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

University of Missouri

Department of Electrical Engineering and Computer Science

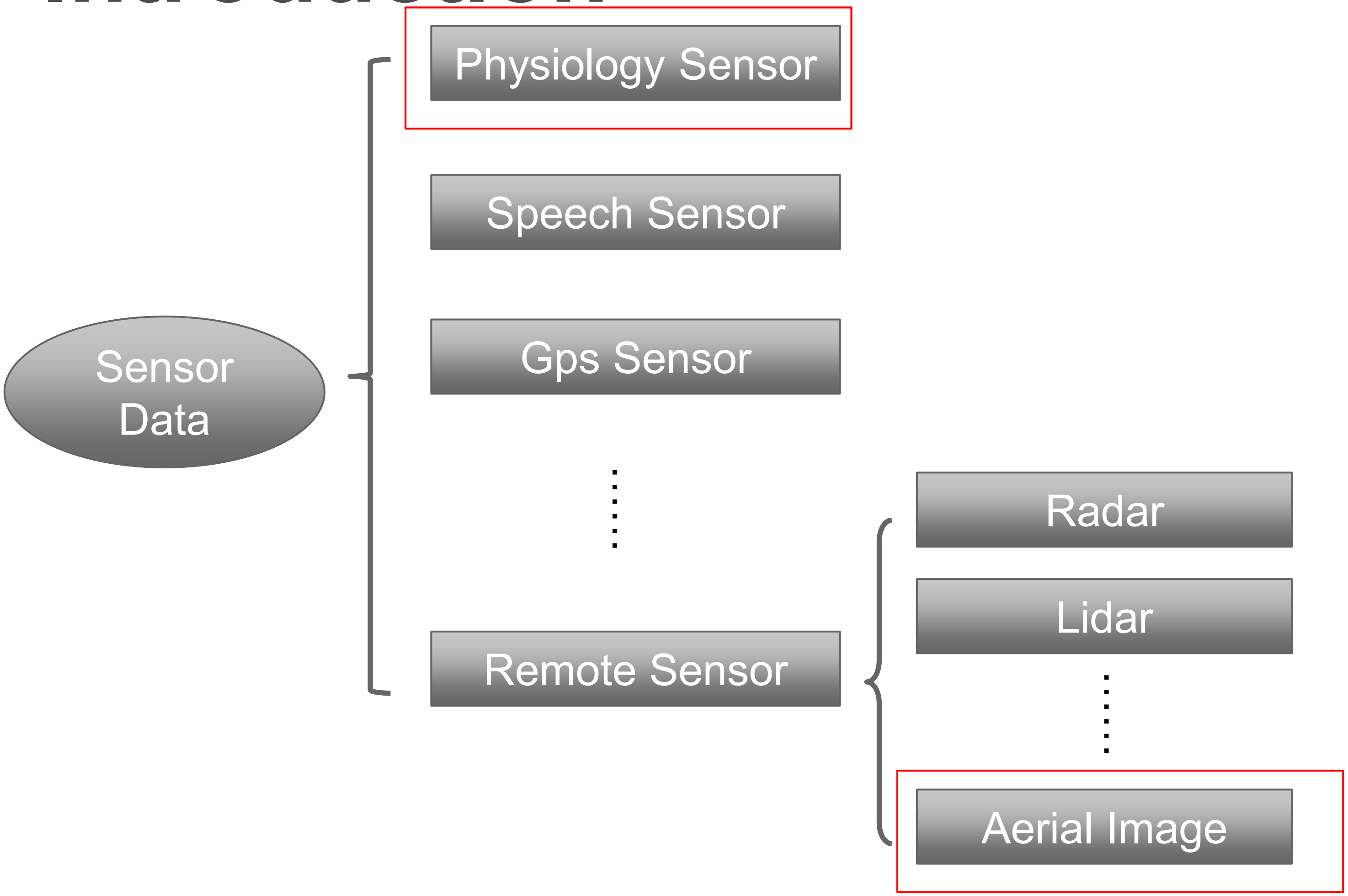
Speaker: Peng Sun

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



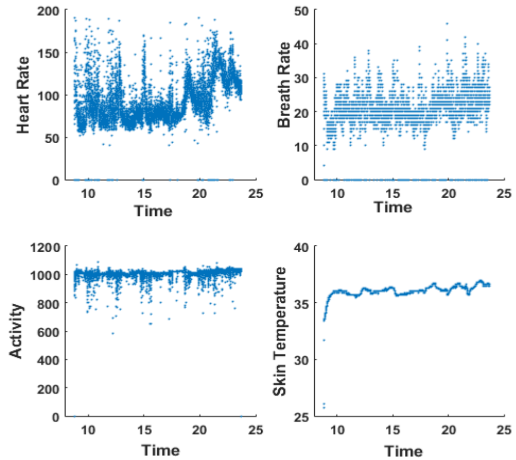
Our Research Focus on

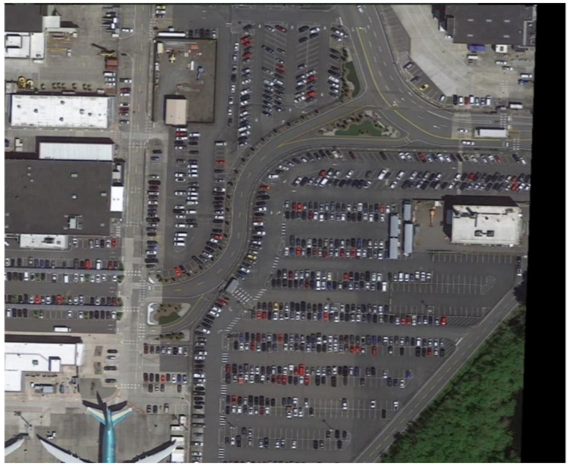
Physiology Sensor Data

Aerial Image Sensor

Ambulatory Assessment

Object Detection in Aerial Images



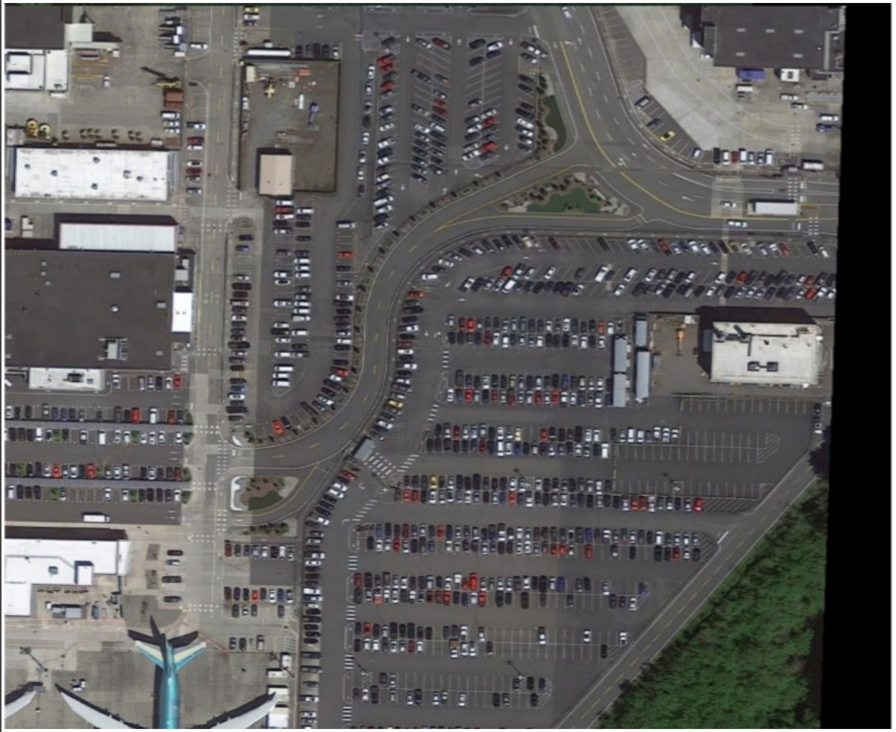


Main Contribution

Problem	Our Proposed Method
General Object Detection in Aerial Images	Novel Adaptive Saliency Biased Loss (ASBL)
Bird Detection in Aerial Images	DNN Object Detection Model Adaption and Analysis
Ambulatory Assessment Analysis	Deep Learning Based Automatic Detection of Alcohol usage (DEEP ADA)

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

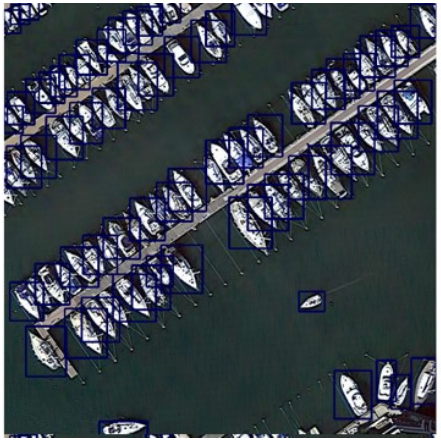
Content

- 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

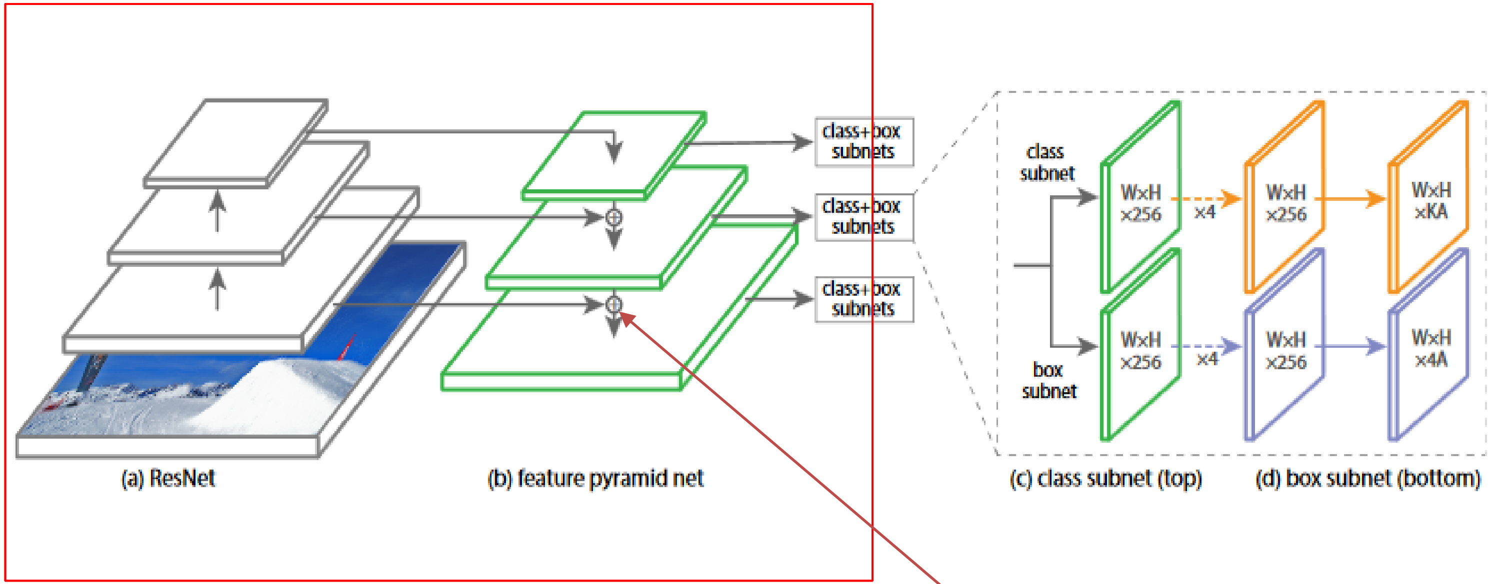
DNN Detector



Related Work

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

Theory Foundation (RetinaNet)



Focal Loss

Feature Pyramid Network

Lateral connection

Lin, Tsung-Yi, et al. "Focal loss for dense object detection." *IEEE transactions on pattern analysis and machine intelligence* (2018).

