Training My Car to See Using Virtual Worlds

Dr. Antonio M. López

www.cvc.uab.es/~antonio

April 5th, 2016
Research: Vision for ADAS & AD

E.g. Pedestrian Detection
Research: Vision for ADAS & AD

E.g. Pedestrian Detection

Cascade

Random Forest of Local Experts

Rejected

Rejected

Rejected

Pedestrian

Local expert
(HOG+LBP / LinSVM)

Internal Node

1242×375 pixels, 12 scales

NVIDIA DRIVE PX

“GPU-based pedestrian detection for autonomous driving”, GTC 2016.
Research: Vision for ADAS & AD

E.g. Semantic Segmentation (pixel-wise classification)
Research: Vision for ADAS & AD

E.g. Semantic Segmentation (pixel-wise classification)

Deep Convolutional Neural Networks
Research: Vision for ADAS & AD

E.g. Semantic Segmentation (pixel-wise classification)
Among other tasks: **object detection** and **semantic segmentation**, e.g., using Computer Vision.
Machine Learning Needs Annotated Data

Manual Annotation
Machine Learning Needs Annotated Data
Machine Learning Needs
Annotated Data
Virtual Worlds for Automatic Annotations
Virtual Worlds for Automatic Annotations
Virtual Worlds for Automatic Annotations
Virtual Worlds for Automatic Annotations
Virtual Worlds for Automatic Annotations

“Close the circle between modern Computer Animation and Computer Vision”
Cool World: Domain Adaptation of Virtual and Real Worlds

Domain shift: training and testing data follow different probability distributions

Domain adaptation: source knowledge (data or models) & target data with few or no manual annotations.
Cool World: Domain Adaptation of Virtual and Real Worlds

Domain shift: training and testing data follow different probability distributions

Source domain: Virtual world

Target domain: Real world

Domain adaptation: source knowledge (data or models) & target data with few or no manual annotations.
Cool World: Domain Adaptation of Virtual and Real Worlds

Realistic Virtual World & Virtual Samples → Prior Model

Domain Adaptation

Self-paced Learning

Human Oracle

Detection (target domain)

Detections by the source model at FPPI = 0.1.

Detections by the adaptive model at FPPI = 0.1.

Training Data (target domain)
Virtual Worlds for Automatic Annotations

RGB

Semantic Segmentation

Depth Map

Powered By unity
Virtual Worlds for Automatic Annotations
SYNTHIA: Synthetic Imagery with Annotations

<table>
<thead>
<tr>
<th>Dataset</th>
<th># Frames (A)</th>
<th># Training (T)</th>
<th># Validation (V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CamVid [5, 6]</td>
<td>701</td>
<td>300</td>
<td>401</td>
</tr>
<tr>
<td>KITTI [11, 19, 29]</td>
<td>547</td>
<td>200</td>
<td>347</td>
</tr>
<tr>
<td>Urban LabelMe [31, 1]</td>
<td>942</td>
<td>200</td>
<td>742</td>
</tr>
<tr>
<td>CBCL StreetScenes [4, 1]</td>
<td>3547</td>
<td>200</td>
<td>3347</td>
</tr>
<tr>
<td>SYNT HIA-Rand</td>
<td>13,400</td>
<td>13,400</td>
<td>0</td>
</tr>
</tbody>
</table>
SYNTHIA: Synthetic Imagery with Annotations

<table>
<thead>
<tr>
<th>Training</th>
<th>Validation</th>
<th>sky</th>
<th>building</th>
<th>road</th>
<th>sidewalk</th>
<th>fence</th>
<th>vegetation</th>
<th>pole</th>
<th>car</th>
<th>sign</th>
<th>pedestrian</th>
<th>cyclist</th>
<th>per-class</th>
<th>global</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camvid (T)</td>
<td>CamVid (V)</td>
<td>99</td>
<td>65</td>
<td>95</td>
<td>52</td>
<td>7</td>
<td>79</td>
<td>5</td>
<td>80</td>
<td>3</td>
<td>26</td>
<td>6</td>
<td>46.3</td>
<td>81.9</td>
</tr>
<tr>
<td>Camvid (T) + SYNTHIA-Rand (A)</td>
<td>CamVid (V)</td>
<td>98</td>
<td>90</td>
<td>91</td>
<td>63</td>
<td>5</td>
<td>83</td>
<td>9</td>
<td>94</td>
<td>0</td>
<td>58</td>
<td>31</td>
<td>56.5 (10.2)</td>
<td>90.7 (8.8)</td>
</tr>
<tr>
<td>KITTI (T)</td>
<td>KITTI (V)</td>
<td>79</td>
<td>83</td>
<td>87</td>
<td>73</td>
<td>0</td>
<td>85</td>
<td>0</td>
<td>69</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>44.2</td>
<td>80.5</td>
</tr>
<tr>
<td>KITTI (T) + SYNTHIA-Rand (A)</td>
<td>KITTI (V)</td>
<td>89</td>
<td>86</td>
<td>90</td>
<td>58</td>
<td>0</td>
<td>72</td>
<td>0</td>
<td>76</td>
<td>0</td>
<td>66</td>
<td>29</td>
<td>51.6 (7.4)</td>
<td>80.8 (0.3)</td>
</tr>
<tr>
<td>U-LabelMe (T)</td>
<td>U-LabelMe (V)</td>
<td>72</td>
<td>80</td>
<td>75</td>
<td>45</td>
<td>0</td>
<td>62</td>
<td>2</td>
<td>53</td>
<td>0</td>
<td>14</td>
<td>2</td>
<td>36.4</td>
<td>62.4</td>
</tr>
<tr>
<td>U-LabelMe (T) + SYNTHIA-Rand (A)</td>
<td>U-LabelMe (V)</td>
<td>69</td>
<td>77</td>
<td>93</td>
<td>33</td>
<td>0</td>
<td>62</td>
<td>11</td>
<td>77</td>
<td>1</td>
<td>67</td>
<td>24</td>
<td>46.7 (10.3)</td>
<td>72.1 (9.7)</td>
</tr>
<tr>
<td>CBCL (T)</td>
<td>CBCL (V)</td>
<td>62</td>
<td>77</td>
<td>86</td>
<td>41</td>
<td>0</td>
<td>74</td>
<td>5</td>
<td>63</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>37.9</td>
<td>73.9</td>
</tr>
<tr>
<td>CBCL (T) + SYNTHIA-Rand (A)</td>
<td>CBCL (V)</td>
<td>72</td>
<td>82</td>
<td>90</td>
<td>39</td>
<td>0</td>
<td>58</td>
<td>26</td>
<td>70</td>
<td>5</td>
<td>52</td>
<td>39</td>
<td>48.4 (10.5)</td>
<td>75.2 (1.3)</td>
</tr>
</tbody>
</table>
Virtual Worlds for Automatic Annotations
Virtual Worlds for Simulation
Virtual Worlds for Debugging
Conclusion
Dr. Antonio M. López, Principal Investigator UAB & CVC ADAS Group

Address
Edifici O, Campus UAB
08193 Bellaterra
Barcelona

Phone & Fax
Direct Line: +34 93 581 2561
Fax: +34 93 581 1670
www.cvc.uab.cat

E-contact
www.cvc.uab.cat/~antonio
antonio@cvc.uab.cat

Thanks to:
- Spanish DGT SPIP2014-01352.
- Spanish MEC TRA2014-57088-C2-1-R.

Of course, many thanks also to:

Many Thanks!!!