University of Missouri

Region Based Object Detectors for Recognizing Birds in Aerial Images

M.S. Thesis Defense

By: Pedro Carrion Castagnola

Advisor: Dr. Yi Shang

Outline

1. Introduction
2. Related Work
3. Design and Implementation
4. Experiments
5. Results
6. Conclusions and Future Work

Outline

1. **Introduction**
    - **Problem**
    - **Motivation**
2. Related Work
3. Design and Implementation
4. Experiments
5. Results
6. Conclusions and Future Work

Problem

- Missouri Department of Conservation: Protection and management of birds in different areas in Missouri
- Recognition and counting of birds
- Data: Aerial images
- Difficult problem
  - Birds look small in the images
  - Blurriness
  - Different shapes and colors
  - Complex backgrounds

Motivation

- Try a new type of Neural Network recognize the birds: Capsule Network
- Capsule Network only does classification. First generate region proposals
- Compare the performances of Capsule Network with a simple CNN and a state of the art neural network
- Explore network parameters that can improve the results

Outline

1. Introduction
2. **Related Work**
   - **Dataset**
   - **Convolutional Neural Network**
   - **Capsule Network**
   - **ResNet50 + FPN**
3. Design and Implementation
4. Experiments
5. Results
6. Conclusions and Future Work

Previous Work

- Crop images into 512 x 512 px
- Divided dataset into Easy and Hard datasets

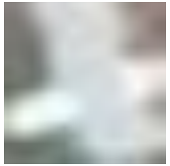
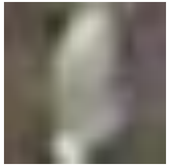
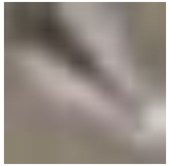
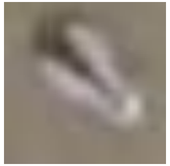
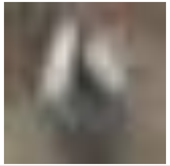
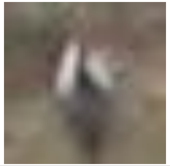
Easy image

Hard image

Labels

- One coordinate point approximately at the center of each bird is known
- Labels generation: 20 by 20 and 30 by 30 pixels bounding box size
- 4 Datasets will be used

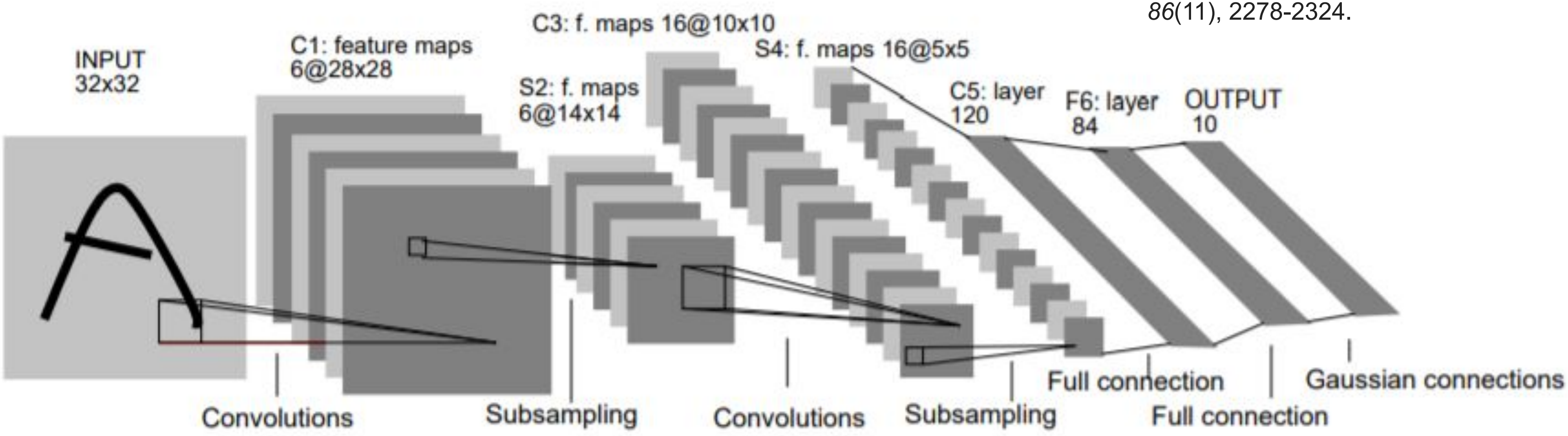
#	Dataset Label Size	Dataset Difficulty
1	20 x 20	Easy
2	20 x 20	Hard
3	30 x 30	Easy
4	30 x 30	Hard

	20x20 Label	30x30 Label
Sample 1		
Sample 2		
Sample 3		
Sample 4		
Sample 5		

Convolutional Neural Network

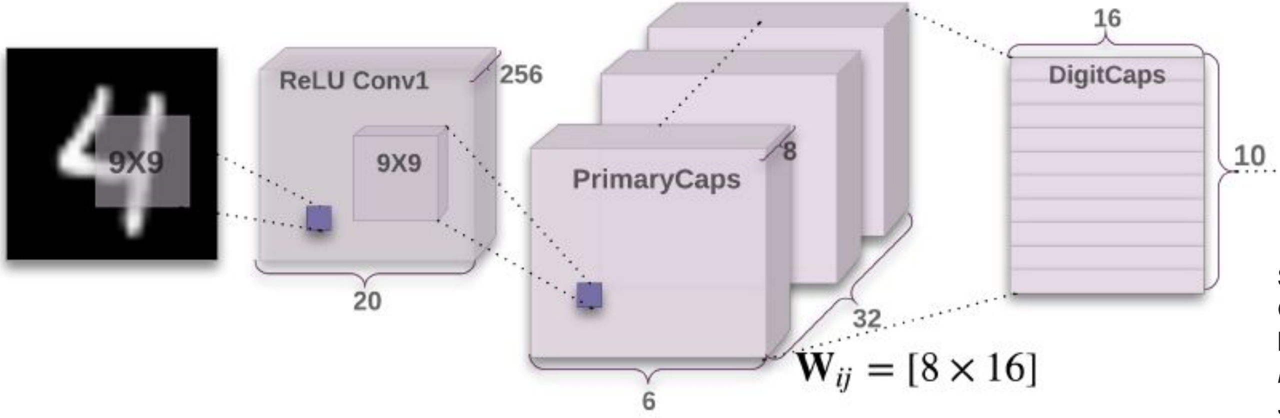
- Convolutional layers - Feature Maps
- Scalar outputs

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE 86*(11), 2278-2324.



