Penguin Monitoring: Can Transfer Learning Facilitate Context-Agnostic Detection

RHODES UNIVERSITY
Where leaders learn

Dane Brown
Karen Bradshaw

South African
NATIONAL PARKS

Importance of Penguin Monitoring

African penguins face severe population decline due to various factors such as habitat destruction, climate change, and food scarcity.

An endangered indicator species, reflecting the health of the Benguela ecosystem.

Effective monitoring informs conservation policies and habitat protection measures.

Sherley, R. B., Underhill, L. G., Barham, B. J., Barham, P. J., Coetzee, J. C., & Crawford, R. J. M. (2020). Defining ecologically relevant scales for spatial protection using long-term tracking data from an endangered seabird. Ecological Applications, 30(5), e02106.

Crawford, R. J. M., Whittington, P. A., Upfold, L., Ryan, P. G., Petersen, S. L., Dyer, B. M., & Cooper, J. (2017). Recent trends in numbers of African Penguins Spheniscus demersus at Robben Island, South Africa. African Journal of Marine Science, 39(3), 269-277.

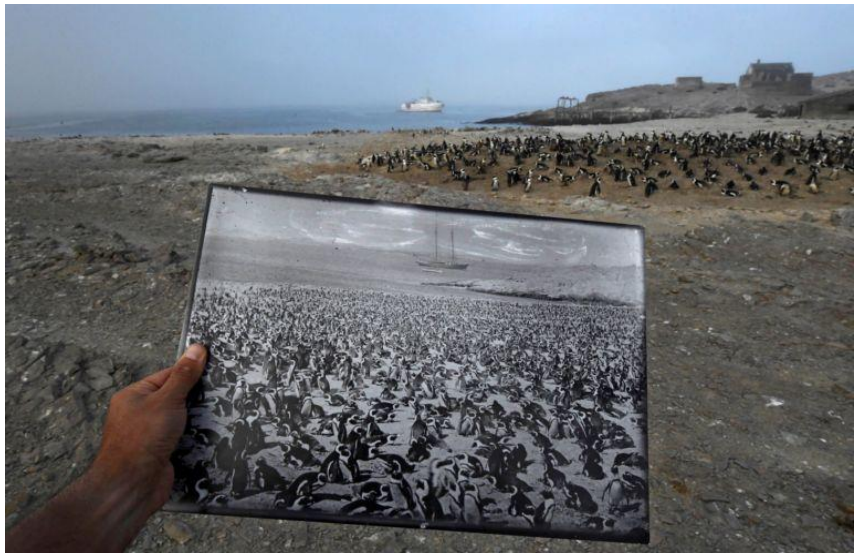
Exploration Methods

Technological advancements in UAV/cameras contribute to our understanding of habitats and can be used to monitor the challenges for intervention

Rachael Herman, Stony Brook University/LSU

Ceramic nests in Algoa Bay (2022)

