IMPROVING OBJECT RECOGNITION IN AERIAL IMAGE AND AMBULATORY ASSESSMENT ANALYSIS BY DEEP LEARNING

Department of Electrical Engineering and Computer Science

Speaker: Peng Sun
Advisor: Yi Shang
Outline

• Introduction

• Proposed Method
  – Novel Adaptive Saliency Biased Loss (ASBL) for Object Detection in Aerial Images
  – Improving Bird Recognition in Aerial Images using Deep Learning
  – A new Deep Learning Based Automatic Detection of Alcohol usage (DEEP ADA)

• Summary
Introduction

Sensor Data

- Physiology Sensor
- Speech Sensor
- Gps Sensor
- Remote Sensor

- Radar
- Lidar
- Aerial Image
Our Research Focus on

Physiology Sensor Data

Ambulatory Assessment

Object Detection in Aerial Images

Aerial Image Sensor
Main Contribution

<table>
<thead>
<tr>
<th>Problem</th>
<th>Our Proposed Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Object Detection in Aerial Images</td>
<td>Novel Adaptive Saliency Biased Loss (ASBL)</td>
</tr>
<tr>
<td>Bird Detection in Aerial Images</td>
<td>DNN Object Detection Model Adaption and Analysis</td>
</tr>
<tr>
<td>Ambulatory Assessment Analysis</td>
<td>Deep Learning Based Automatic Detection of Alcohol usage (DEEP ADA)</td>
</tr>
</tbody>
</table>
Outline

• Introduction

• Proposed Method
  – Novel Adaptive Saliency Biased Loss (ASBL) for Object Detection in Aerial Images
  – Improving Bird Recognition in Aerial Images using Deep Learning
  – A new Deep Learning Based Automatic Detection of Alcohol usage (DEEP ADA)

• Summary
Novel Adaptive Saliency Biased Loss for Object Detection in Aerial Images
Content

• Introduction
• Related Work
• Theory Foundation
• Methodology
  – Image based saliency biased loss
  – Anchor based saliency biased loss
• Experimental Result
• Conclusion
Introduction

- Aerial images characteristics:
  - Multi-scale
  - Multi-angle
  - Crowded
  - Background variant
- Limitation of current DNN detector
  - Face, Human, etc
## Related Work

<table>
<thead>
<tr>
<th>General DNN Object Detection</th>
<th>Authors</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rdemon, 2017</td>
<td>YOLO DNN Object Detector</td>
</tr>
<tr>
<td></td>
<td>Liu, 2016</td>
<td>Single Shot Multibox Detector</td>
</tr>
<tr>
<td></td>
<td>Lin, 2017</td>
<td>RetinaNet</td>
</tr>
<tr>
<td>Aerial Images DNN Object Detection</td>
<td>Han, 2014</td>
<td>WSL and feature learning</td>
</tr>
<tr>
<td></td>
<td>Li, 2017</td>
<td>Rotation-insensitive using RPN and local-contextual feature fusion network</td>
</tr>
<tr>
<td></td>
<td>Cheng, 2019</td>
<td>Rotation-Invariant and Fisher Discriminative (RIFD)</td>
</tr>
</tbody>
</table>
Theory Foundation (RetinaNet)

Feature Pyramid Network

Lateral connection

Theory Foundation (Cont.)

- **Focal loss**

\[ CE(p, y) = \begin{cases} -\log(p) & \text{if } y = 1 \\ -\log(1 - p) & \text{otherwise} \end{cases} \]

\[ \alpha_t = \begin{cases} \alpha, & \text{if } y = 1 \\ 1 - \alpha, & \text{if } y = 0 \end{cases} \]

\[ p_t = \begin{cases} p, & \text{if } y = 1 \\ 1 - p, & \text{otherwise} \end{cases} \]

\[ FL(p, y) = \alpha_t * (1 - p_t)^\gamma * CE(p, y) \]

- **Anchor box**

- **Saliency map:**

  - In computer vision, a saliency map of an image is a value array representing each pixel's importance.
Content

• Introduction
• Related Work
• Theory Foundation
• **Methodology**
  – Image based saliency biased loss
  – Anchor based saliency biased loss
• Experimental Result
• Conclusion
Adaptive Saliency Biased Loss

Anchor Loss

Image Loss

Adaptive Saliency Biased Loss

Adaptive
Adaptive weighting anchor’s loss

Saliency
Saliency map as weighting factor of Loss

Biased
Weighting Loss in biased way

Loss
Novel Objective Loss function

Image Based ASBL

Anchor Based ASBL
Image-based ASBL
Saliency estimator

- Pretrained DNN:
  - Pretrained by general objects.
  - Capability to extract basic shape features

- Use saliency map to define complexity of image
  - Activated cell may have foreground prediction
  - More activated cell = more complex
Formula of Image ASBL

- Normalize

\[ S_I'(x) = \frac{S_I(x) - S_{\text{min}}}{S_{\text{max}} - S_{\text{min}}} (S_{ub} - S_{lb}) + S_{lb} \]

- \( S_{\text{max}}, S_{\text{min}} \) is calculated by training data; \( S_{ub} \) set as 1.