Capsule Network

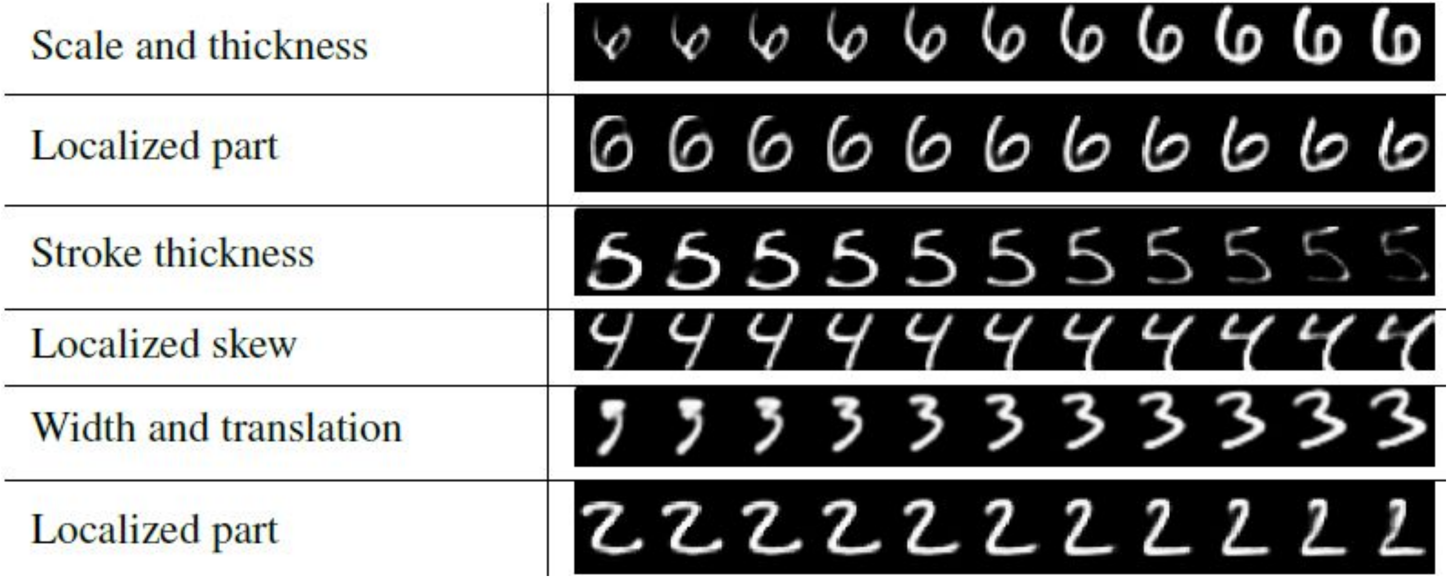
- Spatial relationship between simple and complex objects
- Instead of using scalar outputs, it uses vector outputs (capsules)
  - Length represents the probability that an object exists
  - Orientation represents instantiation parameters of an object, as for example position, size, rotation, etc



Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. In *Advances neural information processing systems* (pp. 3856-3866).

Capsule Network

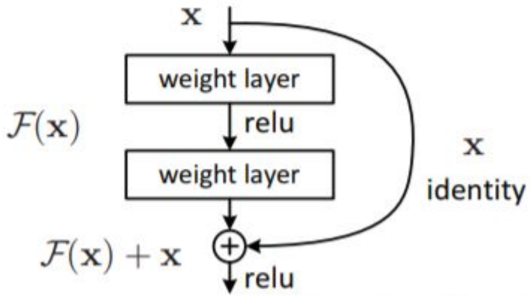
- Reconstruction



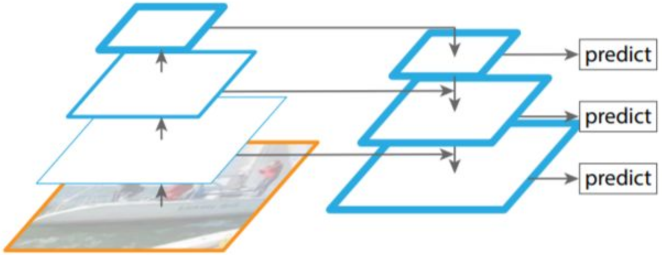
Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. In *Advances neural information processing systems* (pp. 3856-3866).

ResNet50 and Feature Pyramids Network (FPN)

- Residual Network avoid gradient diffusion in deep NN using shortcut connections. Accuracy will not decrease if more layers are added
- Feature Pyramids Networks allows predictions at different scales (good for small objects). Works with ResNets



He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

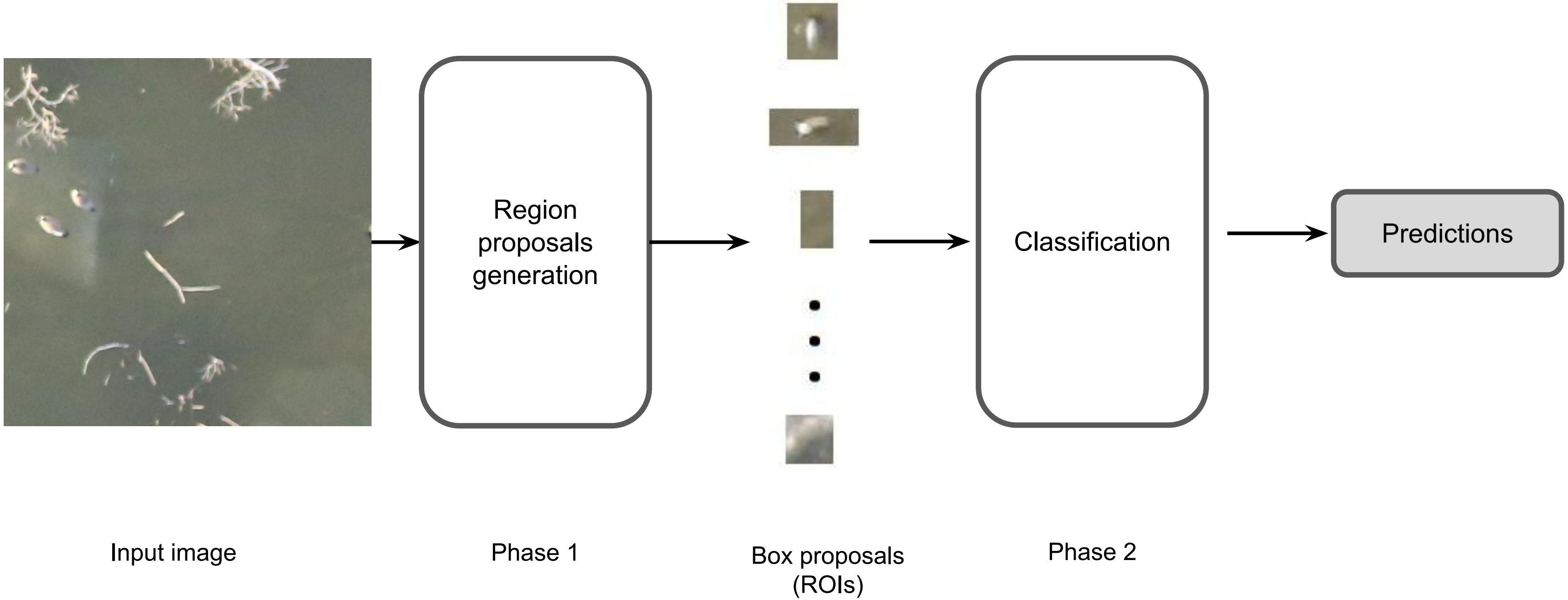


Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2117-2125).

