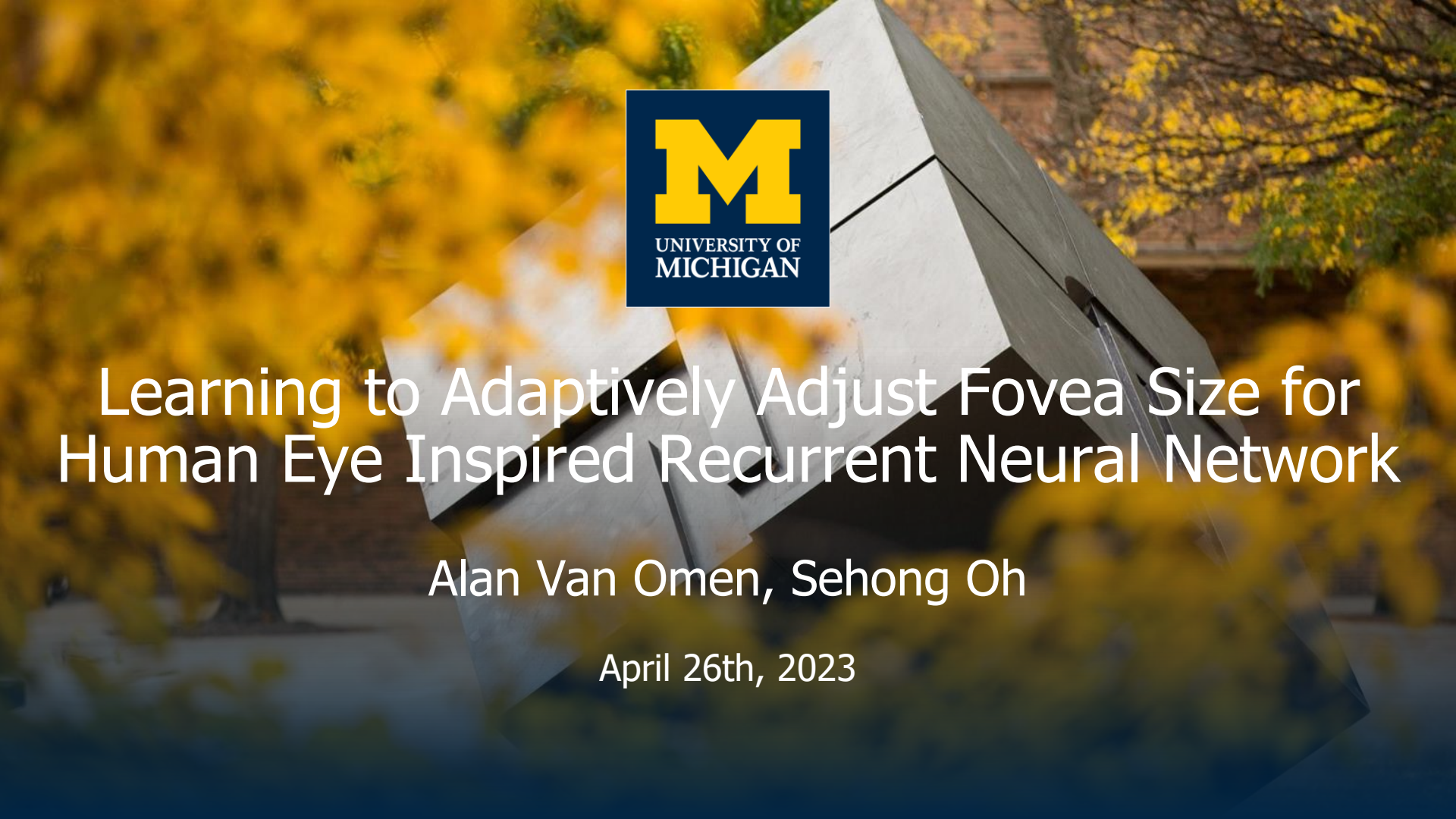


Learning to Adaptively Adjust Fovea Size for Human Eye Inspired Recurrent Neural Network