- Image-based

\[ ASBL_I(x, p, y) = S_I'(x) \times FL(p, y) \]
Image Complexity

Low Level Conv Features

High Level Conv Features

Low $S_I$  ➔  High $S_I$
Anchor-based ASBL
Formula of Anchor ASBL

S: Saliency information; 
A: Positive Anchors; 
I: Images 

\[ Pr(S|A, I) \propto Pr(A|S, I) \times Pr(S|I) \]

\[ \frac{1}{C} \sum_{c=1}^{C} R_{c,u,v}(x) \times \frac{1}{C} \sum_{c=1}^{C} f_{c,u,v}(x) \]

\[ SA_{u,v}(x) \propto \frac{1}{C} \sum_{c=1}^{C} R_{c,u,v}(x) \odot \frac{1}{C} \sum_{c=1}^{C} f_{c,u,v}(x) \]

\[ \frac{1}{C} \sum_{c=1}^{C} R_{c,u,v}(x) \]

Adaptive updated during training

• Anchor ASBL

\[ ASBL_{A}(p, y) = \sum_{a=1}^{A} \sum_{v=1}^{H} \sum_{u=1}^{W} SA_{u,v}(x) \times FL_{u,v,a}(p, y) \]

A is # of Anchors, H: height, W: width
ASBL RetinaNet

- Trained ResNet50 with 2 more conv-block using ImageNet
- Image-based ASBL on RetinaNet
- Half numbers of Epoch
- Anchor-based ASBL on RetinaNet
Content

• Introduction
• Related Work
• Theory Foundation
• Methodology
  – Image based saliency biased loss
  – Anchor based saliency biased loss
• Experimental Result
• Conclusion
Data

DOTA: A Large-scale Dataset for Object Detection in Aerial Images

- **2806** aerial images
- **188, 282** Instances
- **15** categories
- **1024 **× **1024** crop images
- **1/2** training; **1/6** validation; **1/3** test

NWPU VHR-10

- **10** Categories
- **400 **× **400** crop images
- **679** training; **200** validation; **293** test

---


Image-based ASBL Ablation Study

- Saliency map normalization
- How big range for normalization?
  - Too big -> not enough training for easy case
  - Too narrow -> raw Focal loss

- Multi-scale saliency experiments
  - Experiments on C2, C3, C4, C5 saliency map

<table>
<thead>
<tr>
<th>$S_{lb}$</th>
<th>Normalization</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>Y</td>
<td>61.79</td>
</tr>
<tr>
<td>0.5</td>
<td>Y</td>
<td>63.48</td>
</tr>
<tr>
<td>0.7</td>
<td>Y</td>
<td>62.4</td>
</tr>
<tr>
<td>-</td>
<td>N</td>
<td>62.86</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Conv Block Layer of ResNet50</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
</tr>
</thead>
<tbody>
<tr>
<td>mAP</td>
<td>64.77</td>
<td>64.51</td>
<td>63.32</td>
<td>63.48</td>
</tr>
</tbody>
</table>
Anchor-based ASBL Ablation Study

Start from Image-based ASBL

- Adaptive update?
  - 64.82 -> 65.53

- Normalization?
  - 65.53 -> 66.12

<table>
<thead>
<tr>
<th>Method</th>
<th>Dynamic update</th>
<th>Normalization</th>
<th>$S_{lb}$</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ASBL_A$</td>
<td>Y</td>
<td>Y</td>
<td>0.3</td>
<td>65.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.5</td>
<td><strong>66.12</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.7</td>
<td>65.58</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>-</td>
<td></td>
<td>65.53</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>Y</td>
<td>0.5</td>
<td>64.82</td>
</tr>
<tr>
<td>$ASBL_I$</td>
<td>-</td>
<td>Y</td>
<td>0.5</td>
<td>64.77</td>
</tr>
</tbody>
</table>
## Performance on DOTA