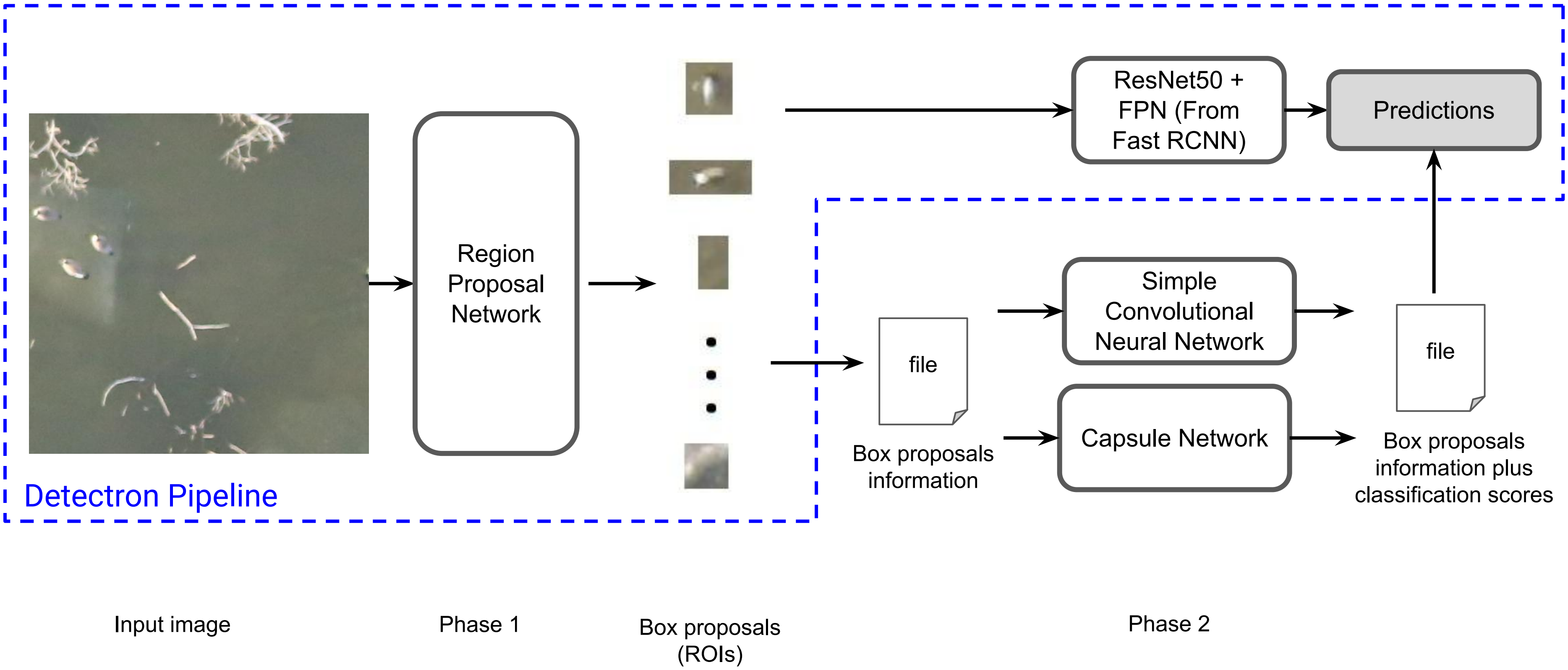
Outline

1. Introduction
2. Related Work
3. **Design and Implementation**
    - **Region Based Object Detectors**
    - **Step1: Region Proposal Generation**
    - **Step 2: Proposals Classification**
4. Experiments
5. Results
6. Conclusions and Future Work