Theory Foundation (Cont.)

- Focal loss

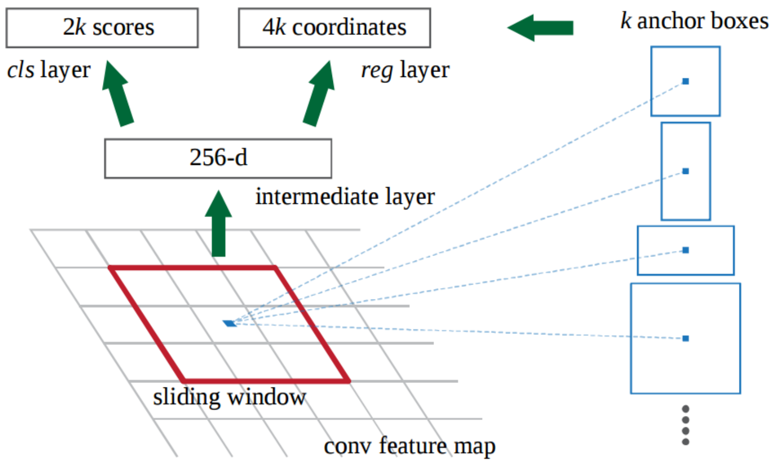
$$CE(p,y) = \begin{cases} -log(p), & if\ y=1 \\ -log(1-p) & otherwise \end{cases}$$

$$\alpha_t = \begin{cases} \alpha, & if\ y=1 \\ 1-\alpha & if\ y=0 \end{cases} \qquad p_t = \begin{cases} p, & if\ y=1 \\ 1-p & otherwise \end{cases}$$

$$FL(p,y) = \alpha_t * (1-p_t)^{\gamma} * CE(p,y)$$

α: foreground weight
γ: scale of important of misclassify
p: prob of classify
y: ground truth

- 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

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

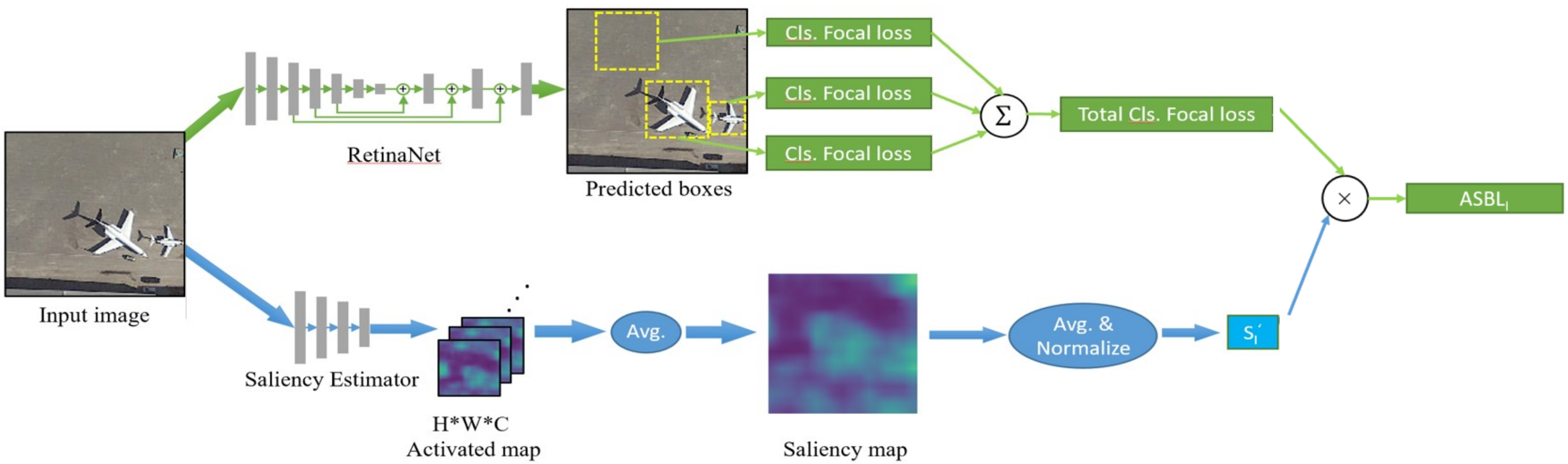
Loss

Novel Objective **Loss** function

Image Based ASBL

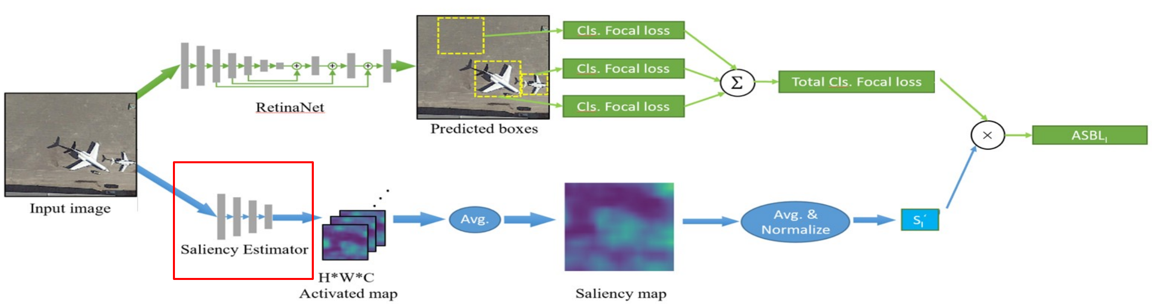
Anchor Based ASBL

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

f: Saliency Estimator $\xrightarrow[\text{H,W Avg}]{\text{Channel Avg}}$ $S_I(x) = \frac{1}{C*W*H}\sum_{c=1}^{C}\sum_{w=1}^{W}\sum_{h=1}^{H} f_{c,w,h}(x)$

- Normalize

$$S'_I(x) = \frac{S_I(x) - S_{min}}{S_{max} - S_{min}}(S_{ub} - S_{lb}) + S_{lb}$$

  – S_{max}, S_{min} is calculated by training data; S_{ub} set as 1.

- Image-based $ASBL_I(x, p, y) = S'_I(x) * FL(p, y)$

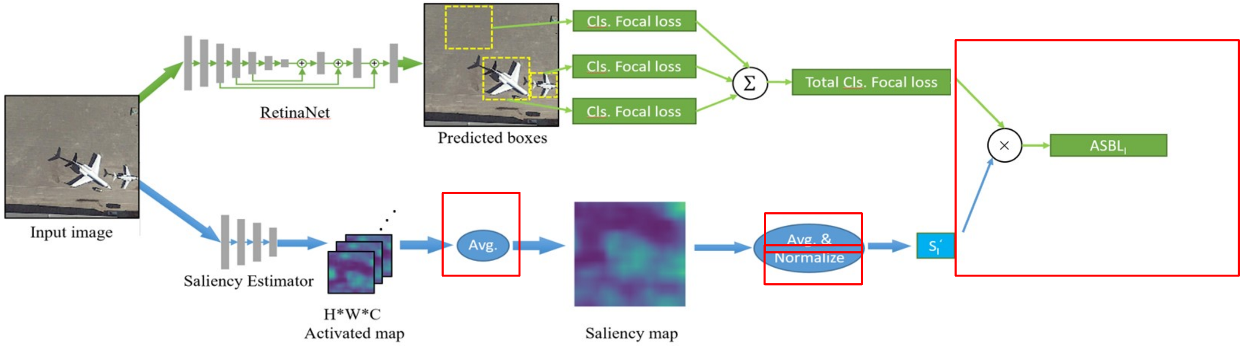
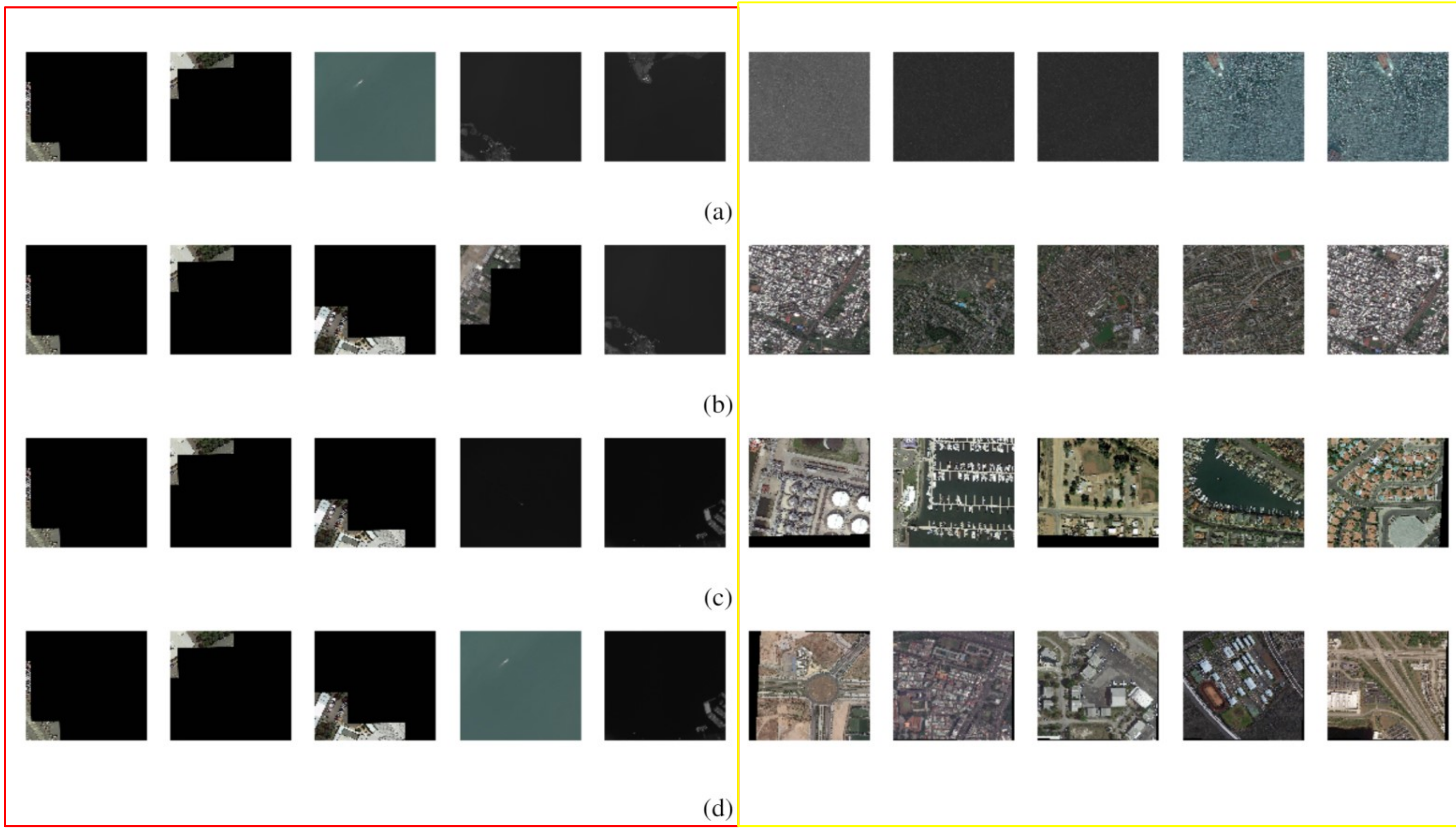


