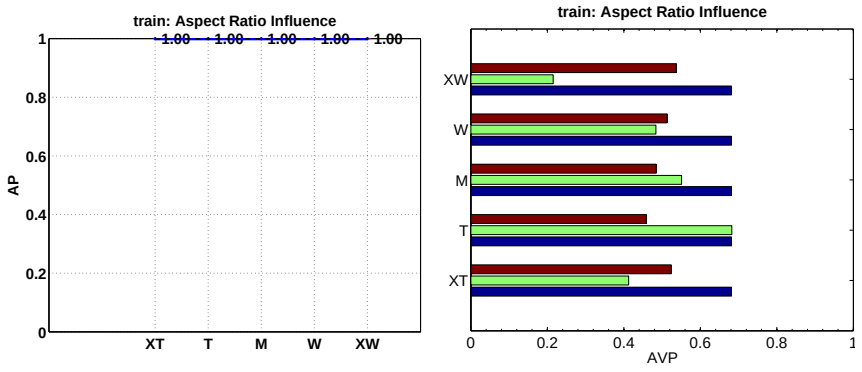
Figure 110: **Precision/Recall Curves.** (a) Performance working on continuous pose estimation. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$. (b) AVP and (c) PEAP performances working on discrete pose estimation. Results obtained by considering 4, 8, 16, and 24 views.

Figures 111, 112, 113, 114 and 115 summarize the main object characteristic influences on detection and pose estimation performances for train class. Figure 111 reports the effect of object size characteristic on object detection and pose estimation performances. Figure 112 shows the influence of aspect ratio characteristic on object detection and pose estimation performances. Figure 113 represents the influence of occluded and truncated objects on detection and pose estimation performances. Figure 114 shows the visible side influence. Figure 115 reports the part visibily effect.



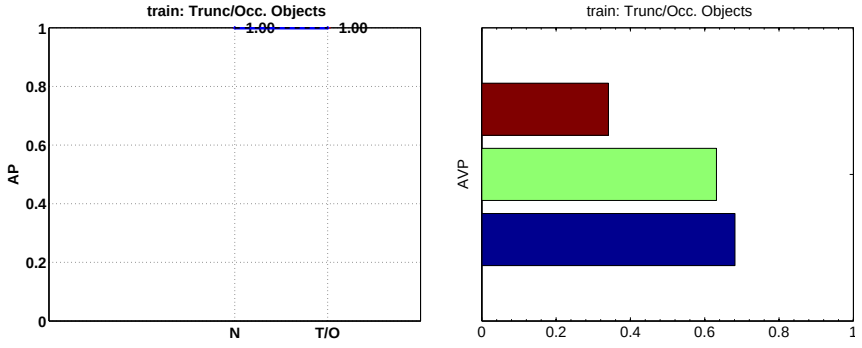
(a) (b)

Figure 111: **Effect of Object Size Characteristic on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP variance. XS and S are considered EXTRA SMALL and SMALL objects, M are MEDIUM objects, L and XL are LARGE and EXTRA LARGE objects. (b) Pose Estimation Results in terms of AVP. Blue Bars show the results when all pose estimations are considered, Green Bars show the results when only the pose estimations corresponding this object size are considered and Brick Red Bars display the results when the pose estimations corresponding this object size are removed.



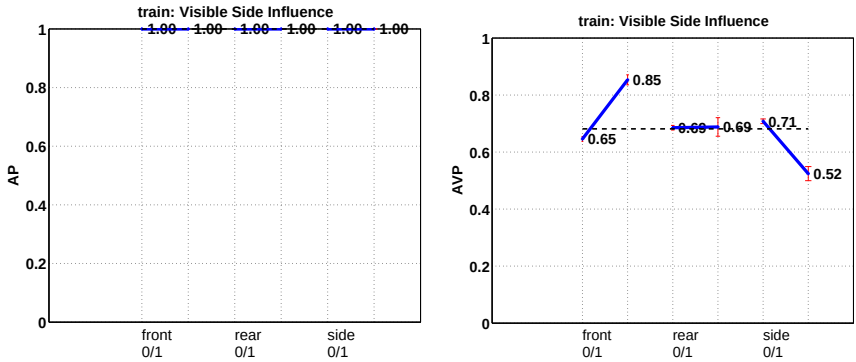
(a) (b)