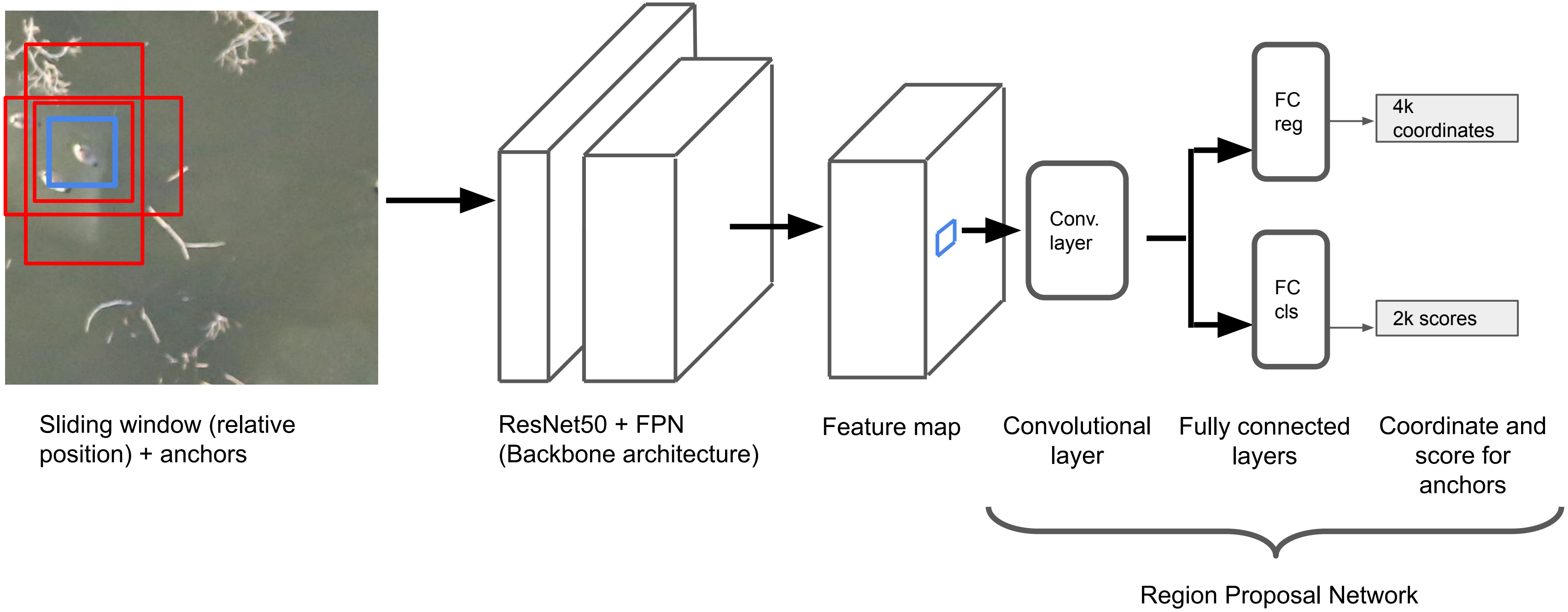
Region Based Object Detectors



Implementation

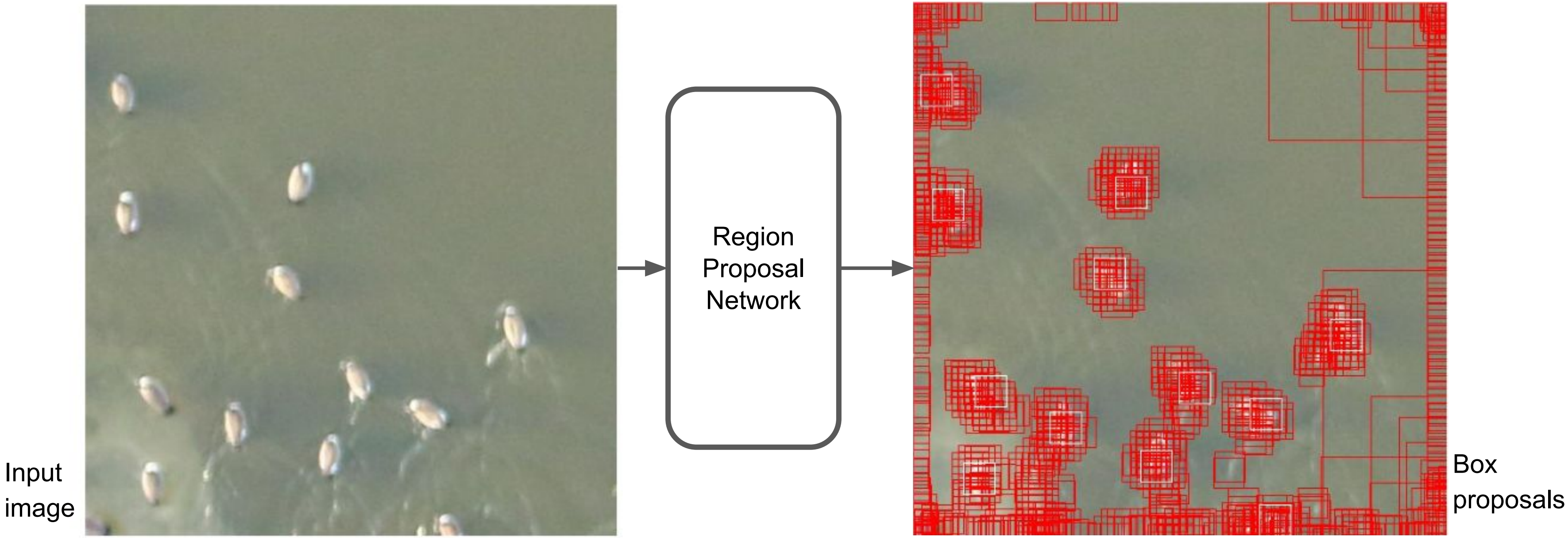


Step 1: Region Proposals Pipeline



Step 1: Region Proposals Samples

- Around 800 region proposals per image
- Contains positive and negative samples that will be used for the training the classification methods.

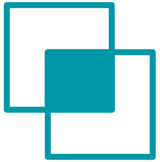


Mini - batches

- No need to use all region proposals. Choose random boxes
- Positive and negative examples according to its Intersection over Union (IoU) with the ground truths

Random samples per mini-batch	128
IoU for positive sample	0.5 or more
Positive / Negative balance	50:50
Augmentation	50% prob. flip image

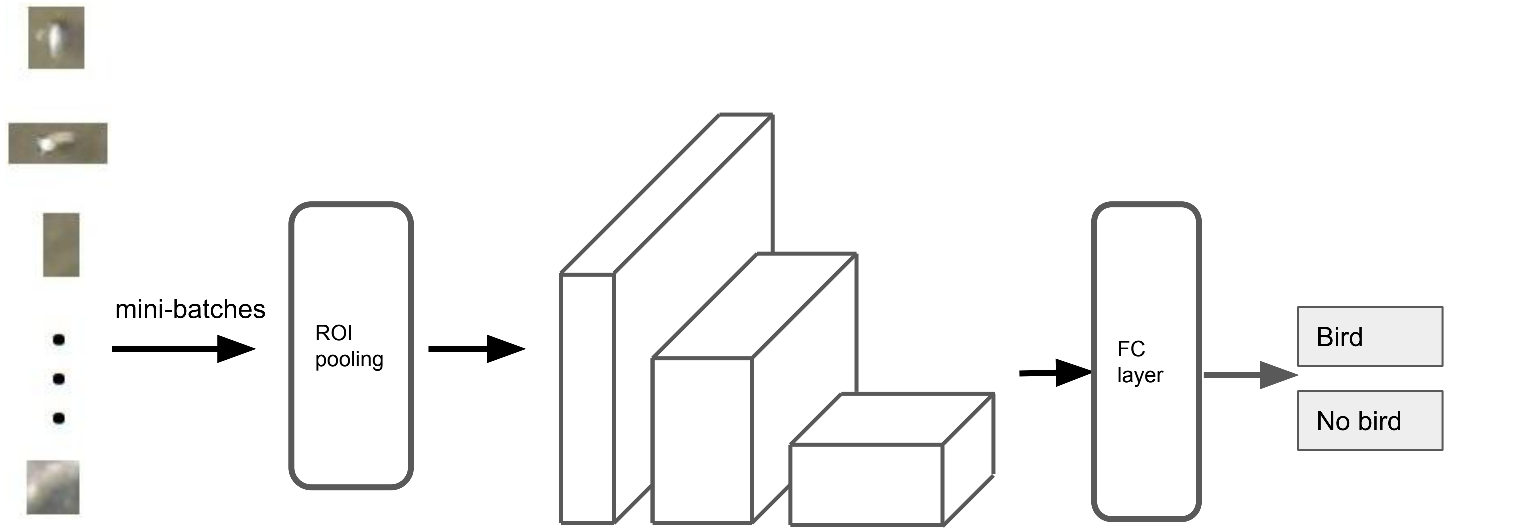
Overlap area =



Area of union =

$$IoU = \frac{Overlap\ area}{Area\ of\ union}$$

Step 2: Classification - ResNet50 + FPN (From Fast RCNN)



Box proposals (ROIs)