Image Complexity

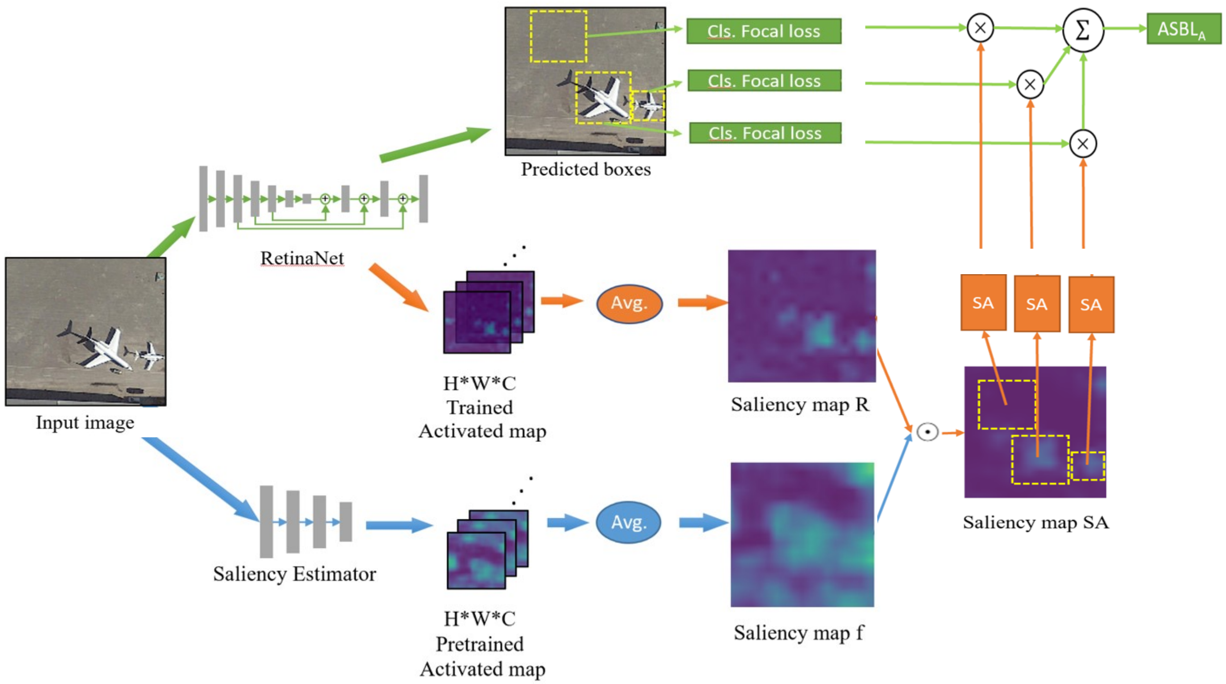
Low Level Conv Features

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

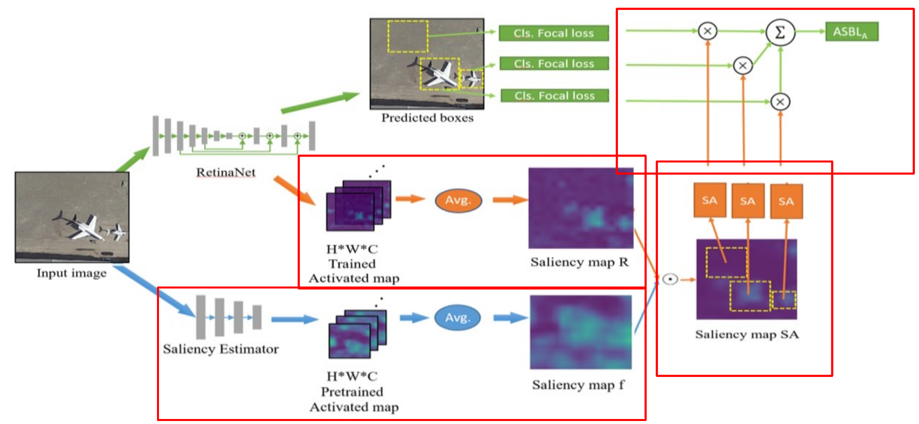
$$Pr(S|A,I) \propto Pr(A|S,I) * Pr(S|I)$$

$$\frac{1}{C}\sum_{c=1}^{C} R_{c,u,v}(x) \qquad \frac{1}{C}\sum_{c=1}^{C} f_{c,u,v}(x)$$

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

$\frac{1}{C}\sum_{c=1}^{C} R_{c,u,v}(x)$ Adaptive updated during training

- Anchor ASBL
  - $ASBL_A(p,y) = \sum_{a=1}^{A}\sum_{v=1}^{H}\sum_{u=1}^{W} SA_{u,v}(x) * FL_{u,v,a}(p,y)$
  - 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

NWPU VHR-10

2806
aerial images

188, 282
Instances

15
categories

1024 *1024
crop images

1/2 training;
1/6 validation;
1/3 test

10
Categories

400 * 400
crop images

679 training;
200 validation;
293 test

Xia, Gui-Song, et al. "DOTA: A large-scale dataset for object detection in aerial images." *Proc. CVPR*. 2018.

G. Cheng, P. Zhou, and J. Han. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. IEEE Trans. Geosci. Remote Sens., 54(12):7405–7415, 2016.


Large-vehicle
Swimming pool
Helicopter
Bridge
Plane
Ship
Soccer-ball field
Basketball court
Ground track field
Small-vehicle
Harbor
Baseball diamond
Tennis court
Roundabout
Storage tank

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

S_{lb}	**Normalization**	**mAP**
0.3	Y	61.79
0.5	Y	**63.48**
0.7	Y	62.4
-	N	62.86

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

Conv Block Layer of ResNet50	C2	C3	C4	C5
mAP	**64.77**	64.51	63.32	63.48

Anchor-based ASBL Ablation Study

Start from Image-based ASBL

- Adaptive update?
  - 64.82 -> 65.53
- Normalization?
  - 65.53 -> 66.12

Method	Dynamic update	Normalization	S_{lb}	mAP
$ASBL_A$	Y	Y	0.3	65.54
			0.5	**66.12**
			0.7	65.58
		N	-	65.53
	N	Y	0.5	64.82
$ASBL_I$	-	Y	0.5	64.77

Performance on DOTA

	Data	Plane	BD	Bridge	GTF	SV	LV	Ship	TC	BC	ST	SBF	RA	Harbor	SP	HC	mAP
YOLO [8]	T+V	76.9	33.87	22.73	34.88	38.73	32.02	52.37	61.65	48.54	33.91	29.27	36.83	36.44	38.26	11.61	39.2
SSD [11]	T+V	44.74	11.21	6.22	6.91	2	10.24	11.34	15.59	12.56	17.94	14.73	4.55	4.55	0.53	1.01	10.94
RFCN [9]	T+V	79.33	44.26	36.58	53.53	39.38	34.15	47.29	45.66	47.74	65.84	37.92	44.23	47.23	50.64	34.9	47.24
Faster RCNN [5]	T+V	80.32	**77.55**	32.86	**68.13**	53.66	52.49	50.04	90.41	**75.05**	59.59	**57**	49.81	61.69	56.46	**41.85**	60.46
RetinaNet [18]	T	78.22	53.41	26.38	42.27	63.64	52.63	73.19	87.17	44.64	57.99	18.03	51	43.39	56.56	7.44	50.39
RetinaNet*	T	89.03	62.14	43.88	47.05	73.57	65.18	78.65	90.86	66.28	70.26	35.07	58.26	68.93	66.34	22.16	62.51
ASBL-RetinaNet	T	89.09	67.96	46.38	57.12	73.55	66.19	**78.67**	90.86	71	**73.88**	45.15	60.92	70.01	68.51	32.49	66.12
ASBL-RetinaNet	T+V	**89.51**	74.07	**46.91**	55.54	**73.78**	**66.87**	78.48	**90.86**	70.09	73.2	46.71	**61.34**	**70.5**	**72.17**	32.84	**66.86**