Environments



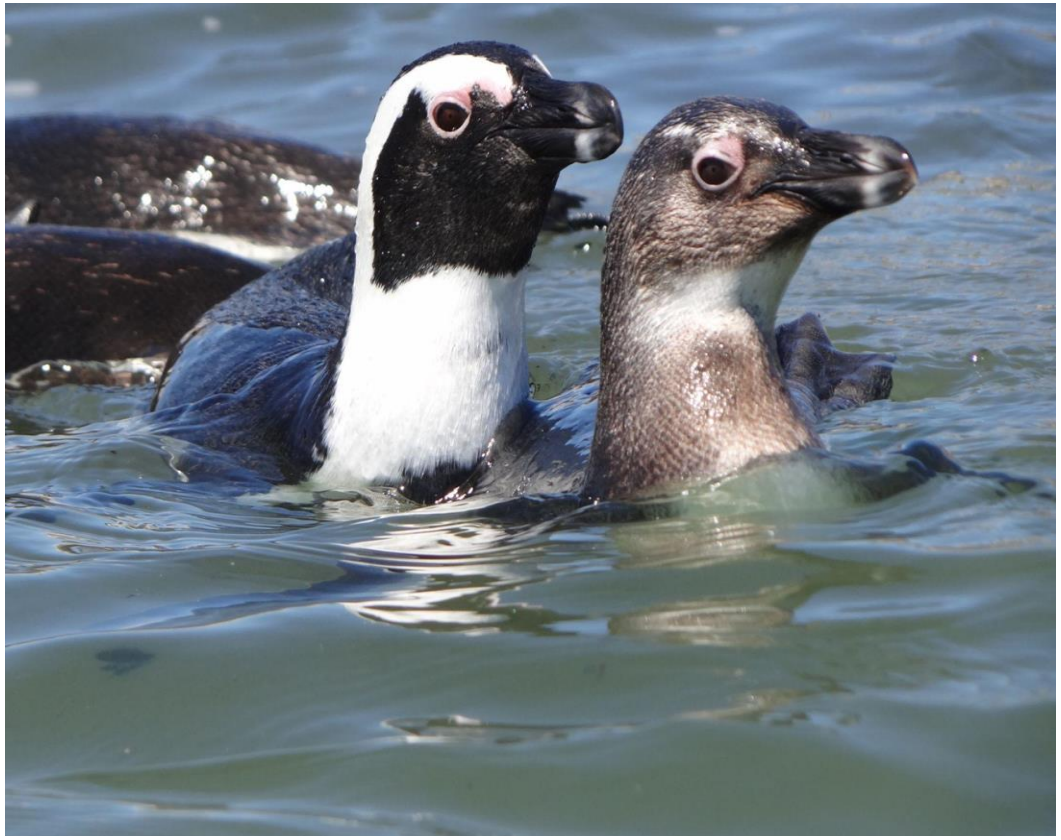
Thomas P. Peschak – National Geographic

African penguin – Bristol Zoo

African penguin – SANParks

African penguin – Daily Maverick

Environments



Adult and Juvenile African penguin

African penguins

Automatic or Assisted Penguin Monitoring

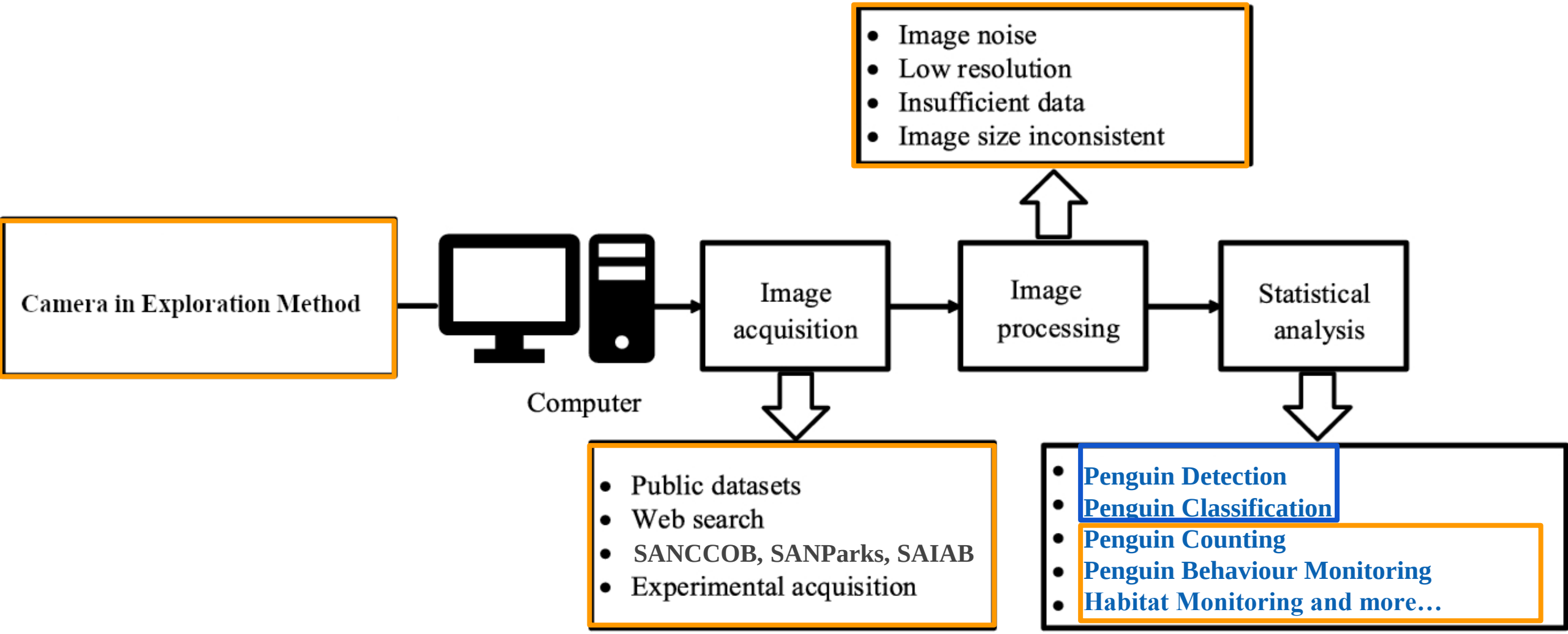
Computer vision techniques have rapidly evolved and are now applied in various fields, offering new ways to various wildlife.