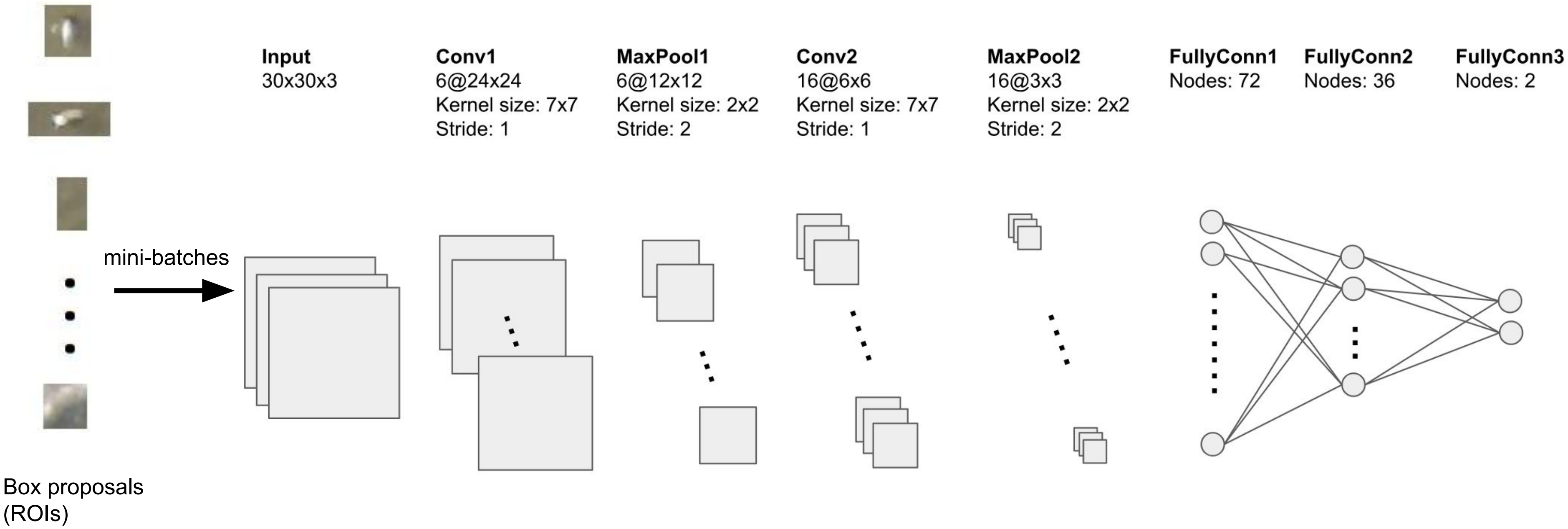
ROI pooling layer (ROI Align)

Backbone architecture (ResNet50 + FPN)

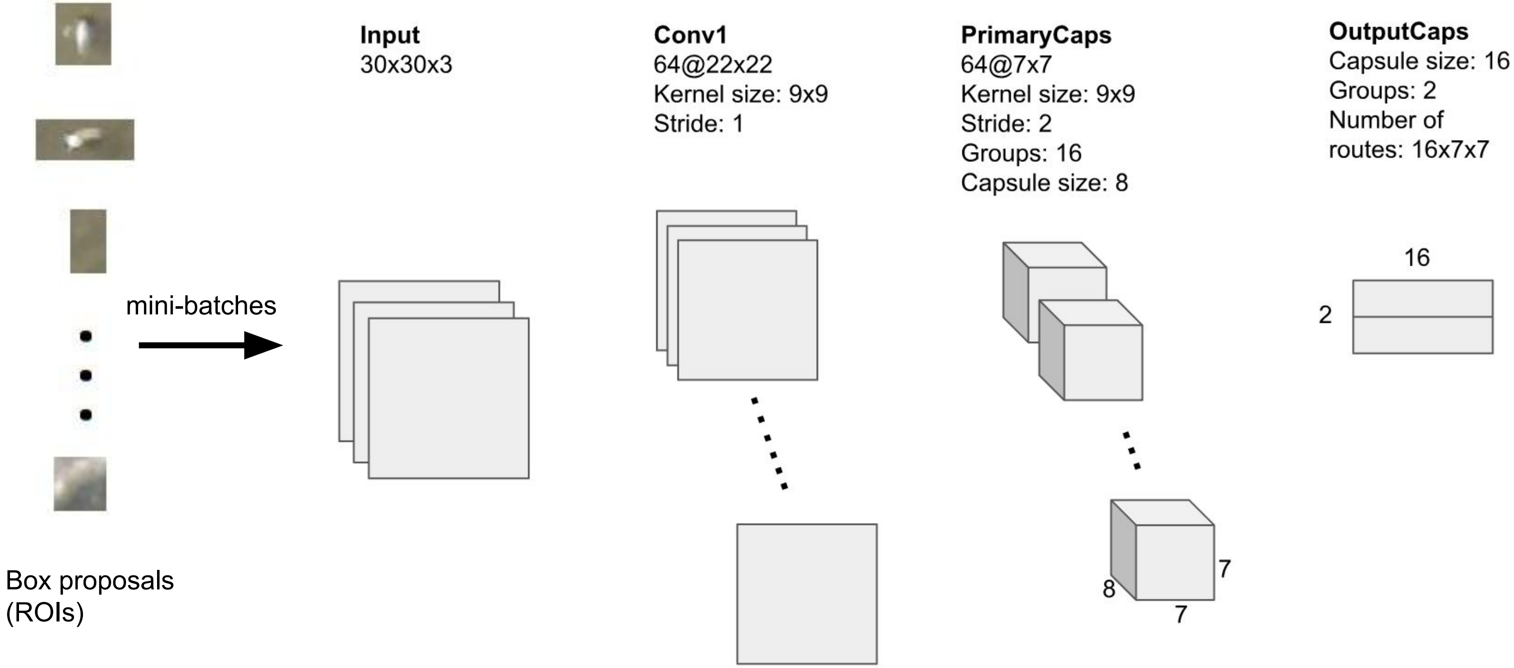
Fully connected layers

Classes prediction (softmax)

Step 2: Classification - Simple Convolutional Neural Network



Step 2: Classification - Capsule Network

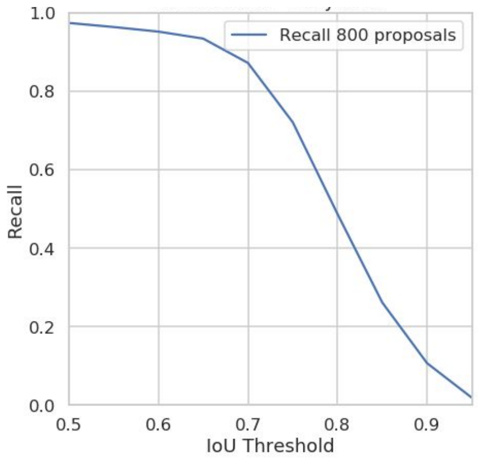


Outline

1. Introduction
2. Related Work
3. Design and Implementation
4. **Experiments**
   - **Step 1: Region Proposal Network Evaluation**
   - **Step 2: Classification Evaluation**
   - **Results in the validation dataset**
5. Results
6. Conclusions and Future Work

Step 1: Region Proposal Network Evaluation

- Region proposals should contain positive and negative samples
- Region proposals should overlap the maximum number of ground truths.



