Here, we provide some supplementary materials for "W2F: A Weakly-Supervised to Fully-Supervised Framework for Object Detection".

**Performance of Fast/Faster-RCNN with strong annotations.** Table 1 shows the performance of Fast-RCNN and Faster-RCNN with strong annotations on the PASCAL VOC 2007 test set.

| | Ours | Fast-RCNN | Faster-RCNN |
|---|---|---|---|
| mAP | 52.4 | 68.7 | 69.9 |

Table 1. The performance of Fast/Faster-RCNN with strong annotations and our method on PASCAL VOC 2007 test set.

**Per-class results.** We also provide detailed per-class average correct location (CorLoc) on VOC 2007 $trainval$ set, as shown in Table 2. Table 3 and and Table 4 show the per-class mAP and CorLoc on VOC2012, respectively.

| Method | aero | bike | bird | boat | bottle | bus | car | cat | chair | cow | table | dog | horse | mbike | person | plant | sheep | sofa | train | tv | Avg. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cinbis *et al*. 2017 | 57.2 | 62.2 | 50.9 | 37.9 | 23.9 | 64.8 | 74.4 | 24.8 | 29.7 | 64.1 | 40.8 | 37.3 | 55.6 | 68.1 | 25.5 | 38.5 | 65.2 | 35.8 | 56.6 | 33.5 | 47.3 |
| Bilen *et al*. 2015 | 66.4 | 59.3 | 42.7 | 20.4 | 21.3 | 63.4 | 74.3 | 59.6 | 21.1 | 58.2 | 14.0 | 38.5 | 49.5 | 60.0 | 19.8 | 39.2 | 41.7 | 30.1 | 50.2 | 44.1 | 43.7 |
| Wang *et al*. 2014 | 80.1 | 63.9 | 51.5 | 14.9 | 21.0 | 55.7 | 74.2 | 43.5 | 26.2 | 53.4 | 16.3 | 56.7 | 58.3 | 69.5 | 14.1 | 38.3 | 58.8 | 47.2 | 49.1 | 60.9 | 48.5 |
| Kantorov *et al*. 2016 | 83.3 | 68.6 | 54.7 | 23.4 | 18.3 | 73.6 | 74.1 | 54.1 | 8.6 | 65.1 | 47.1 | 59.5 | 67.0 | 83.5 | 35.3 | 39.9 | 67.0 | 49.7 | 63.5 | 65.2 | 55.1 |
| Bilen *et al*. 2016[†] | 46.4 | 58.3 | 35.5 | 25.9 | 14.0 | 66.7 | 53.0 | 39.2 | 8.9 | 41.8 | 26.6 | 38.6 | 44.7 | 59.0 | 10.8 | 17.3 | 40.7 | 49.6 | 56.9 | 50.8 | 39.3 |
| Li *et al*. 2016 | 78.2 | 67.1 | 61.8 | 38.1 | 36.1 | 61.8 | 78.8 | 55.2 | 28.5 | 68.8 | 18.5 | 49.2 | 64.1 | 73.5 | 21.4 | 47.4 | 64.6 | 22.3 | 60.9 | 52.3 | 52.4 |
| Tang *et al*. 2017(OICR) | 81.7 | 80.4 | 48.7 | 49.5 | 32.8 | 81.7 | 85.4 | 40.1 | 40.6 | 79.5 | 35.7 | 33.7 | 60.5 | 88.8 | 21.8 | 57.9 | 76.3 | 59.9 | 75.3 | **81.4** | 60.6 |
| Jie *et al*. 2017 | 72.7 | 55.3 | 53.0 | 27.8 | 35.2 | 68.6 | 81.9 | 60.7 | 11.6 | 71.6 | 29.7 | 54.3 | 64.3 | 88.2 | 22.2 | 53.7 | 72.2 | 52.6 | 68.9 | 75.5 | 56.1 |
| Krishna *et al*. 2016 | 58.8 | - | 49.6 | 15.4 | - | - | 64.9 | 59.0 | - | 43.2 | - | 51.2 | 57.5 | 63.1 | - | - | - | - | 54.4 | - | 51.7 |
| Tang *et al*. 2017[†] | 85.8 | 82.7 | 62.8 | 45.2 | **43.5** | 84.8 | **87.0** | 46.8 | 15.7 | 82.2 | **51.0** | 45.6 | **83.7** | 91.2 | 22.2 | 59.7 | 75.3 | 65.1 | 76.8 | 78.1 | 64.3 |
| WSD | 86.7 | 83.1 | 62.2 | 57.5 | 28.6 | 82.7 | 83.4 | 36.6 | 39.7 | 80.1 | 39.9 | 28.4 | 59.5 | 88.0 | 15.6 | 55.3 | 82.5 | 63.7 | 76.1 | 79.2 | 61.4 |
| WSD+FSD1 | **88.8** | 84.3 | **68.5** | **59.6** | 37.4 | 85.3 | 85.4 | 40.4 | **45.6** | 81.5 | 45.6 | 32.6 | 66.3 | **92.0** | 16.9 | **60.4** | 84.5 | 67.7 | 77.6 | 79.6 | 65.0 |
| WSD+PGE+FSD1 | 85.4 | **88.6** | 64.6 | 53.2 | 36.3 | **85.8** | 86.3 | 82.0 | 39.7 | 81.5 | 46.8 | 74.4 | 74.2 | 89.6 | 42.9 | 52.8 | 80.4 | **70.2** | 73.4 | 80.7 | 69.4 |
| WSD+PGE+PGA+FSD2 | 85.4 | 87.5 | 62.5 | 54.3 | 35.5 | 85.3 | 86.6 | **82.3** | 39.7 | **82.9** | 49.4 | **76.5** | 74.8 | 90.0 | **46.8** | 53.9 | **84.5** | 68.3 | **79.1** | 79.9 | **70.3** |