* RetinaNet with modified anchor sizes and ratios

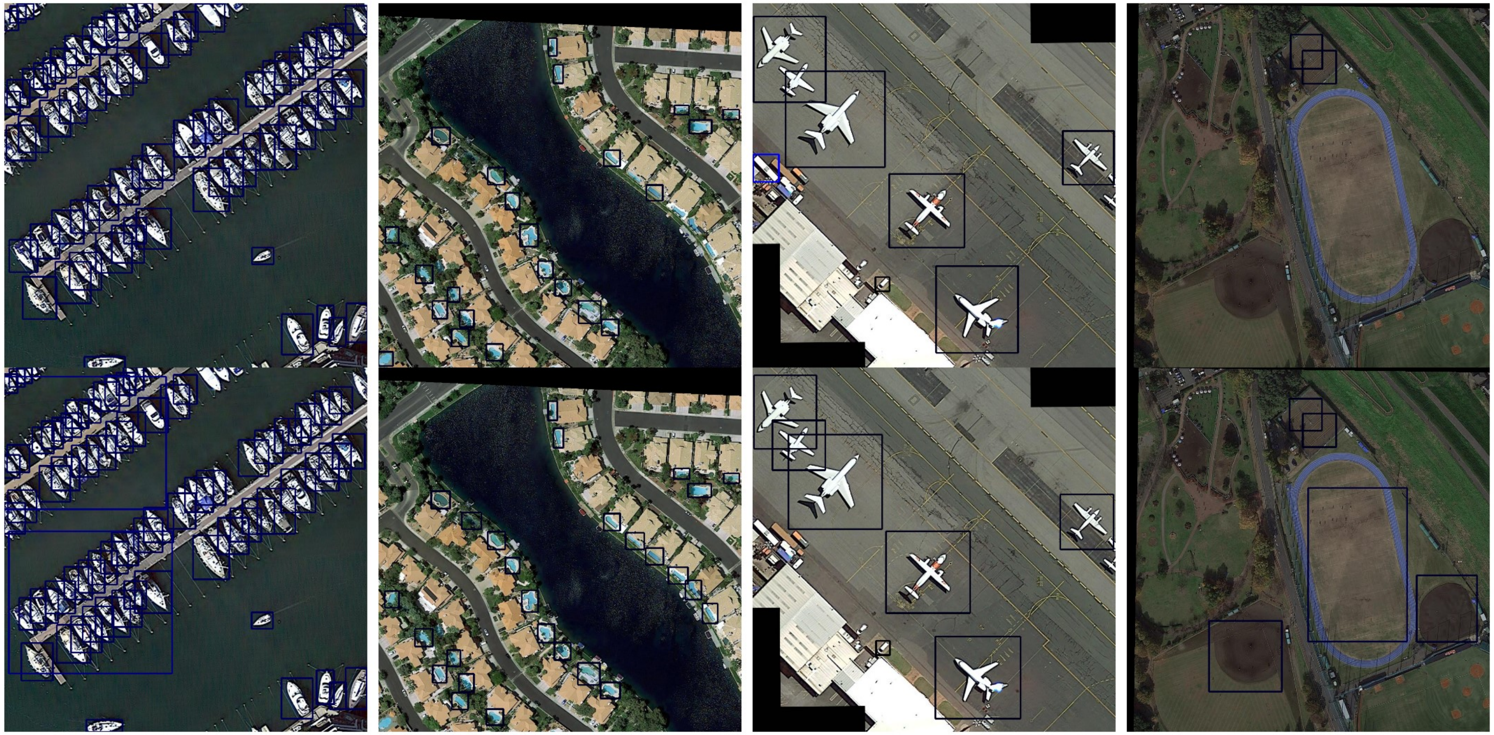
- 10 out of 15 outperform Faster RCNN
- 14 out of 15 outperform modified RetinaNet

Performance on NWPU VHR10

	Airplane	Ship	Storage tank	Baseball diamond	Tennis court	Basketball court	Ground track field	Harbor	Bridge	Vehicle	mAP
COPD [41]	62.25	69.37	64.52	82.13	34.13	35.25	84.21	56.31	16.43	44.28	54.89
Transferred CNN [1]	66.03	57.13	85.01	80.93	35.11	45.52	79.37	62.57	43.17	41.27	59.61
RICNN [29]	88.71	78.34	86.33	89.09	42.33	56.85	87.72	67.47	62.31	72.01	73.11
Faster RCNN [5]	90.9	86.3	90.53	**98.24**	89.72	69.64	100	80.11	61.49	78.14	84.51
Li etc [27]	99.70	90.8	90.61	92.91	**90.29**	80.13	90.81	80.29	68.53	**87.14**	87.12
RetinaNet* [18]	96.58	83.61	74.76	84.32	63.99	59.66	98.33	62.83	65.72	78.31	76.81
ASBL-RetinaNet (VGG16)	**100**	91.27	**96.76**	96.69	68.54	**87.67**	100	77.10	**83.92**	83.45	88.54
ASBL-RetinaNet (ResNet50)	99.34	**93.19**	94.36	97.70	71.19	84.67	**100**	**87.61**	81.49	83.50	**89.31**

Models	Average running time per images (s)
COPD [41]	1.16
Transferred CNN [1]	5.09
RICNN [29]	8.47
Faster RCNN [5]	0.09
Li etc [27]	2.89
ASBL-RetinaNet	**0.045**

Visualization Comparison



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

Visualization

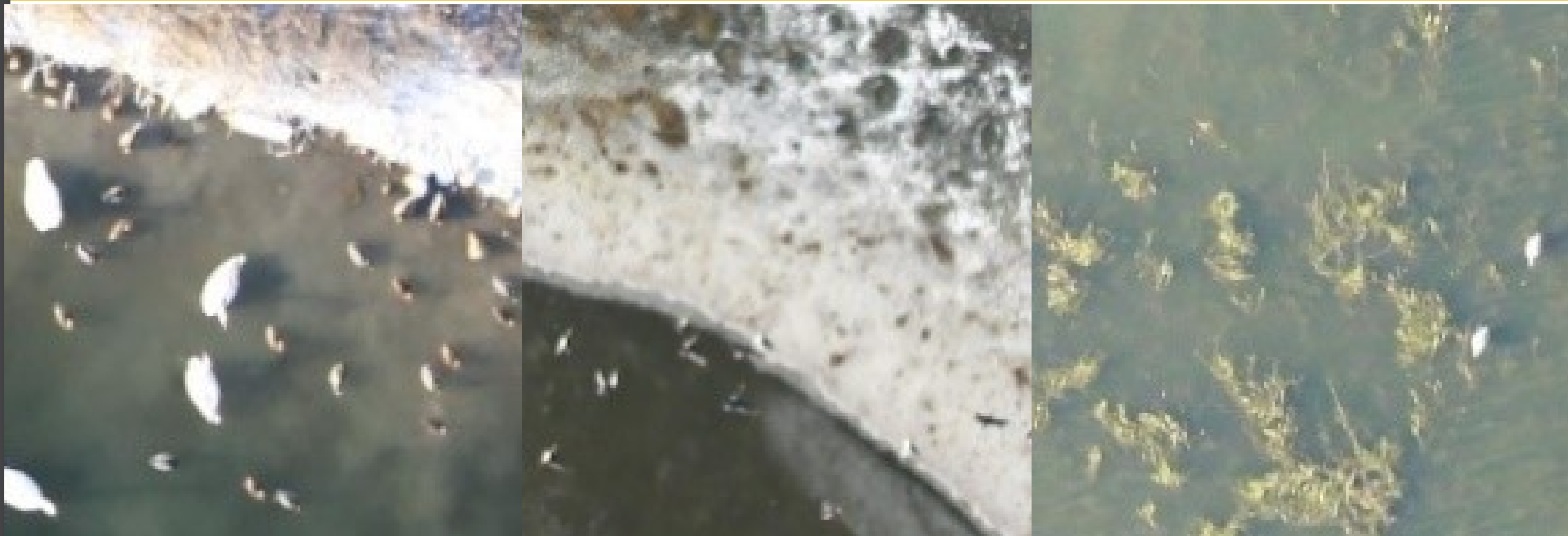
NWPU VHR-10

DOTA

Content

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



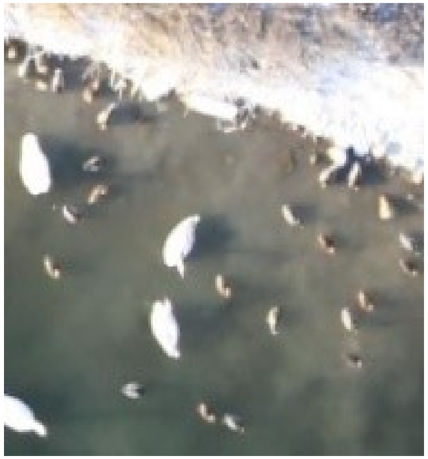
Improving Bird Recognition in Aerial Images using Deep Learning

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

	Authors	Methods
General object detection	Liu, 2016	Single Shot Multibox Detector
	Lin, 2017	RetinaNet
Aerial images object detection	Han, 2014	WSL and feature learning
	Li, 2017	Rotation-insensitive using RPN and local-contextual feature fusion network
	Cheng, 2019	Rotation-Invariant and Fisher Discriminative (RIFD)
Small object detection	Hu, 2017	Tiny face detector
	Yang, 2018	DFL-CNN (double focal loss CNN) for vehicle detection

LIBAI Dataset

Data Collection and Label

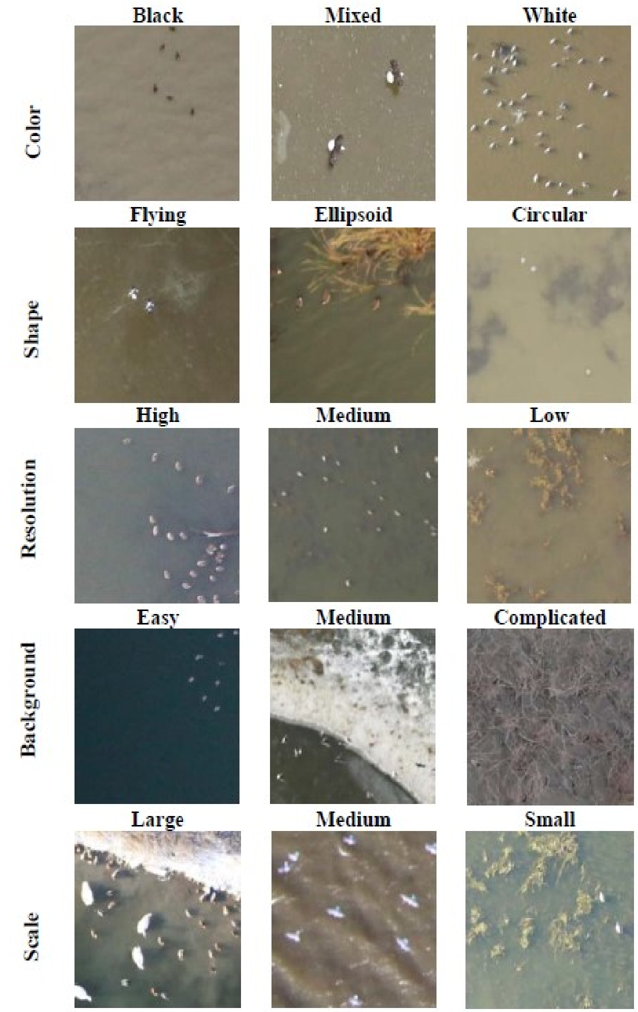
- Collection: Illinois Natural History Survey at the University of Illinois
- Labeling: human label