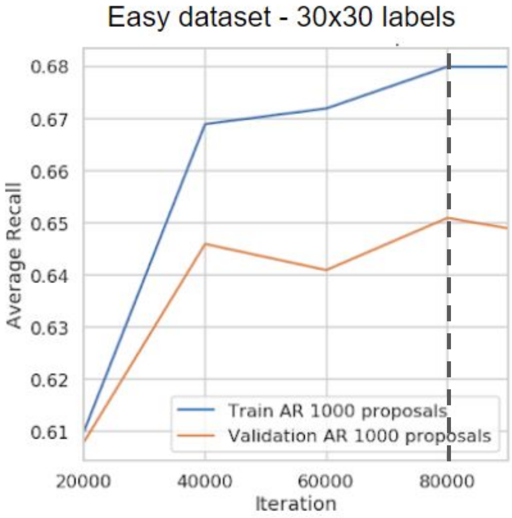
$$Recall = \frac{tp}{tp + fn}$$

Metric: Average Recall

1. Consider true positive if IoU is higher than threshold
2. Calculate Recall at each IoU threshold
3. Calculate the average recall

Step 1: RPN - Model Training

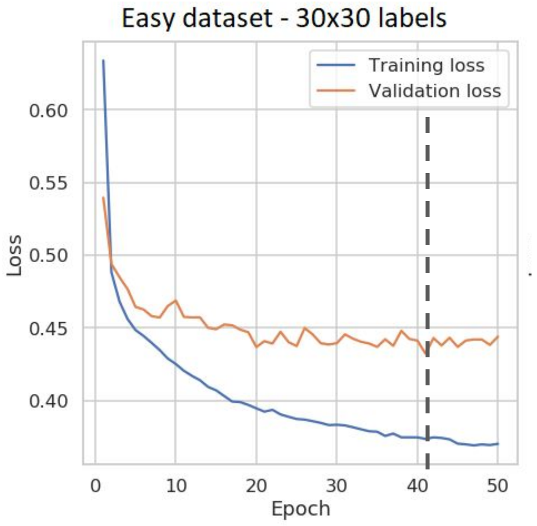
- 30 by 30 labels obtained better average recall scores
- 30 by 30 labels contains more birds.



Dataset Label Size	Dataset Difficulty	Dataset Type	Percentage of birds in boxes	Average recall
20 x 20	Easy	Validation	0.939	0.539
20 x 20	Hard	Validation	0.938	0.543
30 x 30	Easy	Validation	**0.978**	**0.651**
30 x 30	Hard	Validation	0.974	0.630

Step 2: Classification Model Training

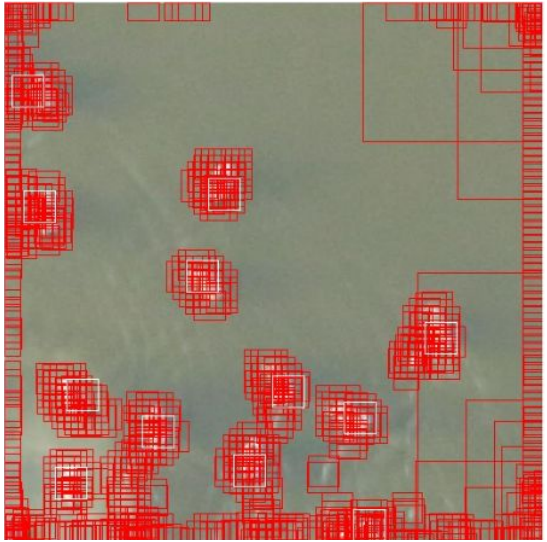
- Select model (weights) with the lowest loss during training
- As in RPN training, 30 by 30 labels and easy datasets have better performance



Dataset Label Size	Dataset Difficulty	Dataset Type	Classification method	Minimum Validation Loss
20 x 20	Easy	Validation	Simple CNN	0.4830
20 x 20	Hard	Validation	Simple CNN	0.4535
30 x 30	Easy	Validation	Simple CNN	**0.4323**
30 x 30	Hard	Validation	Simple CNN	**0.3923**

Step 2: Classification Evaluation

- After making predictions each region proposal will have a “bird score” between 0 and 1
- Remove region proposals **overlapped** and with a **low score**
- Precision, Recall and F1 metrics to measure performance



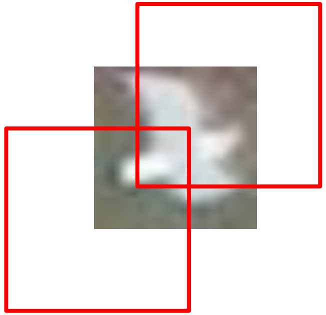
$$Precision = \frac{tp}{tp + tn}$$

$$Recall = \frac{tp}{tp + fn}$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Remove Overlapped Boxes

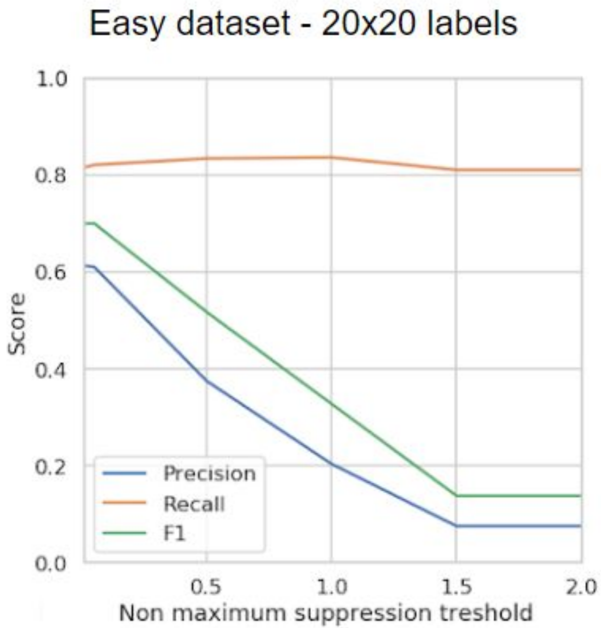
- Non Maximum Suppression after classification. Keep boxes with higher scores