Table 2. Correct localization (CorLoc) (%) of our method and other state-of-the-art methods on the PASCAL VOC 2007 $test$ set. The † denotes the results of combining multiple models, others are the results of using single model. FSD1 means Fast-RCNN, and FSD2 represents Faster-RCNN. The weakly-supervised detectors in the top part are based on MIL learning, and the methods in the middle part are similar to our framework (*i.e*. using pseudo ground-truths to train a fully-supervised detector).

| Method | aero | bike | bird | boat | bottle | bus | car | cat | chair | cow | table | dog | horse | mbike | person | plant | sheep | sofa | train | tv | mAP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Kantorov *et al*. 2016 | 64.0 | 54.9 | 36.4 | 8.1 | 12.6 | 53.1 | 40.5 | 28.4 | 6.6 | 35.3 | **34.4** | 49.1 | 42.6 | 62.4 | 19.8 | 15.2 | 27.0 | 33.1 | 33.0 | 50.0 | 35.3 |
| Tang *et al*. 2017(OICR) | 67.7 | 61.2 | 41.5 | 25.6 | 22.2 | 54.6 | 49.7 | 25.4 | 19.9 | 47.0 | 18.1 | 26.0 | 38.9 | 67.7 | 2.0 | 22.6 | 41.1 | 34.3 | 37.9 | 55.3 | 37.9 |
| Jie *et al*. 2017 | 60.8 | 54.2 | 34.1 | 14.9 | 13.1 | 54.3 | 53.4 | 58.6 | 3.7 | 53.1 | 8.3 | 43.4 | 49.8 | 69.2 | 4.1 | 17.5 | 43.8 | 25.6 | **55.0** | 50.1 | 38.3 |
| Tang *et al*. 2017[†] | 71.4 | 69.4 | **55.1** | 29.8 | 28.1 | 55.0 | 57.9 | 24.4 | 17.2 | **59.1** | 21.8 | 26.6 | 57.8 | 1.3 | 1.0 | 23.1 | **52.7** | 37.5 | 33.5 | 56.6 | 42.5 |
| WSD[†] | 70.0 | 63.3 | 43.0 | 28.0 | 25.4 | 54.1 | 52.5 | 19.8 | 16.1 | 48.6 | 14.3 | 29.9 | 49.9 | 70.2 | **23.4** | **25.3** | 42.4 | 39.1 | 41.5 | 56.7 | 39.6 |
| WSD+FSD1[‡] | 72.3 | 70.3 | 51.8 | **32.4** | 27.5 | 58.6 | 58.7 | 17.6 | 13.3 | 58.1 | 14.0 | 29.5 | 62.2 | **74.3** | 1.2 | 21.6 | 47.6 | 45.9 | 32.6 | **58.1** | 42.4 |
| WSD+PGE+FSD1[§] | 71.5 | **71.0** | 46.6 | 27.6 | 26.6 | 58.1 | **59.1** | **62.1** | 19.4 | 59.0 | 8.9 | **71.4** | 64.1 | 74.2 | 6.7 | 23.6 | 47.4 | 45.2 | 44.9 | 57.5 | 47.3 |
| WSD+PGE+PGA+FSD2[¶] | **73.0** | 69.4 | 45.8 | 30.0 | **28.7** | **58.8** | 58.6 | 56.7 | **20.5** | 58.9 | 10.0 | 69.5 | **67.0** | 73.4 | 7.4 | 24.6 | 48.2 | **46.8** | 50.7 | 58.0 | **47.8** |