<table>
<thead>
<tr>
<th>Method</th>
<th>Data</th>
<th>Plane</th>
<th>BD</th>
<th>Bridge</th>
<th>GTF</th>
<th>SV</th>
<th>LV</th>
<th>Ship</th>
<th>TC</th>
<th>BC</th>
<th>ST</th>
<th>SBF</th>
<th>RA</th>
<th>Harbor</th>
<th>SP</th>
<th>HC</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>YOLO [8]</td>
<td>T+V</td>
<td>76.9</td>
<td>33.87</td>
<td>22.73</td>
<td>34.88</td>
<td>38.73</td>
<td>32.02</td>
<td>52.37</td>
<td>61.65</td>
<td>48.54</td>
<td>33.91</td>
<td>29.27</td>
<td>36.83</td>
<td>36.44</td>
<td>38.26</td>
<td>11.61</td>
<td>39.2</td>
</tr>
<tr>
<td>RFCN [9]</td>
<td>T+V</td>
<td>79.33</td>
<td>44.26</td>
<td>36.58</td>
<td>53.53</td>
<td>39.38</td>
<td>34.15</td>
<td>47.29</td>
<td>45.66</td>
<td>47.74</td>
<td>65.84</td>
<td>37.92</td>
<td>44.23</td>
<td>47.23</td>
<td>50.64</td>
<td>34.9</td>
<td>47.24</td>
</tr>
<tr>
<td>Faster RCNN [5]</td>
<td>T+V</td>
<td>80.32</td>
<td>77.55</td>
<td>32.86</td>
<td>68.13</td>
<td>53.66</td>
<td>52.49</td>
<td>50.04</td>
<td>90.41</td>
<td>75.05</td>
<td>59.59</td>
<td>57</td>
<td>49.81</td>
<td>61.69</td>
<td>56.46</td>
<td>41.85</td>
<td>60.46</td>
</tr>
<tr>
<td>RetinaNet [18]</td>
<td>T</td>
<td>78.22</td>
<td>53.41</td>
<td>26.38</td>
<td>42.27</td>
<td>63.64</td>
<td>52.63</td>
<td>73.19</td>
<td>87.17</td>
<td>44.64</td>
<td>57.99</td>
<td>18.03</td>
<td>51</td>
<td>43.39</td>
<td>56.56</td>
<td>7.44</td>
<td>50.39</td>
</tr>
<tr>
<td>RetinaNet*</td>
<td>T</td>
<td>89.03</td>
<td>62.14</td>
<td>43.88</td>
<td>47.05</td>
<td>73.57</td>
<td>65.18</td>
<td>78.65</td>
<td>90.86</td>
<td>66.28</td>
<td>70.26</td>
<td>35.07</td>
<td>58.26</td>
<td>68.93</td>
<td>66.34</td>
<td>22.16</td>
<td>62.51</td>
</tr>
<tr>
<td>ASBL-RetinaNet</td>
<td>T</td>
<td>89.09</td>
<td>67.96</td>
<td>46.38</td>
<td>57.12</td>
<td>73.55</td>
<td>66.19</td>
<td>78.67</td>
<td>90.86</td>
<td>71</td>
<td>73.88</td>
<td>45.15</td>
<td>60.92</td>
<td>70.01</td>
<td>68.51</td>
<td>32.49</td>
<td>66.12</td>
</tr>
<tr>
<td>ASBL-RetinaNet</td>
<td>T+V</td>
<td><strong>89.51</strong></td>
<td>74.07</td>
<td><strong>46.91</strong></td>
<td>55.54</td>
<td><strong>73.78</strong></td>
<td><strong>66.87</strong></td>
<td>78.48</td>
<td><strong>90.86</strong></td>
<td>70.09</td>
<td>73.2</td>
<td>46.71</td>
<td><strong>61.34</strong></td>
<td><strong>70.5</strong></td>
<td><strong>72.17</strong></td>
<td>32.84</td>
<td><strong>66.86</strong></td>
</tr>
</tbody>
</table>

* RetinaNet with modified anchor sizes and ratios

- 10 out of 15 outperform Faster RCNN
- 14 out of 15 outperform modified RetinaNet
## Performance on NWPU VHR10