Figure 112: **Effect of Object Size Characteristic on Object Detection and Pose Estimation.** (a) Detection analysis: AP variance. XS and S are considered EXTRA SMALL and SMALL objects, M are MEDIUM objects, L and XL are LARGE and EXTRA LARGE objects. (b) Pose Estimation Results in terms of AVP. Blue Bars show the results when all pose estimations are considered, Green Bars show the results when only the pose estimations corresponding this object size are considered and Brick Red Bars display the results when the pose estimations corresponding this object size are removed.



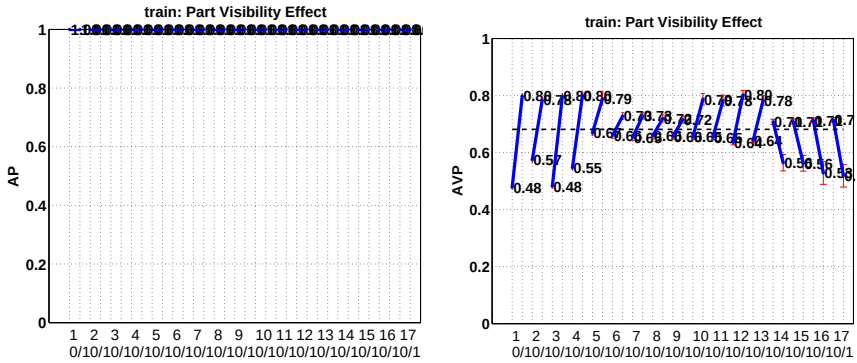
(a) (b)

Figure 113: **Effect of Occluded and Truncated Objects on Detection and Pose Estimation Performances.** (a) Detection analysis: AP variance. N non occluded or truncated object, T/O truncated or occluded object. Results on Pose Estimation in terms of the AVP metric. Blue Bars show the results when all pose estimations are considered. Green Bars display the results when the occluded and truncated objects are removed. Brick Red Bars show the performance when we are only considered the occluded and truncated objects.



(a) (b)

Figure 114: **Effect of Visible Sides on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP performance. Pose Estimation analysis: (b) Pose Estimation analysis: AVP performance. Black dashed lines indicate overall. Visible side: **1** = performance when the side is visible; **0** = performance when the side is not visible.



(a) (b)

Figure 115: **Part Visibility Influence on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP performance. Pose Estimation analysis: (b) Pose Estimation analysis: AVP performance. Black dashed lines indicate overall performance. Visible part: **1** = performance when the part is visible; **0** = performance when the part is not visible. The correspondence between the part number and the part name is indicated at the beginning of this Section (see Statistics for train class).

Figure 116 provides a summary of the sensitivity to each characteristic and the potential impact on improving pose estimation robustness. The worst-performing and best-performing combinations for each object characteristic are averaged over the selected object categories. The difference between the best and the worst performance indicates sensitivity; the difference between the best and the overall indicates the potential impact.