RP1: Score: 0.91

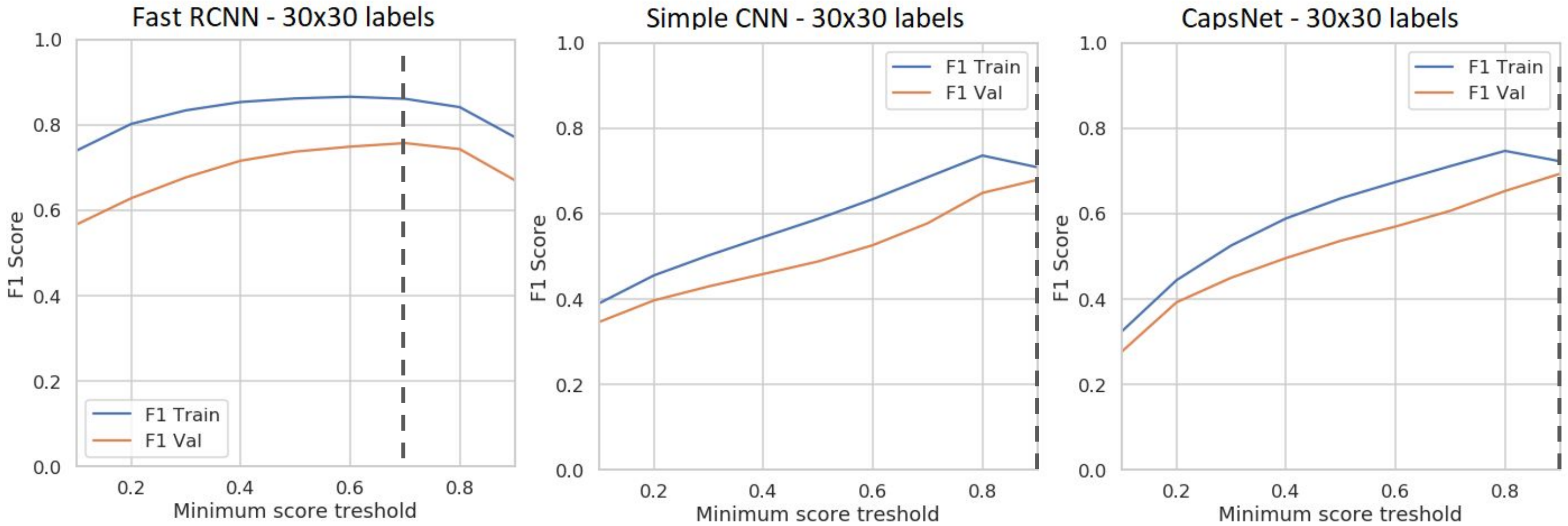
RP2: Score: 0.55

- For this dataset, birds are not highly overlapped.
- F1 increases if non-maximum suppression **removes all overlapped boxes**.



Remove Boxes with low score

- For each classifier, find a threshold that maximizes the F1 score in the validation dataset
- Higher threshold means high Recall (tends to recognize backgrounds as birds)



Capsule Networks - Reconstruction

- Reconstruct the input image from the output capsule vector using a reconstruction network
- This network did not produced a significant improvement

- Difficulty to recognize the shape due to the high variety of bird shapes

Training samples

Results in Validation Dataset

- 30 by 30 labels datasets obtained better scores
- Fast RCNN achieved better scores.

Easy Dataset - 30x30 labels

Classification method	Validation Precision	Validation Recall	Validation F1 score
Fast RCNN	0.7799	0.7364	**0.7575**
Simple CNN	0.7207	0.6422	0.6792
CapsNet	0.6885	0.6984	0.6934

Hard Dataset - 30x30 labels

Classification method	Validation Precision	Validation Recall	Validation F1 score
Fast RCNN	0.6461	0.6276	**0.6367**
Simple CNN	0.5594	0.6502	0.6014
CapsNet	0.4590	0.7565	0.5714

Outline

1. Introduction
2. Related Work
3. Design and Implementation
4. Experiments
5. **Results**
    - **Results on test dataset**
    - **Visualizations**
6. Conclusions and Future Work

Results - Easy Test Dataset

- Fast RCNN classification achieved better results
- Using validation threshold helped to improve F1 test score.
- Simple CNN and Capsule Networks tends to have High Recall

Easy Dataset - 30x30 labels

Classification method	Validation Precision	Validation Recall	Validation F1 score
Fast RCNN	0.9249	0.8809	**0.9024**
Simple CNN	0.9497	0.7335	0.8277
CapsNet	0.9315	0.7964	0.8587

Hard Dataset - 30x30 labels

Classification method	Validation Precision	Validation Recall	Validation F1 score
Fast RCNN	0.6804	0.6492	**0.6644**
Simple CNN	0.4412	0.6033	0.5097
CapsNet	0.3631	0.6000	0.4524