<table>
<thead>
<tr>
<th>Models</th>
<th>Average running time per images (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>COPD [41]</td>
<td>1.16</td>
</tr>
<tr>
<td>Transferred CNN [1]</td>
<td>5.09</td>
</tr>
<tr>
<td>RICNN [29]</td>
<td>8.47</td>
</tr>
<tr>
<td>Faster RCNN [5]</td>
<td>0.09</td>
</tr>
<tr>
<td>Li etc [27]</td>
<td>2.89</td>
</tr>
<tr>
<td>RetinaNet* [18]</td>
<td>0.045</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Airplane</th>
<th>Ship</th>
<th>Storage tank</th>
<th>Baseball diamond</th>
<th>Tennis court</th>
<th>Basketball court</th>
<th>Ground track field</th>
<th>Harbor</th>
<th>Bridge</th>
<th>Vehicle</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>COPD [41]</td>
<td>62.25</td>
<td>69.37</td>
<td>64.52</td>
<td>82.13</td>
<td>34.13</td>
<td>35.25</td>
<td>84.21</td>
<td>56.31</td>
<td>16.43</td>
<td>44.28</td>
</tr>
<tr>
<td>Transferred CNN [1]</td>
<td>66.03</td>
<td>57.13</td>
<td>85.01</td>
<td>80.93</td>
<td>35.11</td>
<td>45.52</td>
<td>79.37</td>
<td>62.57</td>
<td>43.17</td>
<td>41.27</td>
</tr>
<tr>
<td>RICNN [29]</td>
<td>88.71</td>
<td>78.34</td>
<td>86.33</td>
<td>89.09</td>
<td>42.33</td>
<td>56.85</td>
<td>87.72</td>
<td>67.47</td>
<td>62.31</td>
<td>72.01</td>
</tr>
<tr>
<td>Faster RCNN [5]</td>
<td>90.9</td>
<td>86.3</td>
<td>90.53</td>
<td><strong>98.24</strong></td>
<td>89.72</td>
<td>69.64</td>
<td>100</td>
<td>80.11</td>
<td>61.49</td>
<td>78.14</td>
</tr>
<tr>
<td>Li etc [27]</td>
<td>99.70</td>
<td>90.8</td>
<td>90.61</td>
<td>92.91</td>
<td><strong>90.29</strong></td>
<td>80.13</td>
<td>90.81</td>
<td>80.29</td>
<td>68.53</td>
<td><strong>87.14</strong></td>
</tr>
<tr>
<td>RetinaNet* [18]</td>
<td>96.58</td>
<td>83.61</td>
<td>74.76</td>
<td>84.32</td>
<td>63.99</td>
<td>59.66</td>
<td>98.33</td>
<td>62.83</td>
<td>65.72</td>
<td>78.31</td>
</tr>
<tr>
<td>ASBL-RetinaNet (VGG16)</td>
<td><strong>100</strong></td>
<td>91.27</td>
<td><strong>96.76</strong></td>
<td>96.69</td>
<td>68.54</td>
<td><strong>87.67</strong></td>
<td>100</td>
<td>77.10</td>
<td><strong>83.92</strong></td>
<td>83.45</td>
</tr>
<tr>
<td>ASBL-RetinaNet (ResNet50)</td>
<td>99.34</td>
<td><strong>93.19</strong></td>
<td>94.36</td>
<td>97.70</td>
<td>71.19</td>
<td>84.67</td>
<td><strong>100</strong></td>
<td><strong>87.61</strong></td>
<td>81.49</td>
<td>83.50</td>
</tr>
</tbody>
</table>
Visualization Comparison

- Top: RetinaNet; Bottom: ASBL-RetinaNet
- 1\textsuperscript{st} column: crowded; 2\textsuperscript{nd} Column: high image complexity; 3\textsuperscript{rd} column: high anchor complexity; 4\textsuperscript{th} column: clear background
Visualization

NWPU VHR-10

DOTA
Content

• Introduction
• Related Work
• Theory Foundation
• Methodology
  – Image based saliency biased loss
  – Anchor based saliency biased loss
• Experimental Result
• Conclusion
Conclusion and Future Work

• Conclusion:
  – Novel adaptive salience biased loss functions to improve one-stage detector (RetinaNet) with same inference speed on aerial images
  – Two levels ASBL can outperformed other deep learning models on aerial images by 6.4 and 2.19 of mAP on DOTA and NWPU VHR-10
  – Github: https://github.com/ps793/ASBL-RetinaNet

• Future Work:
  – Scale, angles of objects can be considered into weighting factor of objective loss function
Improving Bird Recognition in Aerial Images using Deep Learning
Content

• Introduction
• Related Work
• LIBAI Dataset
• Model Theory and Adaption
• Experimental Result and Analysis
• Summary
Introduction

• How to recognize and count small objects (birds) in aerial images
• Performance of deep learning models on small object in aerial images

DNN Detector
## Related Work

<table>
<thead>
<tr>
<th>General object detection</th>
<th>Authors</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Liu, 2016</td>
<td>Single Shot Multibox Detector</td>
</tr>
<tr>
<td></td>
<td>Lin, 2017</td>
<td>RetinaNet</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Aerial images object detection</th>
<th>Authors</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Han, 2014</td>
<td>WSL and feature learning</td>
</tr>
<tr>
<td></td>
<td>Li, 2017</td>
<td>Rotation-insensitive using RPN and local-contextual feature fusion network</td>
</tr>
<tr>
<td></td>
<td>Cheng, 2019</td>
<td>Rotation-Invariant and Fisher Discriminative (RIFD)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Small object detection</th>
<th>Authors</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hu, 2017</td>
<td>Tiny face detector</td>
</tr>
<tr>
<td></td>
<td>Yang, 2018</td>
<td>DFL-CNN (double focal loss CNN) for vehicle detection</td>
</tr>
</tbody>
</table>
LIBAI Dataset

Data Collection and Label
- Collection: Illinois Natural History Survey at the University of Illinois
- Labeling: human label

Dataset
Dataset: 339 images with 5760 * 3840
Number of birds: 27,607
Bird size: 10px - 40px

Data Division
Based on complexity of Image, divide data into easy and hard
Content

• Introduction
• Related Work
• LIBAI Dataset
• **Model Theory and Adaption**
• Experimental Result and Analysis
• Summary
Proposed Solution