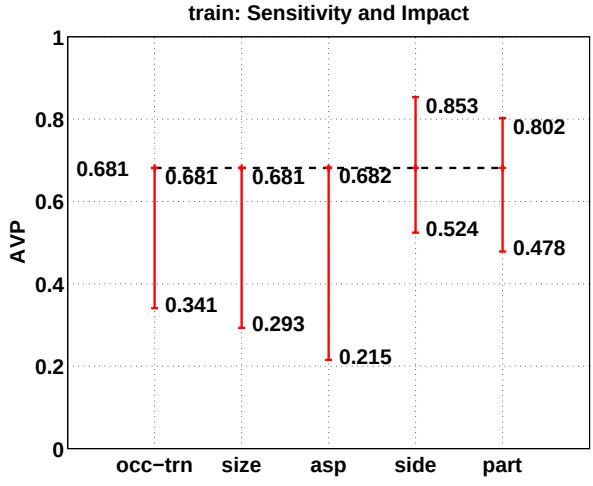


Figure 116: **Summary of Sensitivity and Impact of Object Characteristics.** We show the performance of the highest performing and lowest performing subsets within each characteristic (occlusion/truncation (occ-trn), bounding box area or object size (size), aspect ratio (asp), visible sides (side) and part visibility (part)). Overall accuracy is indicated by the black dashed line. The difference between max and min indicates sensitivity; the difference between max and overall indicates the impact.

Figure 117 shows an analysis of the influence of the overlap criterion considering the AVP metric. For this overlap criterion analysis we follow the PASCAL VOC formulation: to be considered a true positive, the area of overlap between the predicted BB and GT BB must exceed a threshold.

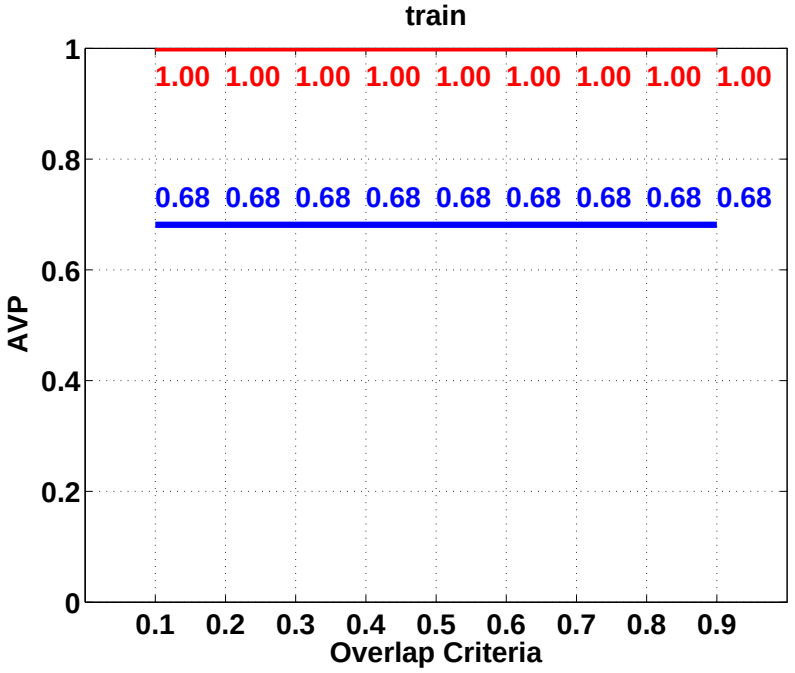


Figure 117: **Simultaneous Object Detection and Pose Estimation.** The detection performance (AP) is represented in red and the pose estimation performance (AVP) in blue.

3.11 tvmonitor

Statistics for tvmonitor class:

Number of objects = 410, No difficult objects = 392
Truncated: None = 269, Truncated = 123
Occluded: None = 305, Occluded = 105
Occluded and Truncated: 43
side visible:

```
bottom: Yes = 0  No = 410
front: Yes = 244  No = 166
rear: Yes = 23  No = 387
side: Yes = 31  No = 379
top: Yes = 0  No = 410
```

part visible:

```
part 1 = front_bottom_left: Yes = 290  No = 120
part 2 = front_bottom_right: Yes = 308  No = 102
part 3 = front_top_left: Yes = 333  No = 77
part 4 = front_top_right: Yes = 328  No = 82
part 5 = back_bottom_left: Yes = 33  No = 377
part 6 = back_bottom_right: Yes = 28  No = 382
part 7 = back_top_left: Yes = 41  No = 369
part 8 = back_top_right: Yes = 33  No = 377
```

Figure 118 summarizes the viewpoint distribution on the test set for tvmonitor class considering the 8 views described in Section 1. Figure 119 shows the analysis and the impact on the pose performance of each type of pose error. We also report in Figure 120 the success rate for tvmonitor class considering each of the 8 viewpoints described in Section 1.