Dataset

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

Data Division

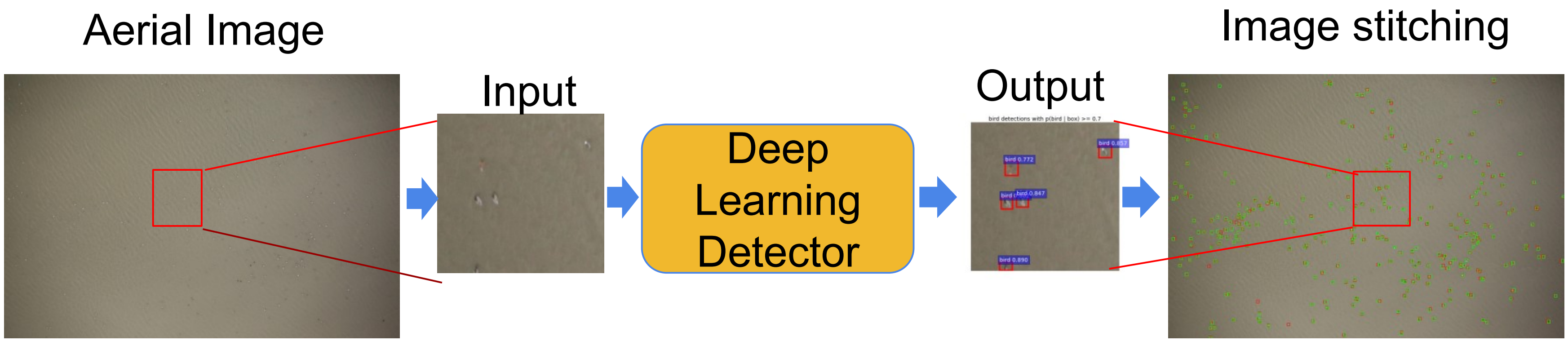
Based on complexity of Image, divide data into easy and hard



Content

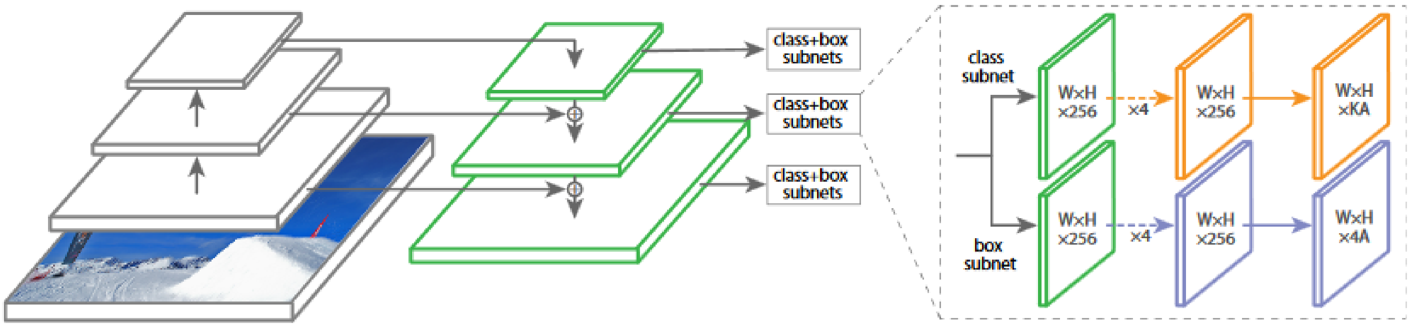
- Introduction
- Related Work
- LIBAI Dataset
- **Model Theory and Adaption**
- Experimental Result and Analysis
- Summary

Proposed Solution



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

RetinaNet

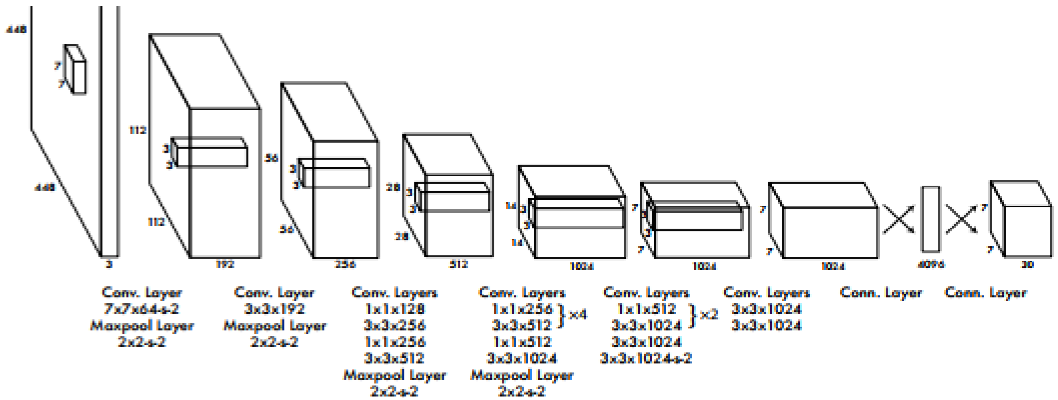


- Theory:
  - Feature Pyramid Network
  - Lateral connection
  - Focal Loss

- 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

Lin, Tsung-Yi, et al. "Focal loss for dense object detection." *IEEE transactions on pattern analysis and machine intelligence* (2018).

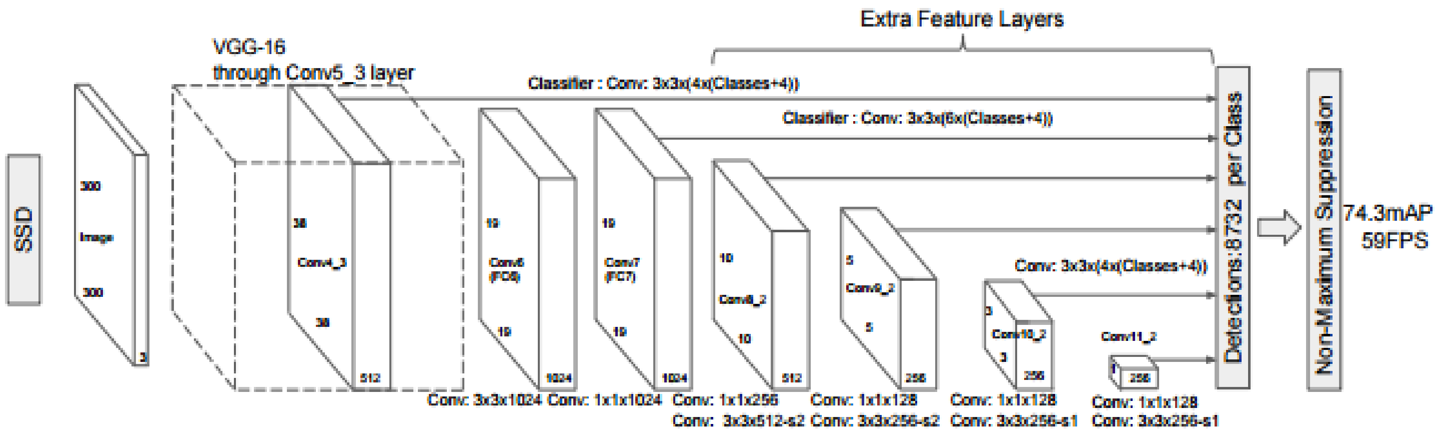
YOLO v3



- 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

Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." *arXiv preprint arXiv:1804.02767* (2018).

SSD

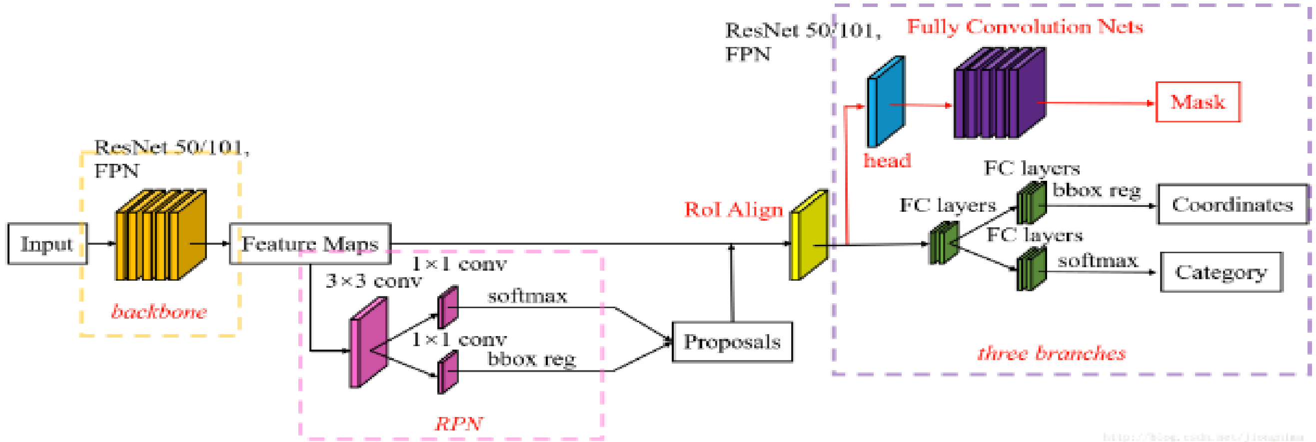


- Theory:
  - Multi-scale feature maps for detection
  - Default boxes and aspect ratios
  - Hard negative mining

- 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

Liu, Wei, et al. "Ssd: Single shot multibox detector." *European conference on computer vision*. Springer, Cham, 2016.