Alan Van Omen, Sehong Oh

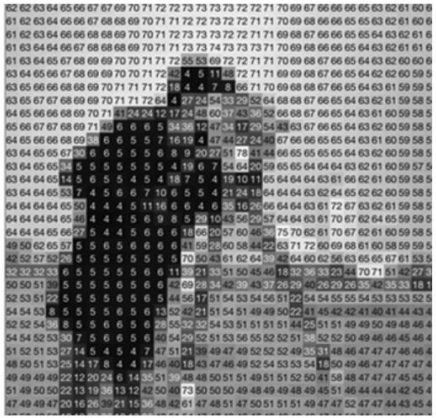
April 26th, 2023

Outline

1. Introduction
2. Motivation & Related Works
3. Methodology
4. Experiments
5. Conclusion

Introduction

1. Human vision and computer vision
   - Computer vision has a different way to see the world from humans
   - Each system has its own pros and cons, and "capacity" and “vulnerability” are the distinct properties that differentiates them



Motivation

1. Computer vision models (CNNs) are vulnerable to adversarial noises [1]

2. Human eyes are more robust than computer vision

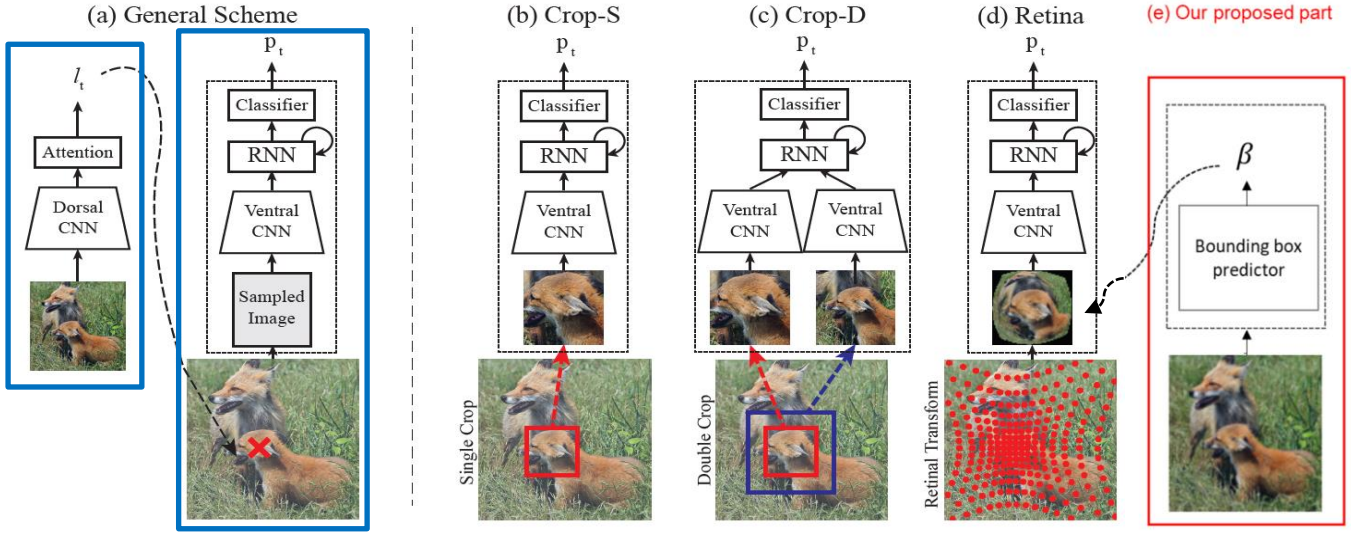
   - Allowing models to fixate to different image regions can alleviate the effect of adversarial noise [2]
   - Non-uniform spatial sampling and varying receptive fields that mimic the retinal transformation in the primate retina can also improve the robustness against adversarial attacks [3]

[1] Aleksander et al. Towards deep learning models resistant to adversarial attacks, 2017.
[2] Ricardo Gattass et al. Visual topography of v2 in the macaque, 1981.
[3] Pouya Bashivan et al. Neural population control via deep image synthesis, 2019.

Methodology

1. Dorsal Stream: or the "where" pathway is responsible for processing information related to spatial awareness and motion.

2. Ventral stream: or the "what" pathway, is responsible for object recognition, face recognition, and determining the color and shape of objects.