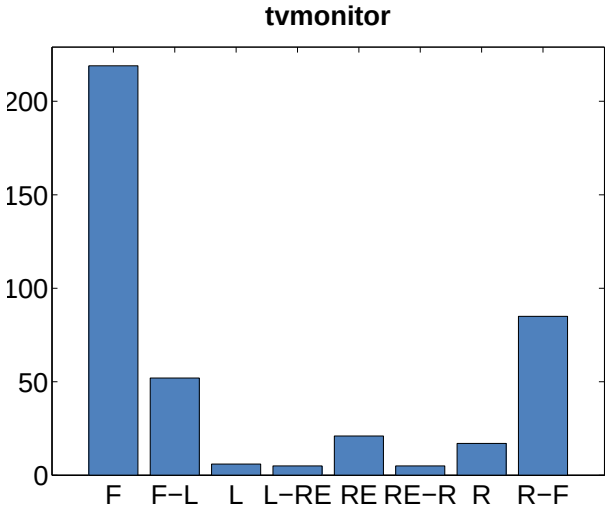
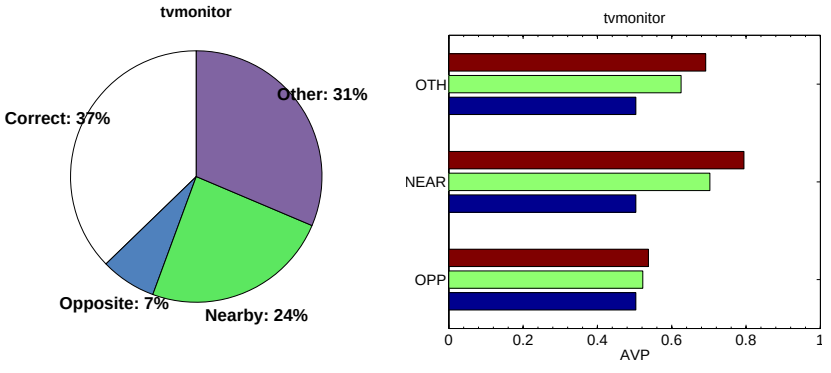


Figure 118: **Analysis of Pose Distribution.** Histogram represents viewpoint distribution on test set for tvmonitor class from PASCAL3D+ dataset.



(a) (b)

Figure 119: **Analysis of the Main Pose Estimation Errors.** (a) Given the correct detections, pie chart shows the fraction of pose error that are due to Opposite viewpoints (Opposite), confusion with Nearby viewpoints (Nearby), confusion with Other rotations (Other), and correct pose estimations (Correct) for tvmonitor class. (b) Impact of pose errors in terms of AVP. Blue Bars show the AVP performance obtained when all estimations are considered. Green Bars display the AVP improvement by removing all estimations of one type: **OTH** removes confusion with other rotation viewpoints; **NEAR** removes confusion with nearby viewpoints; **OPP** removes confusion with opposite viewpoints. Brick Red Bars display the AVP improvement by correcting all estimations of one type: **OTH** corrects confusion with other rotation viewpoints; **NEAR** corrects confusion with nearby viewpoints; **OPP** corrects confusion with opposite viewpoints.