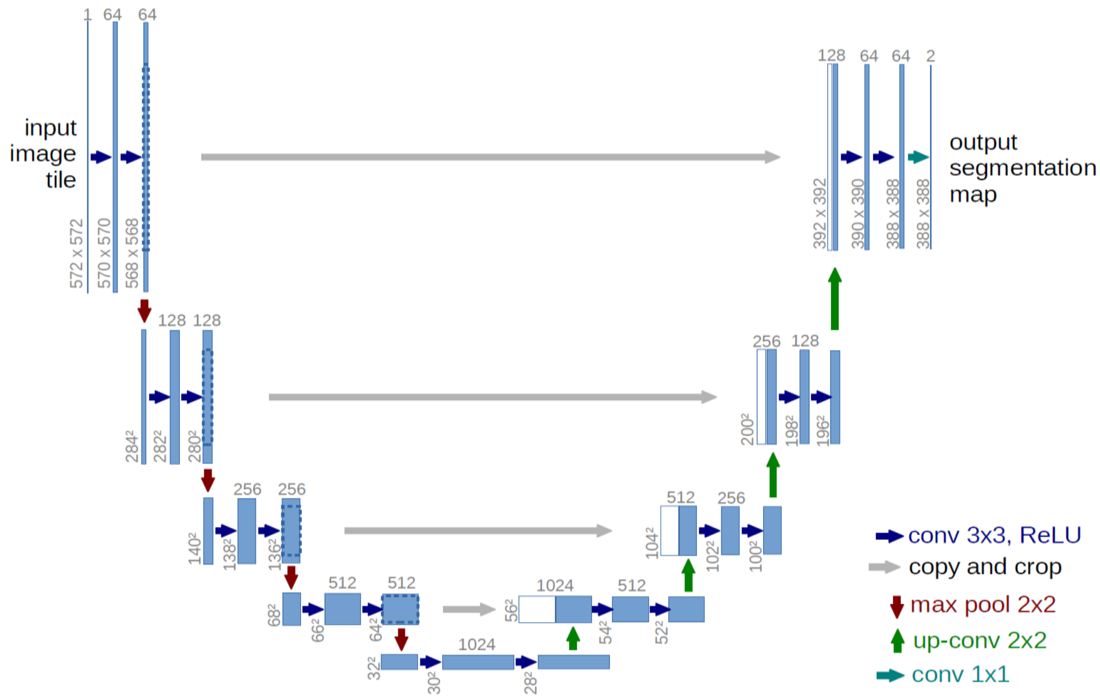
Mask R-CNN



- 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

He, Kaiming, et al. "Mask r-cnn." *Computer Vision (ICCV), 2017 IEEE International Conference on*. IEEE, 2017.

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

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015.

Content

- Introduction
- Related Work
- LIBAI Dataset
- Model Theory and Adaption
- **Experimental Result and Analysis**
- Conclusion

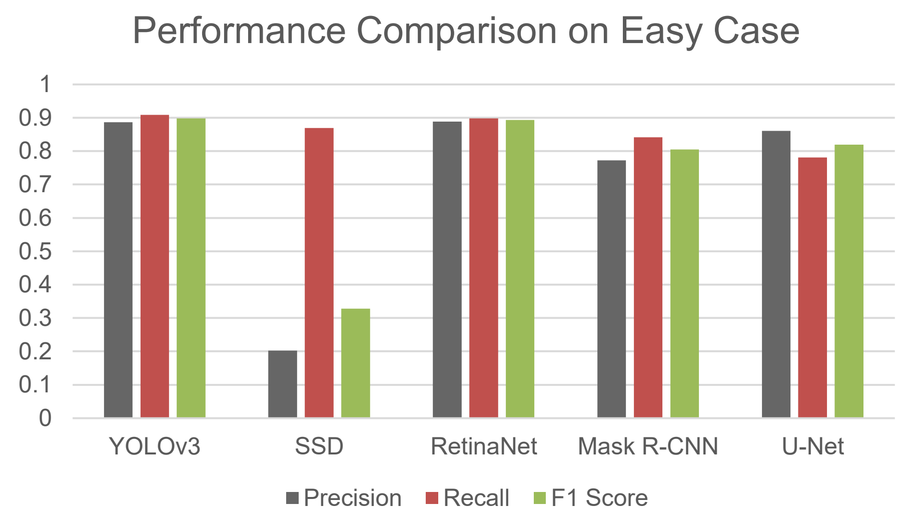
Performance Comparison

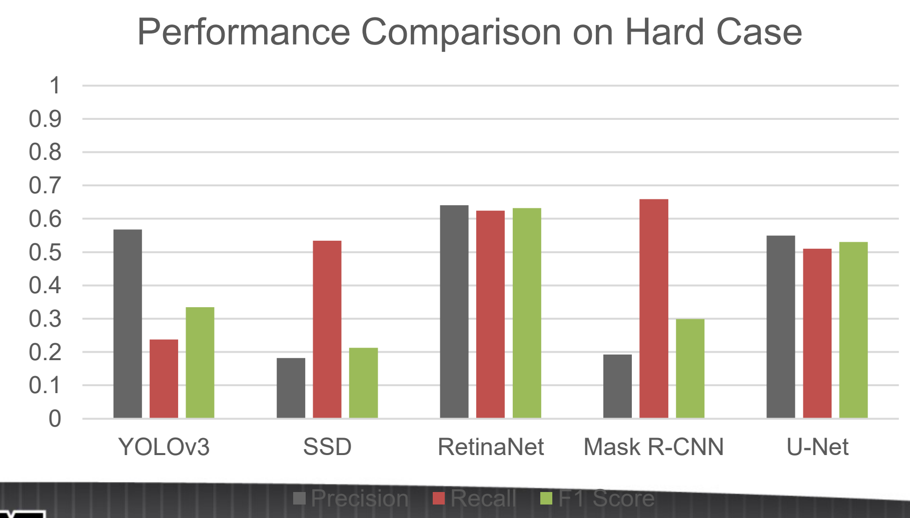
- Easy Data

Train	Val	Test	Total
121	31	10	162

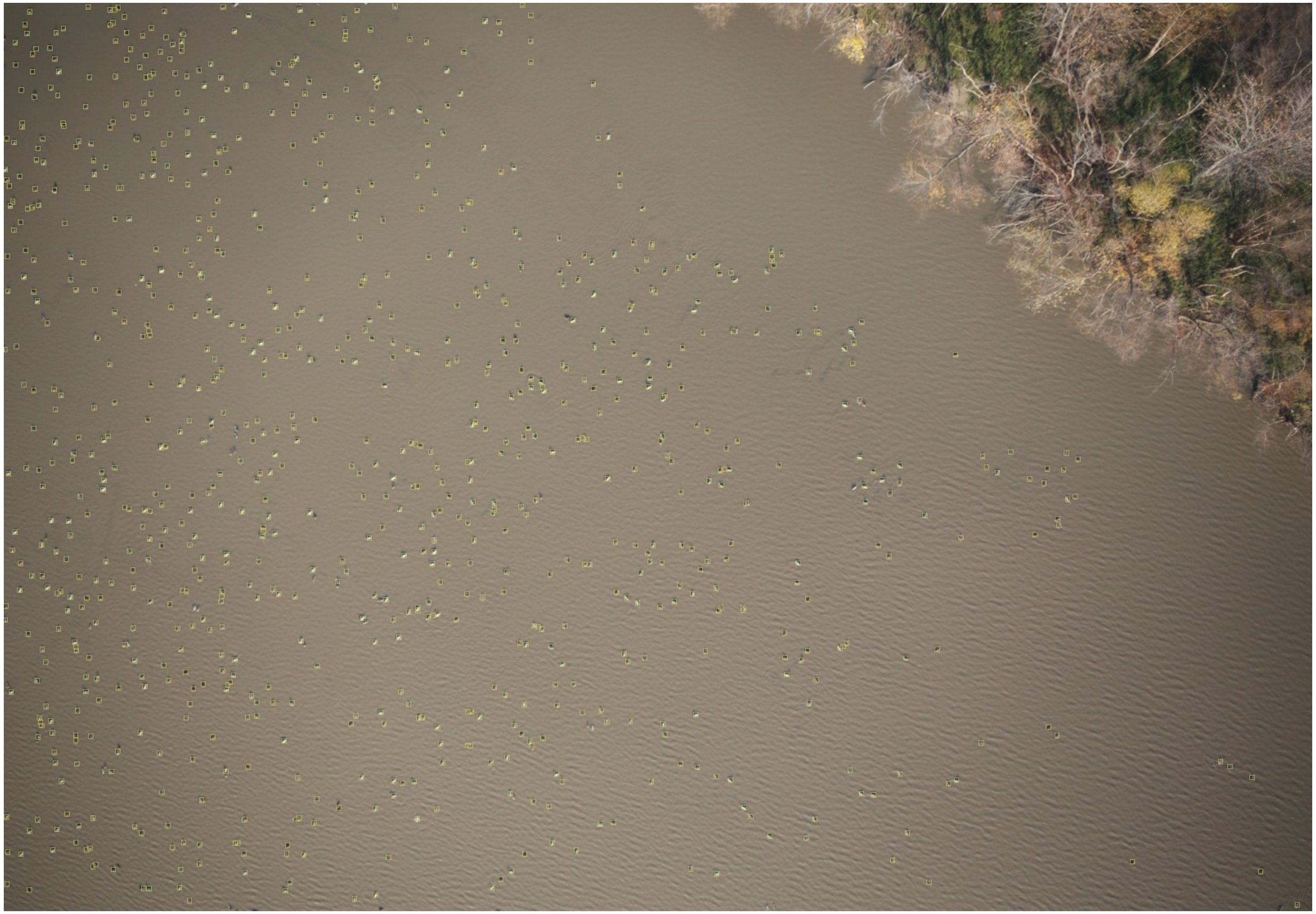
- Hard Data

Train	Val	Test	Total
133	31	10	174





Visualization (Easy)



Model	RetinaNet
# of bird	932
# of count	919
Precision	99.21%
Recall	92.72%

Visualization (Hard)