- Deep Learning Models
  - One-stage models: RetinaNet, YOLO v3, SSD
  - Two-stage model: Mask RCNN
  - Segmentation: U-Net
RetinaNet

- **Theory:**
  - Feature Pyramid Network
  - Lateral connection
  - Focal Loss

- **Model Adaption:**
  - Input resize: 512 -> 600
  - Anchor size: \( \{1, 2^{1/3}, 2^{2/3}\} \) -> \( \{1, 2^{0.5}, 0.3\} \)
  - Optimizer: SGD -> Adam
  - Learning rate: 1e-3 -> 1e-4

YOLO v3

- **Theory:**
  - Multi-scale feature map (FPN)
  - Darknet53: Replace fully connected to convolution, add residual connect

- **Model Adaption:**
  - Anchors sizes and ratio: Kmeans
  - Num of anchors: 9 -> 1
  - Filter size of last layer: 75 -> 18
  - Classes: 20 -> 1

SSD

- Theory:
  - Multi-scale feature maps for detection
  - Default boxes and aspect ratios
  - Hard negative mining

- Model Adaption:
  - Anchor ratio:
    - \( \{1, 2, 3, 1/2, 1/3\} \) -> \( \{1, 2, 1/2\} \)
  - Learning rate: 0.001 -> 0.0001
  - Iteration: 120,000 -> 12,000 (20 hours)
  - SGD -> Adam

Mask R-CNN

- **Theory:**
  - RoI Pooling -> RoI Align
  - Mask Prediction Branch

- **Model Adaption:**
  - Input size: 1024 -> 512
  - Backbone: ResNet101 -> ResNet50
  - Optimizer: SGD -> Adam
  - Learning rate: 1e-3 -> 1e-4

**U-Net**

- **Theory:**
  - Encoder-decoder
  - Connection between encoder and decoder

- **Model Adaption:**
  - Crop operation removal: size matches
  - Input size: 572-> 512
  - Unet shape: (1024,512, 256)
  - Initialization: VGG pretrained weight

Content

• Introduction
• Related Work
• LIBAI Dataset
• Model Theory and Adaption
• Experimental Result and Analysis
• Conclusion
Performance Comparison

• Easy Data

<table>
<thead>
<tr>
<th></th>
<th>Train</th>
<th>Val</th>
<th>Test</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>121</td>
<td>31</td>
<td>10</td>
<td>162</td>
</tr>
</tbody>
</table>

• Hard Data

<table>
<thead>
<tr>
<th></th>
<th>Train</th>
<th>Val</th>
<th>Test</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>133</td>
<td>31</td>
<td>10</td>
<td>174</td>
</tr>
</tbody>
</table>
Visualization (Easy)

<table>
<thead>
<tr>
<th>Model</th>
<th>RetinaNet</th>
</tr>
</thead>
<tbody>
<tr>
<td># of bird</td>
<td>932</td>
</tr>
<tr>
<td># of count</td>
<td>919</td>
</tr>
<tr>
<td>Precision</td>
<td>99.21%</td>
</tr>
<tr>
<td>Recall</td>
<td>92.72%</td>
</tr>
</tbody>
</table>
Visualization (Hard)

Model | RetinaNet
---|---
# of bird | 3
# of count | 37
Precision | 2.71%
Recall | 33.33%
Model Analysis

- Segmentation methods can be used to detect small object in aerial images, better than SSD and RCNN.

- SSD are poor on small object detection since the receptive field is too large.
  - Smallest receptive field for prediction: 30*30

- FPN is more robustness than multi-scale feature extraction.
  - RetinaNet and YOLO outperformed SSD.

- Mask RCNN and Yolo are sensitive to noisy and complex background data.
  - Drop down more than 0.5 on F1 score.

- RetinaNet achieve best performance on bird detection!
Summary

• Model Adaptation and Analysis based on the bird detection problem

• Propose solution for bird detection problem in aerial image
  – RetinaNet achieved 89.3 and 63.2 for F1 score on easy and hard dataset of LBAI, respectively
  – Real time detection. (0.049s/image)

• Comparison and analysis of deep learning objectors for small object detection in aerial images
A new Deep Learning Based Automatic Detection of Alcohol usage (DEEP ADA)
Content

• Introduction
• Related Work
• Deep ADA Pipeline
  – ADA (Automatic Detection of Alcohol usage) Module
  – Deep ADA (DNN Automatic Detection of Alcohol) Module
• Experimental Result
  – ADA
  – Deep ADA
• Summary
Introduction

• Physiological sensor data
  • Noisy
  • Complex env
  • Little label
• How to extract useful features from physiological sensor data?
• Prediction on user events, like alcohol events, is critical.
• How to use unlabeled data to improve classification?
## Related Work