[4] Choi, Minkyu, et al. Human Eyes Inspired Recurrent Neural Networks are More Robust Against Adversarial Noises, 2022

Improvement

1. The hyper-parameter *b* in the model controls the degree of non-uniform sampling for the ventral stream (authors use constant b=12)

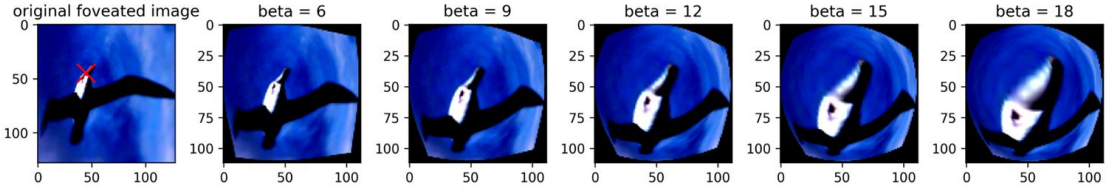
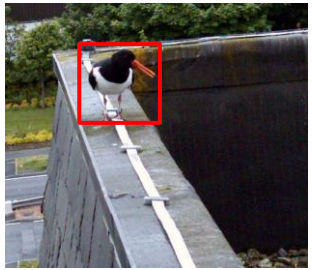
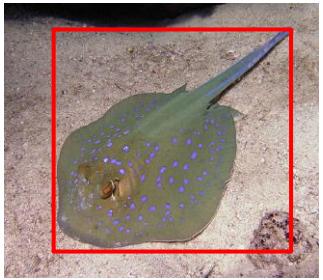


Figure 2. Examples of the effect of *b* on the retinal transformation, with fixation point centered at red x.

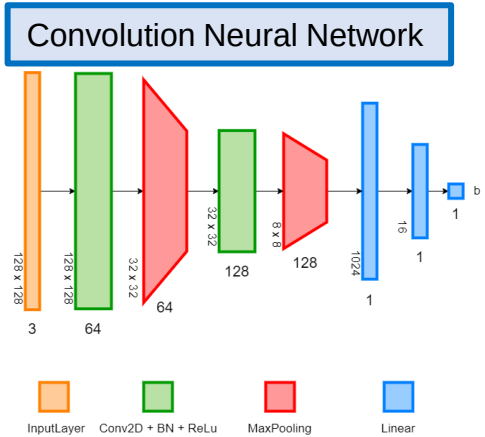
2. We implement 3 methods for adaptively changing the fovea size to improve robustness

Bounding Box Methods (YOLOv5)



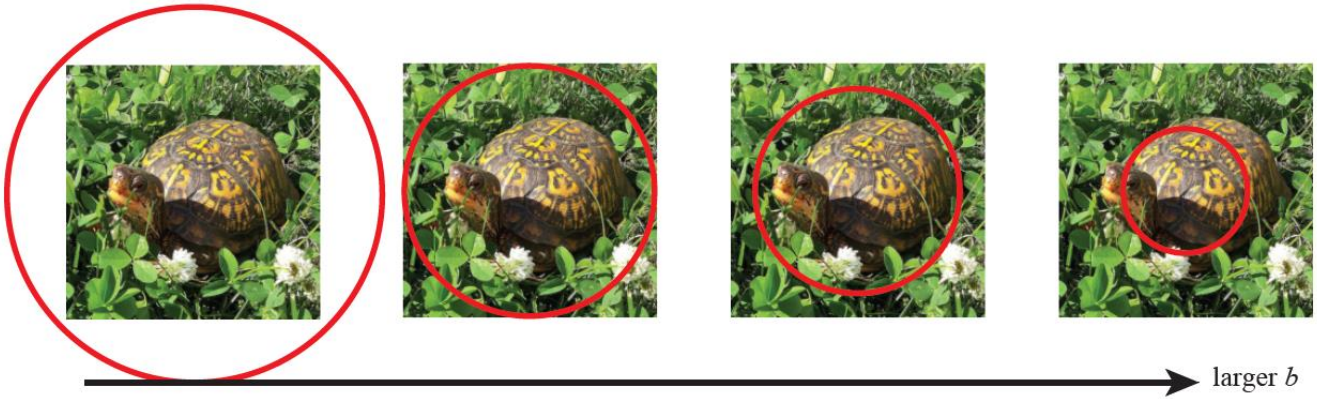


Convolution Neural Network



Effect of b on the fovea size

(a) Size of Fovea



(b) Retinal images

Improvement

1. The hyper-parameter *b* in the model controls the degree of non-uniform sampling for the ventral stream (authors use constant b=12)