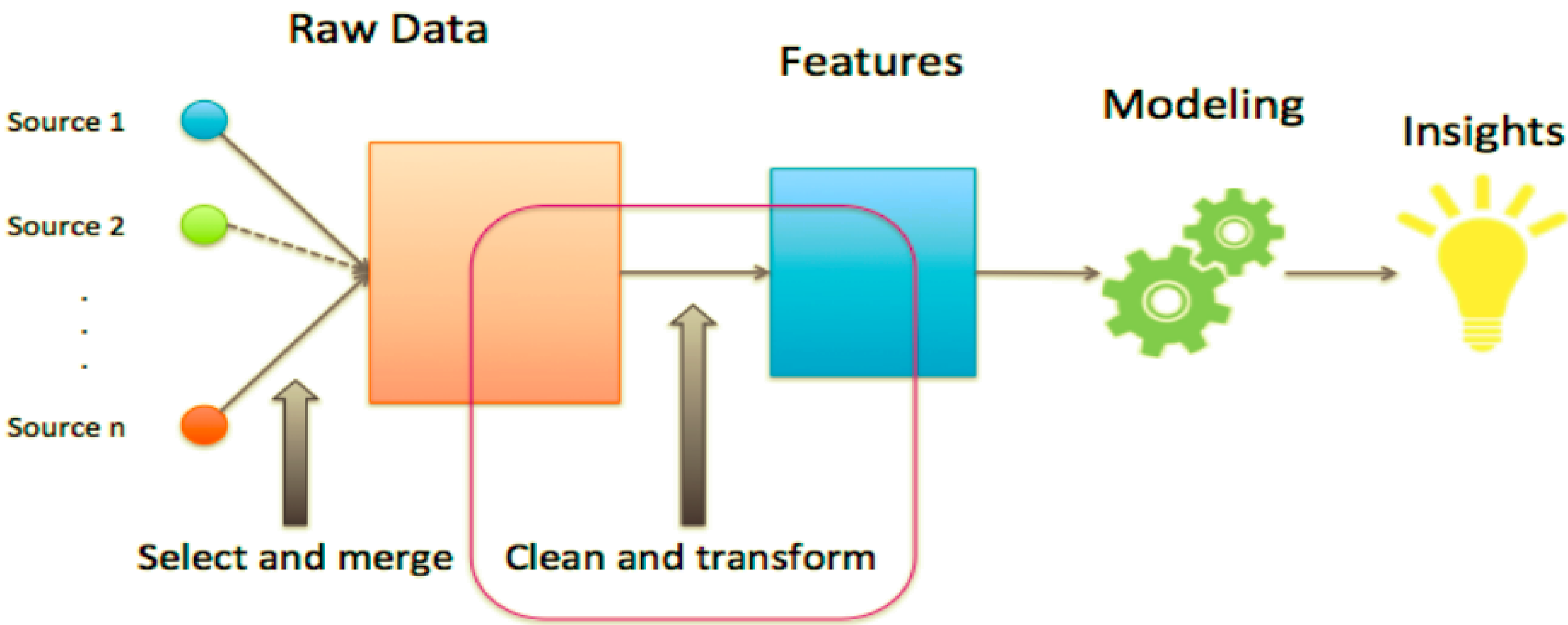
Model	RetinaNet
# of bird	3
# of count	37
Precision	2.71%
Recall	33.33%

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 Adaption 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



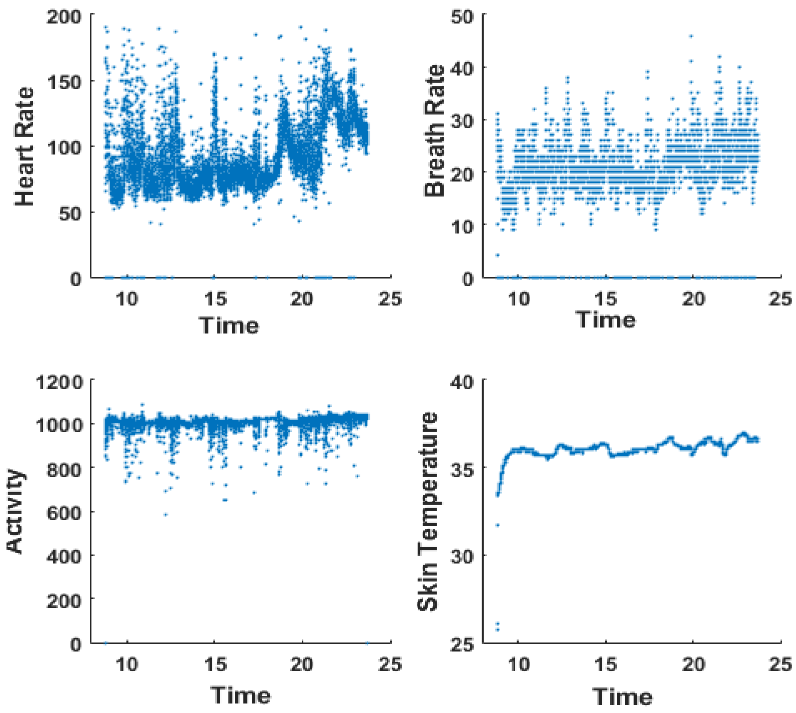
A new Deep Learning Based Automatic Detection of Alcohol usage (DEEP ADA)

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?



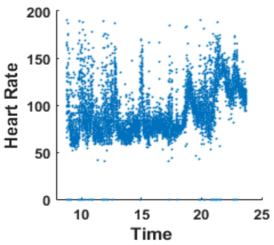
Related Work

	Authors	Methods
Physiological sensor analysis	Hossain, 2014	Cocaine intake detection
	Wei, 2016	Time-Frequency CNN for Automatic Sleep Stage Classification
	Zheng, 2017	Time series classification using multi-channels CNN
Unsupervised Feature Learning	Xu, 2016	Deep auto encoder (unsupervised feature leraning) for audio tagging
	Masci, 2011	Stacked CNN auto-encoders for hierarchical feature extraction on MNIST images

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



Automatic Detection of Alcohol usage (ADA)

DNN Automatic Detection of Alcohol (Deep ADA)