<table>
<thead>
<tr>
<th>Authors</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Physiological sensor analysis</strong></td>
<td></td>
</tr>
<tr>
<td>Hossain, 2014</td>
<td>Cocaine intake detection</td>
</tr>
<tr>
<td>Wei, 2016</td>
<td>Time-Frequency CNN for Automatic Sleep Stage Classification</td>
</tr>
<tr>
<td>Zheng, 2017</td>
<td>Time series classification using multi-channels CNN</td>
</tr>
<tr>
<td><strong>Unsupervised Feature Learning</strong></td>
<td></td>
</tr>
<tr>
<td>Xu, 2016</td>
<td>Deep auto encoder (unsupervised feature learning) for audio tagging</td>
</tr>
<tr>
<td>Masci, 2011</td>
<td>Stacked CNN auto-encoders for hierarchical feature extraction on MNIST images</td>
</tr>
</tbody>
</table>
Content

• Introduction
• Related Work

• Deep ADA Pipeline
  – ADA (Automatic Detection of Alcohol usage) Module
  – Deep ADA (DNN Automatic Detection of Alcohol) Module

• Experimental Result
  – ADA
  – Deep ADA

• Summary
Deep ADA Pipeline

1. Sensor data cleaning
2. Sensor analysis
3. Survey analysis
4. Automatic Detection of Alcohol usage (ADA)
5. DNN feature extraction
6. ML classifier
7. DNN Automatic Detection of Alcohol (Deep ADA)
ADA Sensor Data Cleaning

• Outlier removal
  – Unreliable data removal;
  – Loess smoothing (local regression)
    \[ w_i = (1 - \frac{|x - x_i|}{d(x)}^3)^3 \]

• Smoothness of data
  – Smoothing spline
    \[ \min p \sum_i w_i (y_i - s(x_i))^2 + (1 - p) \int \left( \frac{d^2 s}{dx^2} \right)^2 dx \]
  – Data tendency
Deep ADA Module

Cleaned Physiological Sensor data

Stats Feature
- mean, standard deviation, covariance, skewness, range, root mean square, zero crossing rate, and mean crossing rate

DNN Feature
- 8 channels features from 1D CNN

ML Classifier
- naïve bayes, decision tree, RF, adaboost and SVM

ML Classifier
- SVM
- ResNet50
Novel DNN Feature Extraction
Content

• Introduction
• Related Work
• Deep ADA Pipeline
  – ADA (Automatic Detection of Alcohol usage) Module
  – Deep ADA (DNN Automatic Detection of Alcohol) Module
• Experimental Result
  – ADA
  – Deep ADA
• Summary
Survey Data Analysis of ADA

- Individual case:
  - Unbalanced nested ANOVA
  - Drinking effect is main effect
  - Time effect is the effect within drinking effect

- General case:
  - Shapiro-Wilk test + matched paired t-test/ Wilcoxon signed ranked test