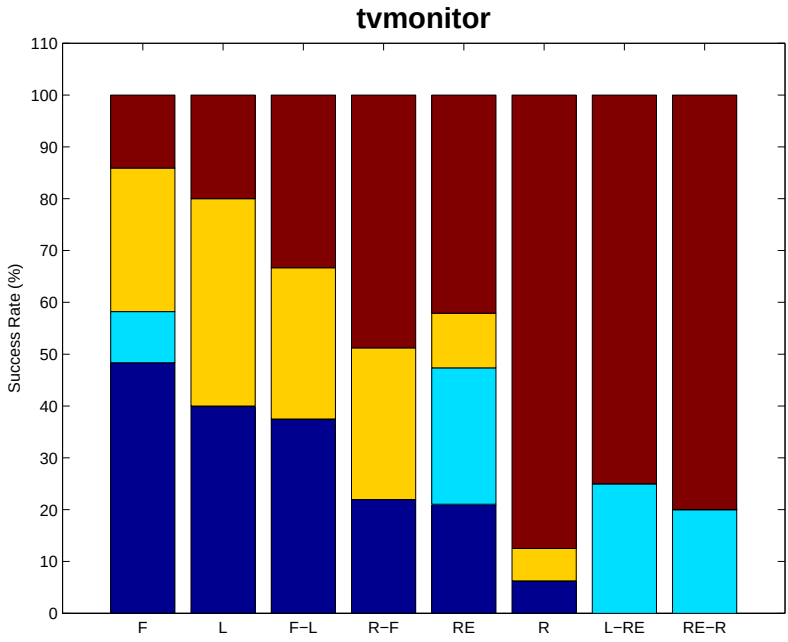
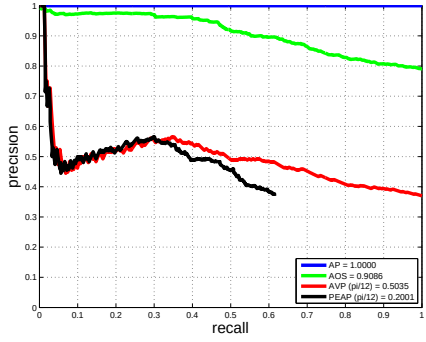
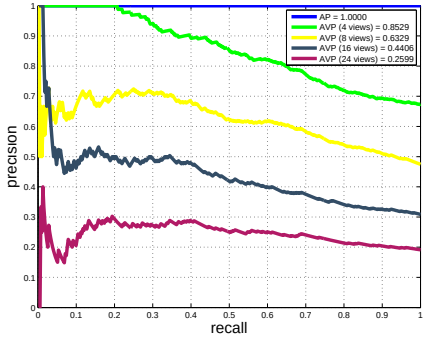


Figure 120: **Success rate for tvmonitor class considering 8 views.** For each of the 8 viewpoints, we report: In Blue, the percentage of the correct pose estimations. In Cyan, the percentage of confusion with the opposite viewpoints. In Yellow, the percentage of confusion with nearby viewpoints. In Brick Red, the percentage of the confusions with other rotations.

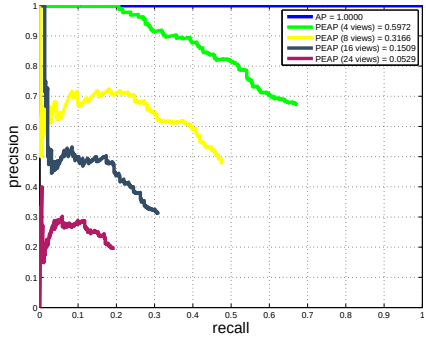
Figure 121 summarizes the detection and pose estimation performances for tvmonitor class. Figure 121a shows the model performance working on continuous pose estimation. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$. Figures 121b and 121c report the AVP and PEAP performances achieve working on discrete pose estimation. The AVP and PEAP metrics are obtained considering different number of the views: 4, 8, 16 and 24.



(a)



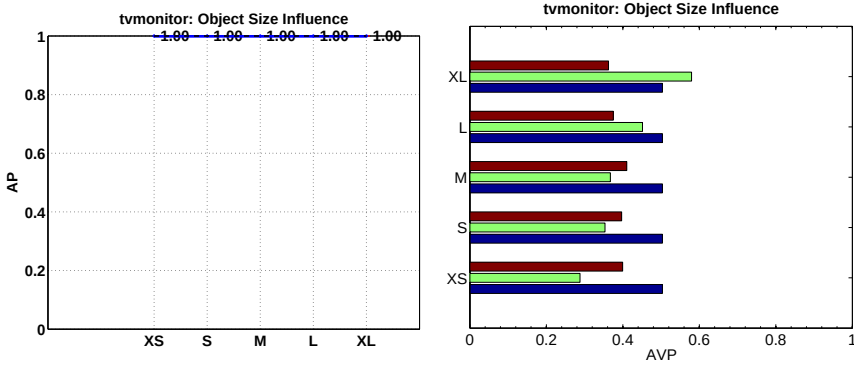
(b)