Sensor data cleaning

Sensor analysis

Survey analysis

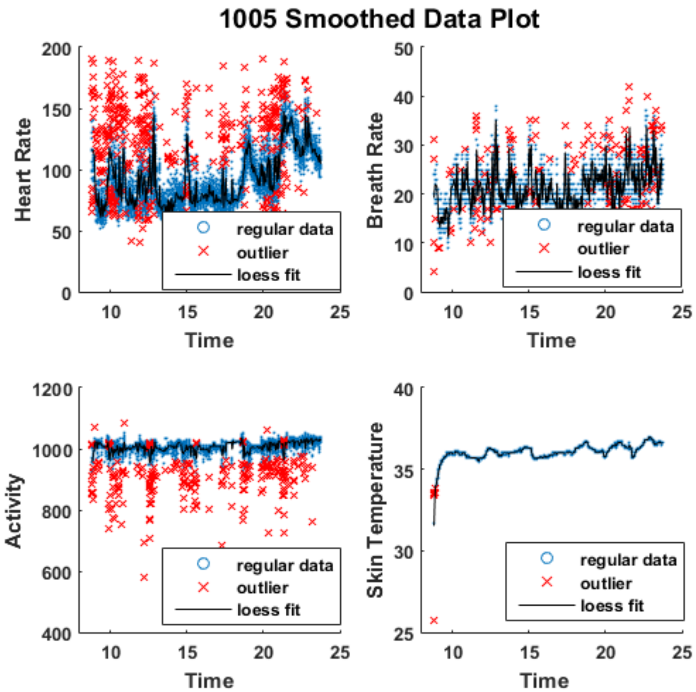
DNN feature extraction

ML classifier

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

$$\min \ p \sum_i w_i (y_i - s(x_i))^2 + (1-p) \int (\frac{d^2 s}{dx^2})^2 dx$$

  - 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

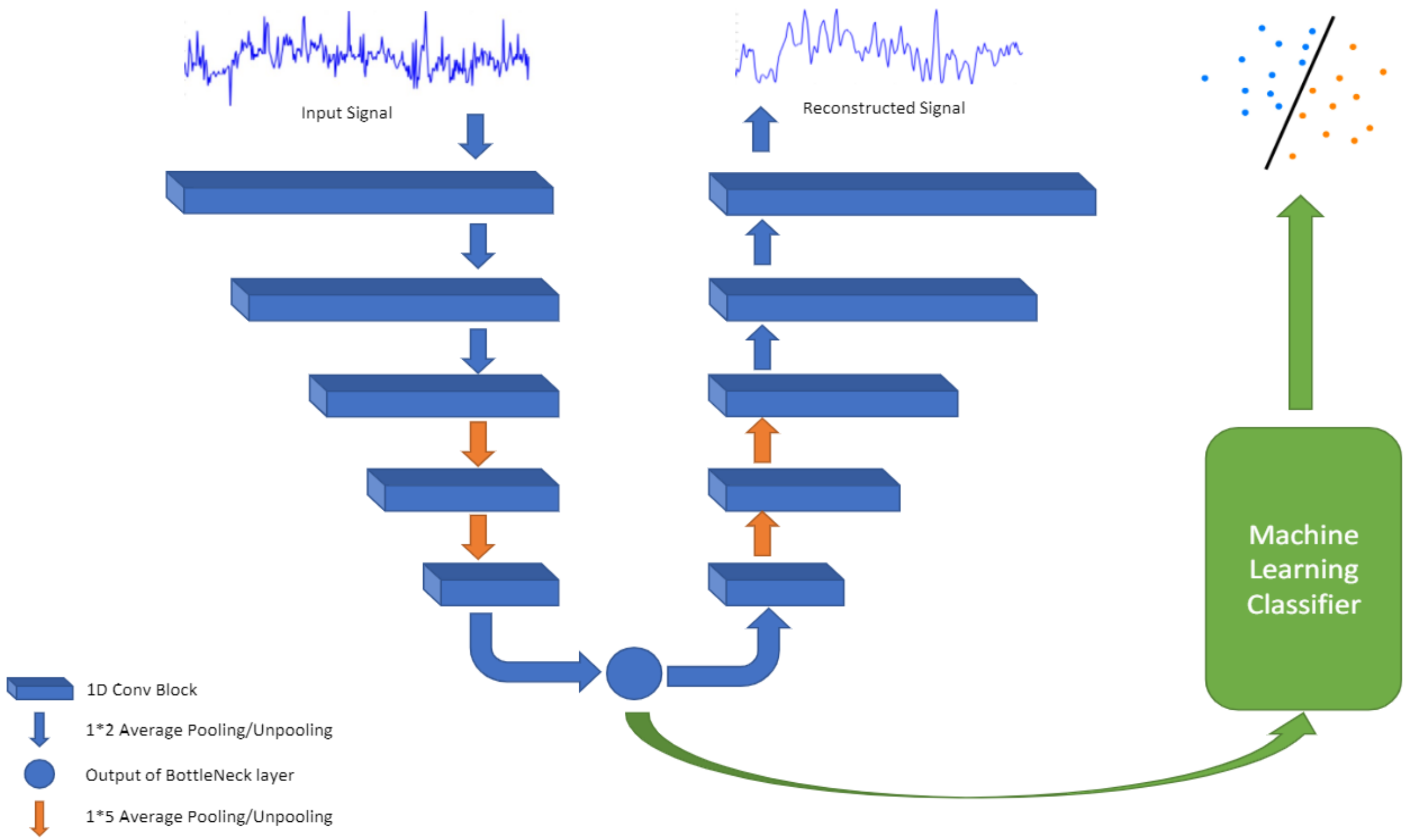
ML Classifier

- SVM
- ResNet50

ML Classifier

naïve bayes, decision tree, RF, adaboost and SVM

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/ Wilcox signed ranked test

P Value of Individual in Drink Day/Time

ID	Positive		Negative		Fear		Hostility		Sadness	
1001	0.05	0.29	0.00	0.01	0.33	0.25	0.00	0.03	0.00	0.02
1003	0.88	0.82	0.04	0.05	0.00	0.01	0.24	0.27	0.51	0.38
1004	0.76	0.54	0.74	0.03	0.30	0.20	0.27	0.27	0.15	0.13
1005	0.01	0.87	0.62	0.84	0.09	0.24	0.20	0.72	0.95	0.85
1007	0.37	0.26	0.46	0.98	0.09	0.68	0.64	0.36	0.78	0.92
1008	0.06	0.00	0.19	0.04	0.89	0.98	0.10	0.02	0.17	0.03
1009	0.58	0.82	0.16	0.23	0.37	0.17	0.08	0.25	0.72	0.44
1010	0.00	Null	0.00	Null	0.02	Null	0.00	Null	0.00	Null
1013	0.15	0.30	0.06	0.08	0.12	0.19	0.08	0.17	0.27	0.30
1014	0.04	0.92	0.30	0.51	0.48	0.50	0.58	0.65	0.32	0.56
1017	0.94	0.79	0.84	0.44	0.38	0.74	0.99	0.30	0.49	0.95
1019	0.59	0.73	0.00	0.05	0.00	0.08	0.03	0.13	0.60	0.38
1020	0.17	0.06	0.91	0.64	0.28	0.08	0.47	0.27	0.71	0.77
1021	0.16	0.00	0.16	0.03	0.66	0.33	0.40	0.19	0.00	0.01
1022	0.82	0.14	0.07	0.62	0.30	0.40	0.50	0.97	0.04	0.65
1024	0.18	0.05	0.69	0.25	0.50	0.24	0.83	0.61	0.46	0.15

P Value of All Subjects in Drink Day/Time

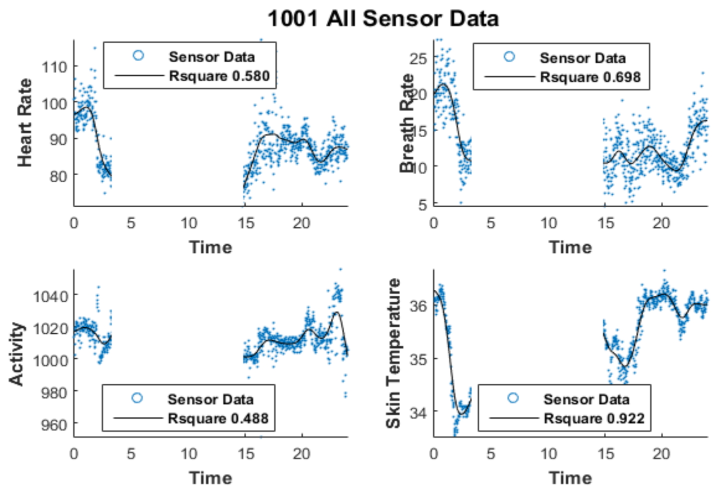
	Positive		Negative		Fear		Hostility		Sadness	
Mean	0.57	0.76	0.03	0.00	0.30	0.04	0.28	0.00	0.41	0.03
Variance	0.30	0.03	0.62	0.30	0.68	0.23	0.87	0.10	0.85	0.74

Increasing Ratio(%) in Drink Day/Time

	Positive		Negative		Fear		Hostility		Sadness	
Mean	2.17	3.46	-4.97	-7.86	-1.76	-4.26	-5.22	-9.68	-5.71	-8.57
Variance	0.07	-27.46	17.83	-12.97	15.43	-21.56	5.00	-24.83	7.32	-10.08

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



P Value of Drinking Effect for Each Individual				
ID	Heart Rate	Breath Rate	Activity	Skin Temp
1001	0.201	0.182	0.352	0.066
1004	0.000	0.001	0.224	0.432
1005	0.001	0.014	0.001	0.639
1007	0.741	0.263	0.163	0.186
1008	0.000	0.004	0.006	0.797
1013	0.970	0.158	0.035	0.386
1019	0.450	0.162	0.578	0.011
1020	0.000	0.949	0.051	0.035

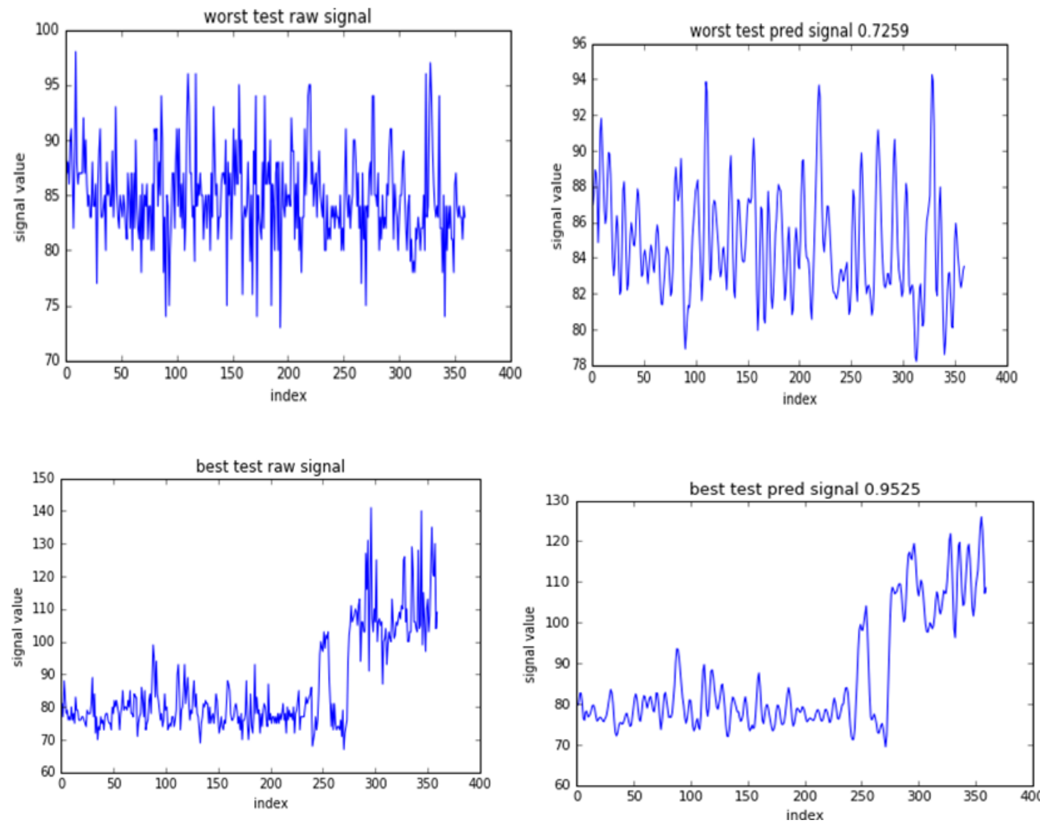
Dataset of Deep ADA Module

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

	C-train	C-test	Note
Within subject	214	50	80 : 20 (8 subjects)
Cross subject	212	52	Test: 1005, 1008, 1020

Experimental Result

- 1D DNN reconstruction
  - Within subject:
    - 0.85 correlation
  - Cross subject:
    - 0.81 correlation



Experimental Results (Cont.)

- Classification results:
  - Within subjects (accuracy)

	Models	**Train**	**Test**
Unsupervised Learning	Stats Features	0.72	0.55
	CNN Features	0.88	**0.74**
Supervised Learning	SVM	0.91	0.53
	ResNet50	0.94	0.52

  - Cross subjects (accuracy)

	Models	**Train**	**Test**
Unsupervised Learning	Stats Features	0.68	0.52
	CNN Features	0.87	**0.73**
Supervised Learning	SVM	0.92	0.49
	ResNet50	0.94	0.51

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

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

- Deep learning for Image recognition :
  - **Peng Sun**, Chen, Guang, and Yi Shang. 'Adaptive Saliency Biased Loss for Object Detection in Aerial Images' *IEEE Transactions on Geoscience and Remote Sensing*. IEEE. (**Submitted after Revision**)
  - Liu, Y., **Sun, P**., Highsmith, M. R., Wergeles, N. M., Sartwell, J., Raedeke, A., ... & Shang, Y. (2018, June). Performance comparison of deep learning techniques for recognizing birds in aerial images. In *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)* (pp. 317-324). IEEE.
  - Yang Liu, **Peng Sun**, and Yi Shang. 'A Survey and Performance Evaluation of Deep Learning Methods for Small Object Detection' (**To be submitted to Journal**)
  - **Sun, P.***, Guerdan, L.*, Rowland, C., Harrison, L., Tang, Z., Wergeles, N., & Shang, Y. (2019, October). Deep Learning vs. Classical Machine Learning: A Comparison of Methods for Fluid Intelligence Prediction. In *Challenge in Adolescent Brain Cognitive Development Neurocognitive Prediction* (pp. 17-25). Springer, Cham.
  - Chen, Guang, **Peng Sun**, and Yi Shang. "Automatic Fish Classification System Using Deep Learning." *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, 2017.
- 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**)
  - **Sun, Peng,** et al. "Ada-automatic detection of alcohol usage for mobile ambulatory assessment." *2016 IEEE International Conference on Smart Computing (SMARTCOMP)*. IEEE, 2016.
  - Shi, Ruiqi, **Peng, Sun** et al. "mAAS--A Mobile Ambulatory Assessment System for Alcohol Craving Studies." *2015 IEEE 39th Annual Computer Software and Applications Conference*. Vol. 3. IEEE, 2015.
  - Bernstein, J. P., Mendez, B. J., **Sun, P**., Liu, Y., & Shang, Y. (2017, January). Using deep learning for alcohol consumption recognition. In *2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC)* (pp. 1020-1021). IEEE.

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!