<table>
<thead>
<tr>
<th>ID</th>
<th>Positive</th>
<th>Negative</th>
<th>Fear</th>
<th>Hostility</th>
<th>Sadness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1001</td>
<td>0.05</td>
<td>0.29</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>1003</td>
<td>0.88</td>
<td>0.04</td>
<td>0.00</td>
<td>0.24</td>
<td>0.51</td>
</tr>
<tr>
<td>1004</td>
<td>0.74</td>
<td>0.27</td>
<td>0.30</td>
<td>0.27</td>
<td>0.15</td>
</tr>
<tr>
<td>1005</td>
<td>0.01</td>
<td>0.62</td>
<td>0.09</td>
<td>0.20</td>
<td>0.95</td>
</tr>
<tr>
<td>1007</td>
<td>0.37</td>
<td>0.46</td>
<td>0.68</td>
<td>0.44</td>
<td>0.78</td>
</tr>
<tr>
<td>1008</td>
<td>0.06</td>
<td>0.19</td>
<td>0.89</td>
<td>0.10</td>
<td>0.17</td>
</tr>
<tr>
<td>1009</td>
<td>0.58</td>
<td>0.16</td>
<td>0.37</td>
<td>0.08</td>
<td>0.72</td>
</tr>
<tr>
<td>1010</td>
<td>0.00</td>
<td>Null</td>
<td>Null</td>
<td>0.00</td>
<td>Null</td>
</tr>
<tr>
<td>1013</td>
<td>0.15</td>
<td>0.06</td>
<td>0.12</td>
<td>0.08</td>
<td>0.27</td>
</tr>
<tr>
<td>1014</td>
<td>0.04</td>
<td>0.30</td>
<td>0.48</td>
<td>0.58</td>
<td>0.32</td>
</tr>
<tr>
<td>1017</td>
<td>0.94</td>
<td>0.84</td>
<td>0.38</td>
<td>0.99</td>
<td>0.49</td>
</tr>
<tr>
<td>1019</td>
<td>0.59</td>
<td>0.73</td>
<td>0.00</td>
<td>0.03</td>
<td>0.60</td>
</tr>
<tr>
<td>1020</td>
<td>0.17</td>
<td>0.91</td>
<td>0.28</td>
<td>0.47</td>
<td>0.71</td>
</tr>
<tr>
<td>1021</td>
<td>0.16</td>
<td>0.00</td>
<td>0.66</td>
<td>0.40</td>
<td>0.00</td>
</tr>
<tr>
<td>1022</td>
<td>0.82</td>
<td>0.07</td>
<td>0.50</td>
<td>0.83</td>
<td>0.61</td>
</tr>
<tr>
<td>1024</td>
<td>0.18</td>
<td>0.69</td>
<td>0.50</td>
<td>0.83</td>
<td>0.46</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID</th>
<th>Positive</th>
<th>Negative</th>
<th>Fear</th>
<th>Hostility</th>
<th>Sadness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1001</td>
<td>0.57</td>
<td>0.76</td>
<td>0.03</td>
<td>0.30</td>
<td>0.41</td>
</tr>
<tr>
<td>1003</td>
<td>0.30</td>
<td>0.04</td>
<td>0.68</td>
<td>0.23</td>
<td>0.87</td>
</tr>
<tr>
<td>1004</td>
<td>0.03</td>
<td>0.30</td>
<td>0.68</td>
<td>0.23</td>
<td>0.87</td>
</tr>
<tr>
<td>1005</td>
<td>0.03</td>
<td>0.30</td>
<td>0.68</td>
<td>0.23</td>
<td>0.87</td>
</tr>
<tr>
<td>1007</td>
<td>0.03</td>
<td>0.30</td>
<td>0.68</td>
<td>0.23</td>
<td>0.87</td>
</tr>
<tr>
<td>1008</td>
<td>0.03</td>
<td>0.30</td>
<td>0.68</td>
<td>0.23</td>
<td>0.87</td>
</tr>
<tr>
<td>1009</td>
<td>0.03</td>
<td>0.30</td>
<td>0.68</td>
<td>0.23</td>
<td>0.87</td>
</tr>
<tr>
<td>1010</td>
<td>0.03</td>
<td>0.30</td>
<td>0.68</td>
<td>0.23</td>
<td>0.87</td>
</tr>
<tr>
<td>1013</td>
<td>0.03</td>
<td>0.30</td>
<td>0.68</td>
<td>0.23</td>
<td>0.87</td>
</tr>
<tr>
<td>1014</td>
<td>0.03</td>
<td>0.30</td>
<td>0.68</td>
<td>0.23</td>
<td>0.87</td>
</tr>
<tr>
<td>1017</td>
<td>0.03</td>
<td>0.30</td>
<td>0.68</td>
<td>0.23</td>
<td>0.87</td>
</tr>
<tr>
<td>1019</td>
<td>0.03</td>
<td>0.30</td>
<td>0.68</td>
<td>0.23</td>
<td>0.87</td>
</tr>
<tr>
<td>1020</td>
<td>0.03</td>
<td>0.30</td>
<td>0.68</td>
<td>0.23</td>
<td>0.87</td>
</tr>
<tr>
<td>1021</td>
<td>0.03</td>
<td>0.30</td>
<td>0.68</td>
<td>0.23</td>
<td>0.87</td>
</tr>
<tr>
<td>1022</td>
<td>0.03</td>
<td>0.30</td>
<td>0.68</td>
<td>0.23</td>
<td>0.87</td>
</tr>
<tr>
<td>1024</td>
<td>0.03</td>
<td>0.30</td>
<td>0.68</td>
<td>0.23</td>
<td>0.87</td>
</tr>
</tbody>
</table>
Sensor Data Analysis of ADA

• Individual case
  – Cleaned data visualization
  – Statistical analysis for drinking and non-drinking
  – Unbalanced nested ANOVA
  – Drinking effect is main effect
  – Time effect is the effect within drinking effect

• 50% percent of subject has significant different in heart rate

<table>
<thead>
<tr>
<th>ID</th>
<th>Heart Rate</th>
<th>Breath Rate</th>
<th>Activity</th>
<th>Skin Temp</th>
</tr>
</thead>
<tbody>
<tr>
<td>1001</td>
<td>0.201</td>
<td>0.182</td>
<td>0.352</td>
<td>0.066</td>
</tr>
<tr>
<td>1004</td>
<td>0.000</td>
<td>0.001</td>
<td>0.224</td>
<td>0.432</td>
</tr>
<tr>
<td>1005</td>
<td>0.001</td>
<td>0.014</td>
<td>0.001</td>
<td>0.639</td>
</tr>
<tr>
<td>1007</td>
<td>0.741</td>
<td>0.263</td>
<td>0.163</td>
<td>0.186</td>
</tr>
<tr>
<td>1008</td>
<td>0.000</td>
<td>0.004</td>
<td>0.006</td>
<td>0.797</td>
</tr>
<tr>
<td>1013</td>
<td>0.970</td>
<td>0.158</td>
<td>0.035</td>
<td>0.386</td>
</tr>
<tr>
<td>1019</td>
<td>0.450</td>
<td>0.162</td>
<td>0.578</td>
<td>0.011</td>
</tr>
<tr>
<td>1020</td>
<td>0.000</td>
<td>0.949</td>
<td>0.051</td>
<td>0.035</td>
</tr>
</tbody>
</table>
Dataset of Deep ADA Module