(c)

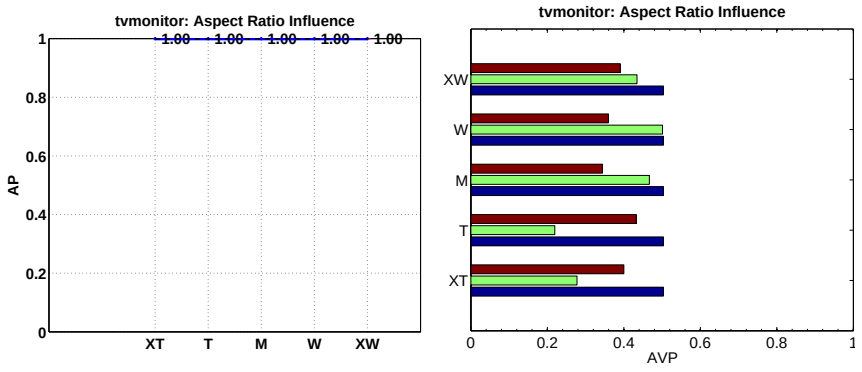
Figure 121: **Precision/Recall Curves.** (a) Performance working on continuous pose estimation. For AVP and PEAP the results are obtained considering a threshold equal to $\frac{\pi}{12}$. (b) AVP and (c) PEAP performances working on discrete pose estimation. Results obtained by considering 4, 8, 16, and 24 views.

Figures 122, 123, 124, 125 and 126 summarize the main object characteristic influences on detection and pose estimation performances for tvmonitor class. Figure 122 reports the effect of object size characteristic on object detection and pose estimation performances. Figure 123 shows the influence of aspect ratio characteristic on object detection and pose estimation performances. Figure 124 represents the influence of occluded and truncated objects on detection and pose estimation performances. Figure 125 shows the visible side influence. Figure 126 reports the part visibily effect.



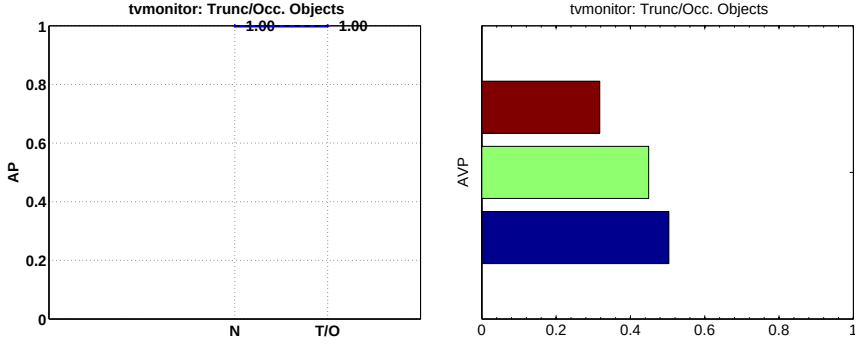
(a) (b)

Figure 122: **Effect of Object Size Characteristic on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP variance. XS and S are considered EXTRA SMALL and SMALL objects, M are MEDIUM objects, L and XL are LARGE and EXTRA LARGE objects. (b) Pose Estimation Results in terms of AVP. Blue Bars show the results when all pose estimations are considered, Green Bars show the results when only the pose estimations corresponding this object size are considered and Brick Red Bars display the results when the pose estimations corresponding this object size are removed.