Table 3. Average precision(AP) (%) of our method and other state-of-the-art methods on the PASCAL VOC 2012 $test$ set. †, FSD1 and FSD2 have the same meanings as Table1. The weakly-supervised detectors in the top part are based on MIL learning, and the methods in the middle part are similar to our framework (*i.e*. using pseudo ground-truths to train a fully-supervised detector). [†] `http://host.robots.ox.ac.uk:8080/anonymous/6UBIHR.html`, [‡] `http://host.robots.ox.ac.uk:8080/anonymous/YXCMZ7.html`, [§] `http://host.robots.ox.ac.uk:8080/anonymous/3DXIHR.html`, [¶] `http://host.robots.ox.ac.uk:8080/anonymous/CHJKOG.html`

| Method | aero | bike | bird | boat | bottle | bus | car | cat | chair | cow | table | dog | horse | mbike | person | plant | sheep | sofa | train | tv | Avg. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Kantorov *et al*. 2016 | 78.3 | 70.8 | 52.5 | 34.7 | 36.6 | 80.0 | 58.7 | 38.6 | 27.7 | 71.2 | 32.3 | 48.7 | 76.2 | 77.4 | 16.0 | 48.4 | 69.9 | 47.5 | 66.9 | 62.9 | 54.8 |
| Tang *et al*. 2017(OICR) | 86.2 | 84.2 | 68.7 | 55.4 | 46.5 | 82.8 | 74.9 | 32.2 | 46.7 | 82.8 | 42.9 | 41.0 | 68.1 | 89.6 | 9.2 | 53.9 | 81.0 | 52.9 | 59.5 | 83.2 | 62.1 |
| Jie *et al*. 2017 | 82.4 | 68.1 | 54.5 | 38.9 | 35.9 | 84.7 | 73.1 | 4.8 | 17.1 | 78.3 | 22.5 | 57.0 | 70.8 | 86.6 | 18.7 | 49.7 | 80.7 | 45.3 | 70.1 | 77.3 | 58.8 |
| Tang *et al*. 2017[†] | 89.3 | 86.3 | **75.2** | 57.9 | 53.5 | 84.0 | 79.5 | 35.2 | **47.2** | 87.4 | **43.4** | 43.8 | 77.0 | 91.0 | 10.4 | 60.7 | **86.8** | 55.7 | 62.0 | **84.7** | 65.6 |
| WSD | 87.0 | 83.2 | 69.0 | 56.6 | 50.5 | 84.4 | 75.8 | 28.0 | 41.9 | 85.1 | 37.3 | 43.6 | 77.2 | 89.2 | 11.2 | 55.8 | 80.7 | 59.0 | 62.6 | 82.5 | 63.0 |
| WSD+FSD1 | **89.5** | 86.2 | 73.9 | **58.3** | 54.2 | **89.3** | 78.2 | 30.4 | 42.8 | **87.4** | 37.1 | 45.8 | 81.8 | **92.2** | 11.7 | **61.1** | 83.4 | 61.6 | 61.8 | 83.7 | 65.5 |
| WSD+PGE+FSD1 | 88.3 | **86.3** | 65.3 | 55.6 | 52.5 | 88.8 | **79.8** | **70.1** | 44.0 | 86.1 | 26.7 | **79.7** | **87.6** | 91.4 | 26.0 | 56.7 | 85.0 | 61.9 | 62.9 | 84.4 | 69.0 |
| WSD+PGE+PGA+FSD2 | 88.8 | 85.8 | 64.9 | 56.0 | **54.3** | 88.1 | 79.1 | 67.8 | 46.5 | 86.1 | 26.7 | 77.7 | 87.2 | 89.7 | **28.5** | 56.9 | 85.6 | **63.7** | **71.3** | 83.0 | **69.4** |

Table 4. Correct localization (CorLoc) (%) of our method and other state-of-the-art methods on the PASCAL VOC 2012 $trainval$ set. †, FSD1 and FSD2 have the same meanings as Table1. The weakly-supervised detectors in the top part are based on MIL learning, and the methods in the middle part are similar to our framework (*i.e*. using pseudo ground-truths to train a fully-supervised detector).