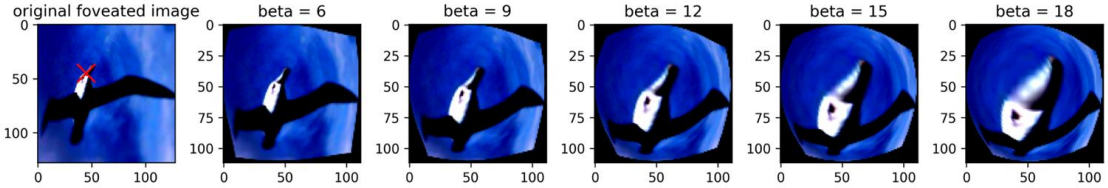
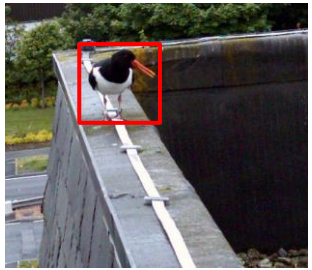
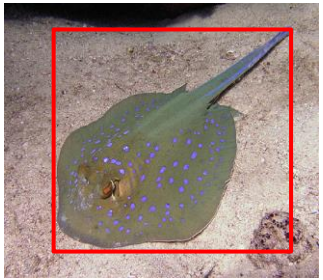


Figure 2. Examples of the effect of *b* on the retinal transformation, with fixation point centered at red x.

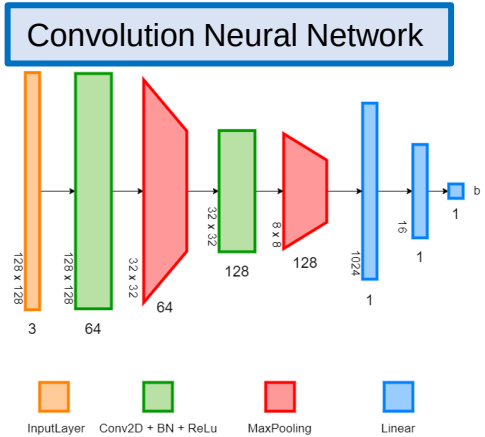
2. We implement 3 methods for adaptively changing the fovea size to improve robustness

Bounding Box Methods (YOLOv5)





Convolution Neural Network



Experiments

- Trained four models using various strategies for adapting the fovea size
  1. **b=12**: held beta constant like original paper
  2. **B-CNN**: learned beta concurrently in an end-to-end manner
  3. **C-BB**: estimated beta from the closest valid bounding box
  4. **L-BB:** estimated beta from the largest valid bounding box

- Learned adversarial noise with 100 steps of PGD over 1600 test images

Parameter	Value
batch size	4
optimizer	Adam
initial learning rate	1e-3
epochs	10
retinal sampling grid size	112
time steps (1 fixation per step)	4
default b	12

		ASR ($\epsilon = 5e-3$)	
Model	Top-1 Acc	Targeted	Untargeted
$b = 12$	35.9%	80.0%	80.2%
b-CNN	38.0%	79.6%	82.9%
C-BB	**39.8%**	78.9%	80.6%
L-BB	28.6%	**78.4%**	**78.9%**

Conclusion

1. Successfully implemented 3 methods for adaptively setting the fovea size
2. Found that C-BB gave slightly higher robustness and better accuracy

Future work

1. Explore how to fine-tune bounding box model on target dataset.
2. Optimize the model efficiency and memory usage to allow for larger batch sizes and better stability.
3. Explore using some reward to promote robustness for learning beta in the end-to-end manner.

Reference

[1] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. arXiv preprint arXiv:1706.06083, 2017.

[2] Ricardo Gattass, CG Gross, and JH Sandell. Visual topography of v2 in the macaque. Journal of Comparative Neurology, 201(4):519–539, 1981.

[3] Pouya Bashivan, Kohitij Kar, and James J DiCarlo. Neural population control via deep image synthesis. Science, 364(6439), 2019.

[4] Choi, Minkyu, et al. "Human Eyes Inspired Recurrent Neural Networks are More Robust Against Adversarial Noises." *arXiv preprint arXiv:2206.07282* (2022).

Q & A

Thank you !