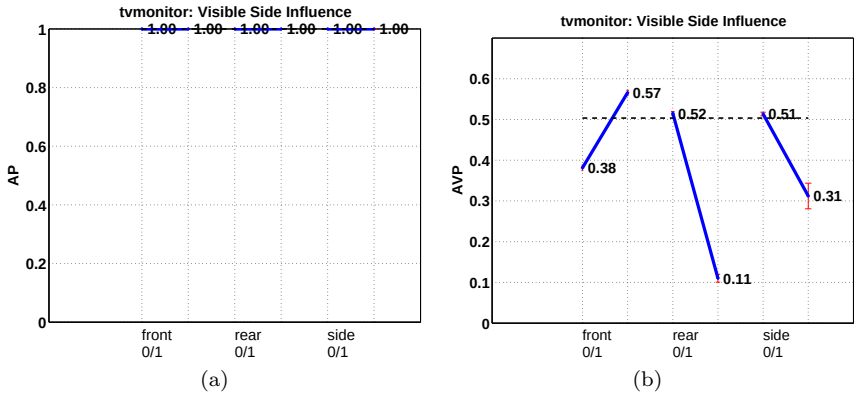
(a) (b)

Figure 123: **Effect of Object Size Characteristic on Object Detection and Pose Estimation.** (a) Detection analysis: AP variance. XS and S are considered EXTRA SMALL and SMALL objects, M are MEDIUM objects, L and XL are LARGE and EXTRA LARGE objects. (b) Pose Estimation Results in terms of AVP. Blue Bars show the results when all pose estimations are considered, Green Bars show the results when only the pose estimations corresponding this object size are considered and Brick Red Bars display the results when the pose estimations corresponding this object size are removed.



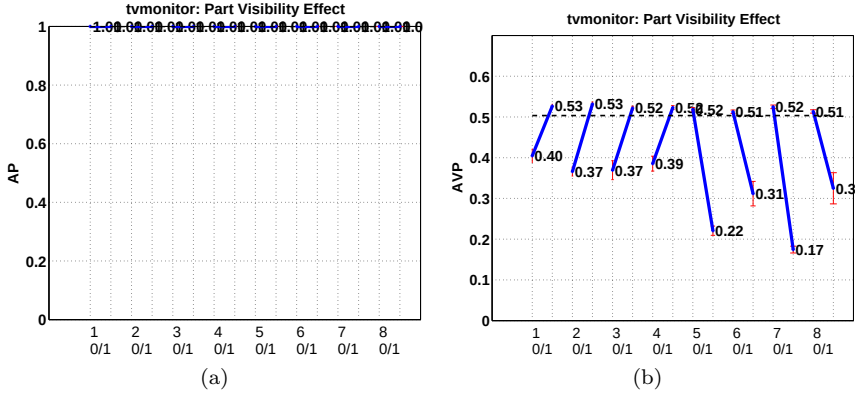
(a) (b)

Figure 124: **Effect of Occluded and Truncated Objects on Detection and Pose Estimation Performances.** (a) Detection analysis: AP variance. N non occluded or truncated object, T/O truncated or occluded object. Results on Pose Estimation in terms of the AVP metric. Blue Bars show the results when all pose estimations are considered. Green Bars display the results when the occluded and truncated objects are removed. Brick Red Bars show the performance when we are only considered the occluded and truncated objects.



(a) (b)

Figure 125: **Effect of Visible Sides on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP performance. Pose Estimation analysis: (b) Pose Estimation analysis: AVP performance. Black dashed lines indicate overall. Visible side: **1** = performance when the side is visible; **0** = performance when the side is not visible.



(a) (b)

Figure 126: **Part Visibility Influence on Object Detection and Pose Estimation Performances.** (a) Detection analysis: AP performance. Pose Estimation analysis: (b) Pose Estimation analysis: AVP performance. Black dashed lines indicate overall performance. Visible part: **1** = performance when the part is visible; **0** = performance when the part is not visible. The correspondence between the part number and the part name is indicated at the beginning of this Section (see Statistics for tvmonitor class).