- Dataset:
  - Data cleaning: ADA data cleaning
    - Labeled and unlabeled
  - 30 minutes data blocks
  - 3 types of signals: heart rate, skin temperature and accelerator,

<table>
<thead>
<tr>
<th></th>
<th>C-train</th>
<th>C-test</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within subject</td>
<td>214</td>
<td>50</td>
<td>80 : 20 (8 subjects)</td>
</tr>
<tr>
<td>Cross subject</td>
<td>212</td>
<td>52</td>
<td>Test: 1005, 1008, 1020</td>
</tr>
</tbody>
</table>
Experimental Result

- 1D DNN reconstruction
  - Within subject:
    - 0.85 correlation
  - Cross subject:
    - 0.81 correlation
Experimental Results (Cont.)

• Classification results:
  - Within subjects (accuracy)

<table>
<thead>
<tr>
<th>Models</th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unsupervised Learning</td>
<td>0.72</td>
<td>0.55</td>
</tr>
<tr>
<td>CNN Features</td>
<td>0.88</td>
<td>0.74</td>
</tr>
<tr>
<td>Supervised Learning</td>
<td>0.91</td>
<td>0.53</td>
</tr>
<tr>
<td>SVM</td>
<td>0.94</td>
<td>0.52</td>
</tr>
<tr>
<td>ResNet50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Cross subjects (accuracy)

<table>
<thead>
<tr>
<th>Models</th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unsupervised Learning</td>
<td>0.68</td>
<td>0.52</td>
</tr>
<tr>
<td>CNN Features</td>
<td>0.87</td>
<td>0.73</td>
</tr>
<tr>
<td>Supervised Learning</td>
<td>0.92</td>
<td>0.49</td>
</tr>
<tr>
<td>SVM</td>
<td>0.94</td>
<td>0.51</td>
</tr>
<tr>
<td>ResNet50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Content

- Introduction
- Related Work
- Deep ADA Pipeline
  - ADA (Automatic Detection of Alcohol usage) Module
  - Deep ADA (DNN Automatic Detection of Alcohol) Module
- Experimental Result
  - ADA
  - Deep ADA
- Summary
Summary

• Proposed a novel feature extraction using deep learning on physiological data
  – Encoder-decoder architecture can be used to extract features
    • Bottleneck as features
    • Reconstruction results are around 0.8 Pearson correlation

• The method proposed outperforms with other traditional feature extraction methods by ~ 20% accuracy
Outline

• Introduction

• Proposed Method
  – Novel Adaptive Saliency Biased Loss (ASBL) for Object Detection in Aerial Images
  – Improving Bird Recognition in Aerial Images using Deep Learning
  – A new Deep Learning Based Automatic Detection of Alcohol usage (DEEP ADA)

• Summary
## Summary

<table>
<thead>
<tr>
<th>Problem</th>
<th>Our Proposed Method</th>
<th>Result achieved</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Object Detection in Aerial Images</td>
<td>Novel Adaptive Saliency Biased Loss (ASBL)</td>
<td>Other models: 6.4 mAP (DOTA); 2.19 mAP (NWPU VHR-10)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Original RetinaNet: : 3.61 mAP (DOTA); 12.5 mAP (NWPU VHR-10)</td>
</tr>
<tr>
<td>Bird Detection in Aerial Images</td>
<td>DNN Object Detection Model Adaption and Analysis</td>
<td>State-of-the-art solution: 91.2 F1, 17.8 MAE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Improving from original by ~ 30 F1.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Model analysis based on bird problem</td>
</tr>
<tr>
<td>Ambulatory Assessment Analysis</td>
<td>Deep Learning Based Automatic Detection of Alcohol usage (DEEP ADA)</td>
<td>Comprehensive statistically analysis on alcohol drinking.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Improving supervised classification: 21 % accuracy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Improving other feature extraction: 19 % accuracy</td>
</tr>
</tbody>
</table>
Deep learning for Image recognition:


Ambulatory assessment analysis:

- **Peng Sun**, Tim Trull and Yi Shang. ‘A new DEEP LEARNING BASED Automatic Detection of Alcohol usage (DEEP ADA)’ *(To be submitted to Journal)*


Acknowledge

• Committee members:
  – Dr. Yi Shang
  – Dr. Jianlin Cheng
  – Dr. Dong Xu
  – Dr. Tim Trull

• My family
  – Shuhui Jia, Jason Sun, Jenny Sun
  – Our parents

• All members in Distributed and Intelligent Computing Lab
Thank you for your attending!