Visualizations - Easy Image 1

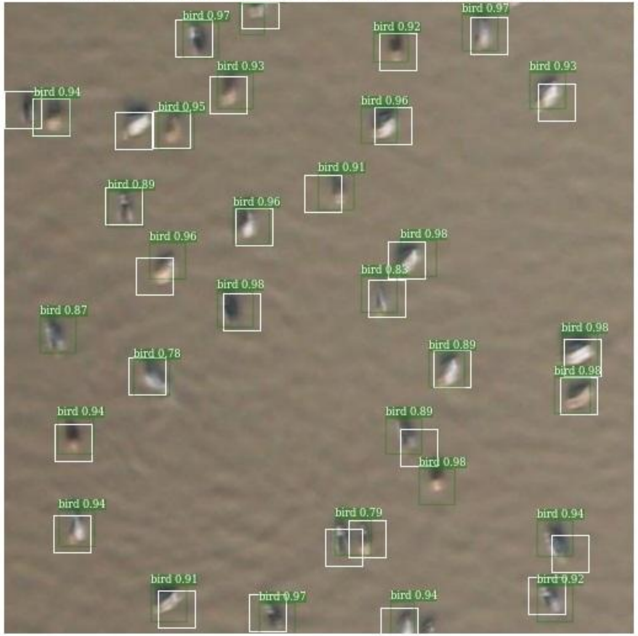
Easy image 1 - Fast RCNN classification

Easy image 1 - Simple CNN classification

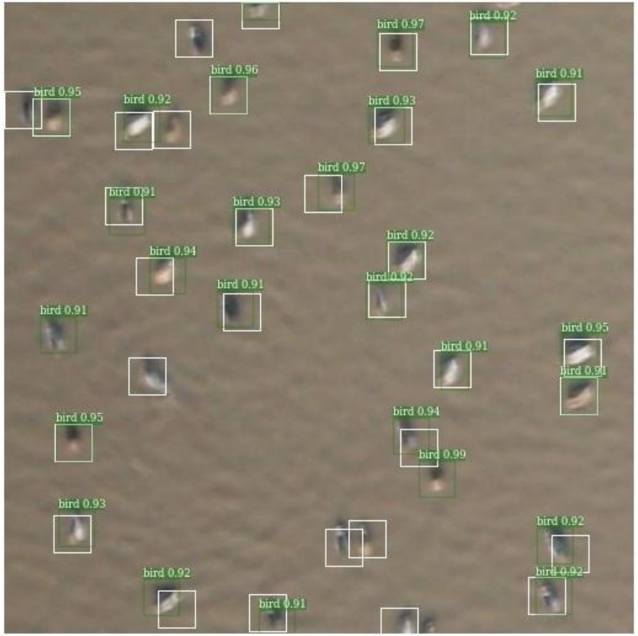
Easy Image 1 - CapsNet classification

Visualizations - Easy Image 2

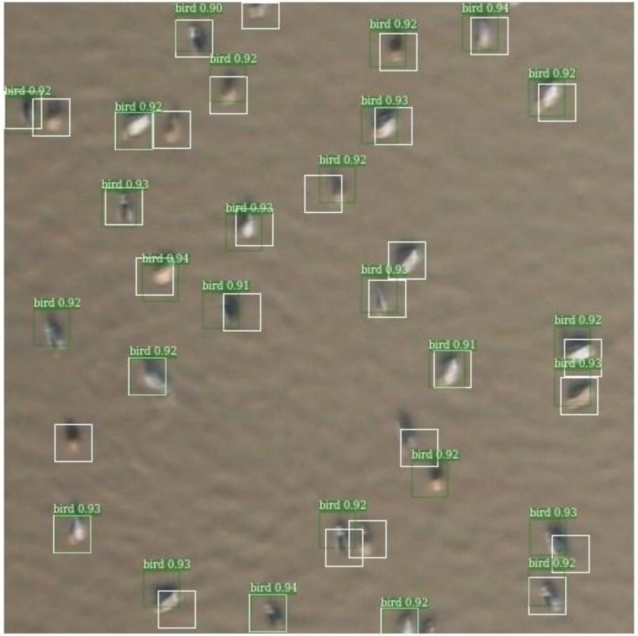
Easy image 2 - Fast RCNN classification



Easy image 2 - Simple CNN classification

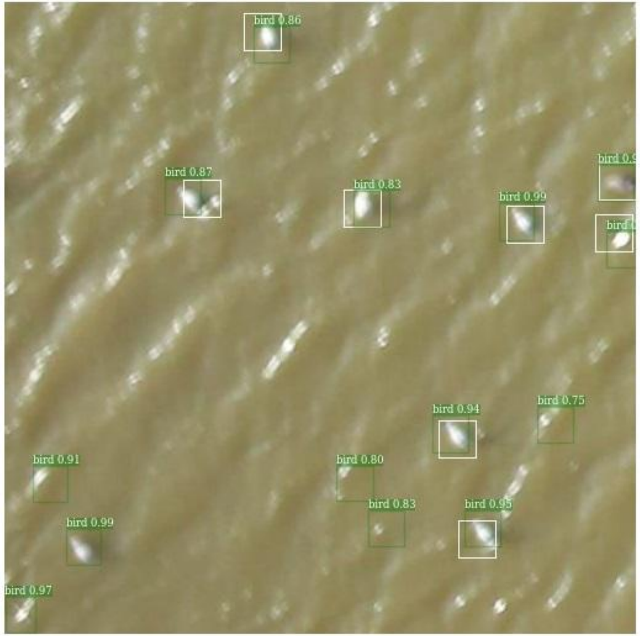


Easy image 2 - CapsNet classification



Visualizations - Hard Image 1

Hard image 1 - Fast RCNN classification



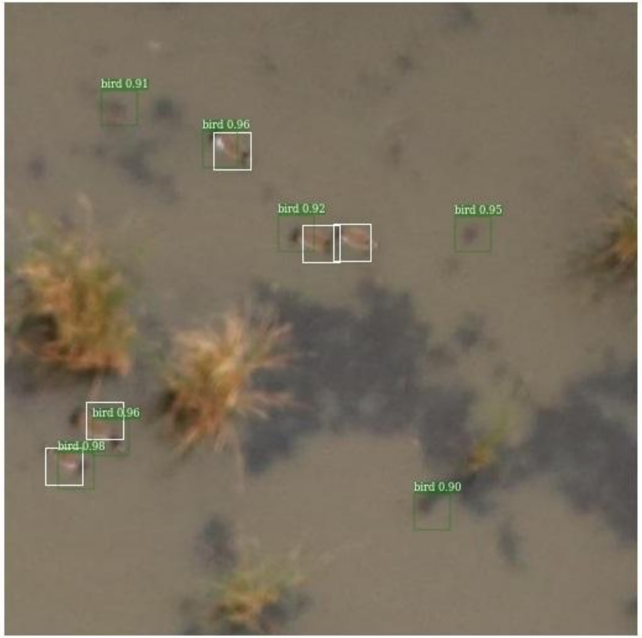
Hard image 1 - Simple CNN classification

Hard image 1 - CapsNet classification

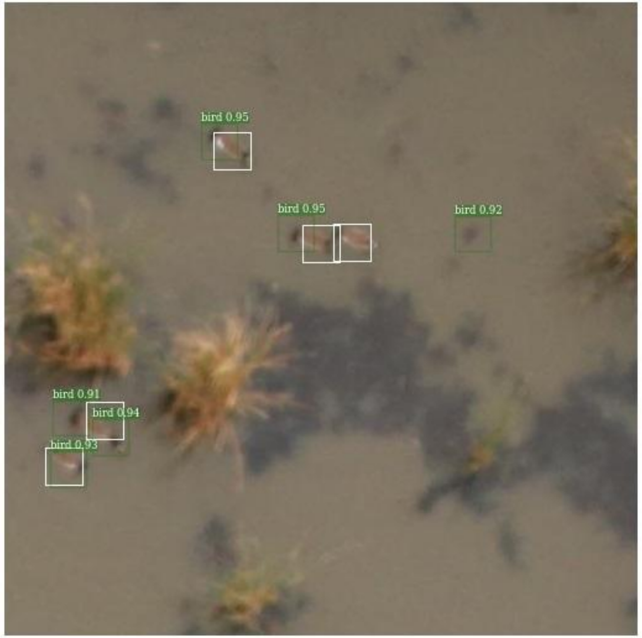
Visualizations - Hard Image 2

Hard image 2 - Fast RCNN classification

Hard image 2 - Simple CNN classification



Hard image 2 - CapsNet classification



Outline

1. Introduction
2. Related Work
3. Design and Implementation
4. Experiments
5. Results
6. **Conclusions and Future Work**
    - **Conclusions**
    - **Future Work**

Conclusions

- 30 by 30 labels achieved better scores. This bounding box size captures more information of the birds
- RPN does works for finding the regions of interest, including most of the birds.
- With the proposed pipeline the box proposals are available for any other classifier
- All overlapped region proposals must be removed for this dataset
- Classifiers tends to have a high recall. All score metrics increases when increasing the minimum score threshold.

Conclusions

- Fast RCNN classification achieved the best F1 score in all the datasets. Simple CNN and Capsule Networks achieved similar scores
- Fast RCNN are the results of applying many advanced techniques like Feature Pyramid Network.
- The high variety of birds shapes made capsule network unable to learn instantiation parameters features.
- CapsNet should work better with more classes and no backgrounds

Future work

- Accurate bounding boxes.
- Add more labels. For example, for flying birds.
- Matrix capsules with EM routing: instead of using vectors to represent the instantiation parameters of an object, use matrices that allows a more complex representation of an object

University of Missouri

Thank You

Questions