Figure 127 provides a summary of the sensitivity to each characteristic and the potential impact on improving pose estimation robustness. The worst-performing and best-performing combinations for each object characteristic are averaged over the selected object categories. The difference between the best and the worst performance indicates sensitivity; the difference between the best and the overall indicates the potential impact.

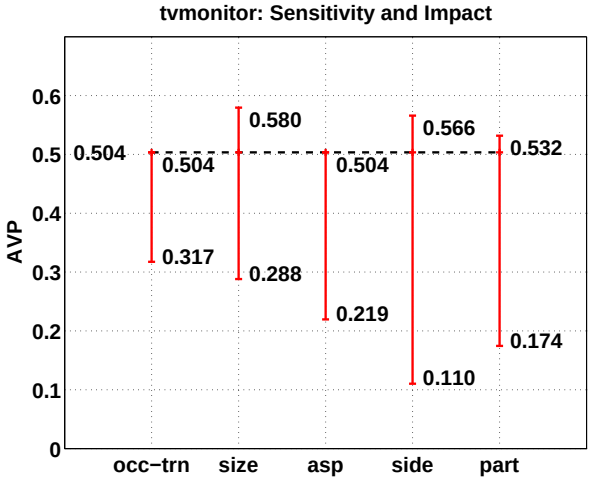


Figure 127: **Summary of Sensitivity and Impact of Object Characteristics.** We show the performance of the highest performing and lowest performing subsets within each characteristic (occlusion/truncation (occ-trn), bounding box area or object size (size), aspect ratio (asp), visible sides (side) and part visibility (part)). Overall accuracy is indicated by the black dashed line. The difference between max and min indicates sensitivity; the difference between max and overall indicates the impact.

Figure 128 shows an analysis of the influence of the overlap criterion considering the AVP metric. For this overlap criterion analysis we follow the PASCAL VOC formulation: to be considered a true positive, the area of overlap between the predicted BB and GT BB must exceed a threshold.

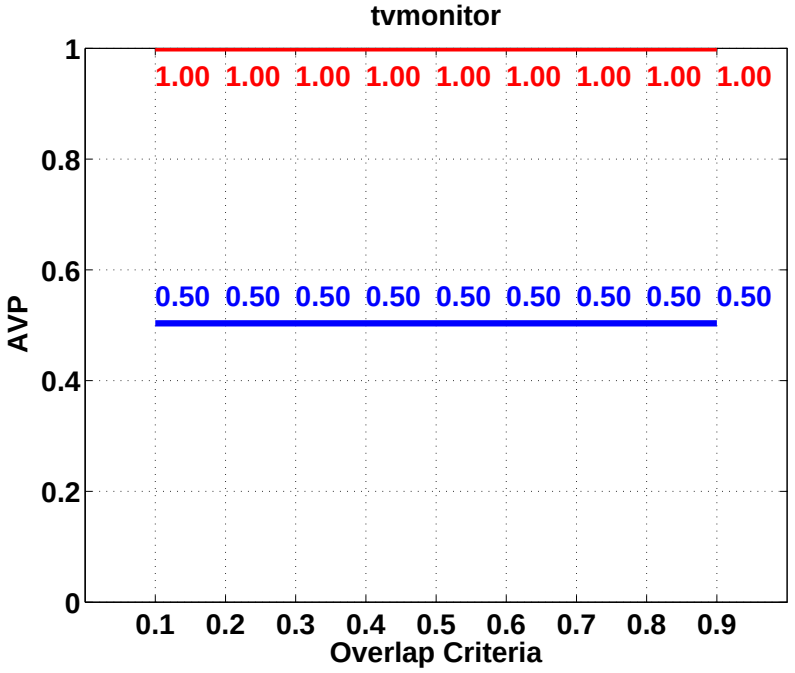


Figure 128: **Simultaneous Object Detection and Pose Estimation.** The detection performance (AP) is represented in red and the pose estimation performance (AVP) in blue.