Cameras and other sensors allows wildlife monitoring following relatively simple but accurate data collection

Computer vision allows for automated **detection** and **classification** of species towards their assessments and of their habitats

Van den Berg, C. P., Rogers, A. C., & Burghardt, T. (2021). Advances in computer vision for wildlife monitoring and conservation. Biological Conservation, 253, 108844.

Computer Vision: Detection and Classification



But we will focus on detection and classification for now

History of Object Detection Models

Traditional Object Detection Models – prehistoric age

- Early object detection models employed Haar-like features and AdaBoost popularized by face detection tasks
- The introduction of Histogram of Oriented Gradients (HOG) brought advancements in detecting humans and other objects of interest

Deep Learning Detection Models

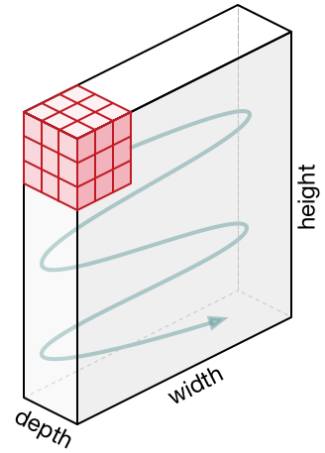
- Region-based Convolutional Neural Networks (R-CNN) – 2014
- Single Shot Detectors (SSD) – 2016

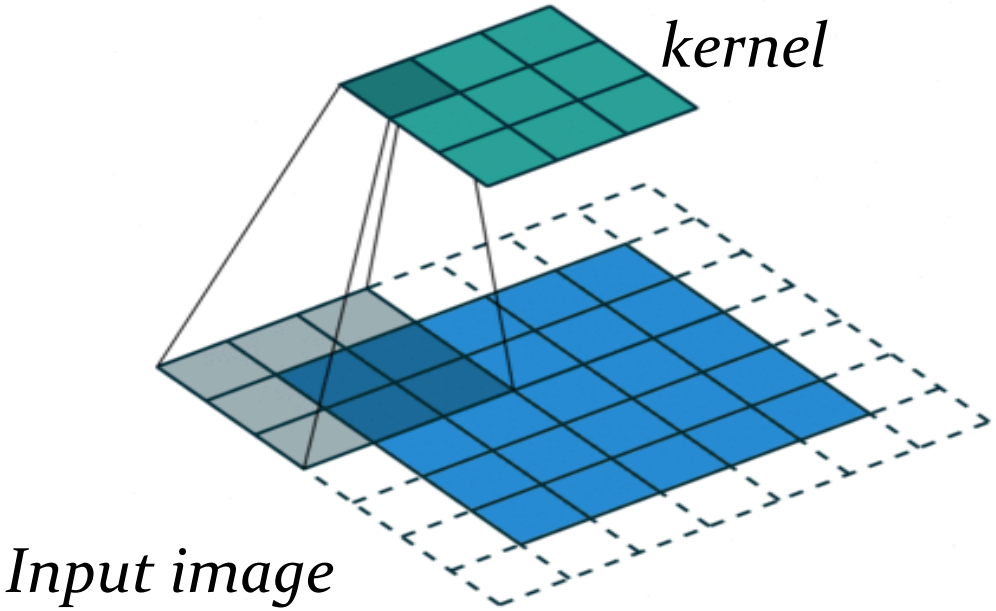
Adouani, A., Henia, W. M. B., & Lachiri, Z. (2019, March). Comparison of Haar-like, HOG and LBP approaches for face detection in video sequences. In *2019 16th International Multi-Conference on Systems, Signals & Devices (SSD)* (pp. 266-271). IEEE.

Convolutional Neural Network Crash Course

Convolution Layer

- Kernel shifted from left to right – top to bottom until it covers whole input image
- Matrix multiplication at each shift – sum is the convoluted output.





Convoluted Feature Output:

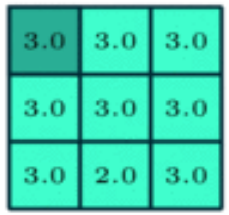
- Conv 1: e.g. blobs and edges
- Conv 2: e.g. pattern of body/flippers
- Conv 3: e.g. shape
- … e.g. flipper location

Convolutional Neural Network Crash Course

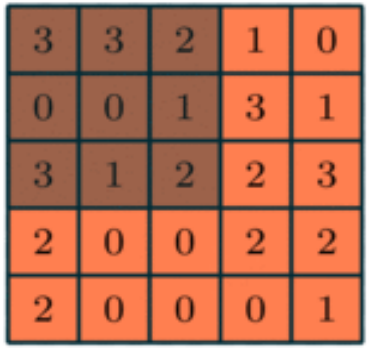
Max Pooling Layer

Reduce the spatial size of the convoluted output from the layer before it.

- Computational power required to process the data decreases
- Acts as noise suppressant



Max pooling output



Convolved feature

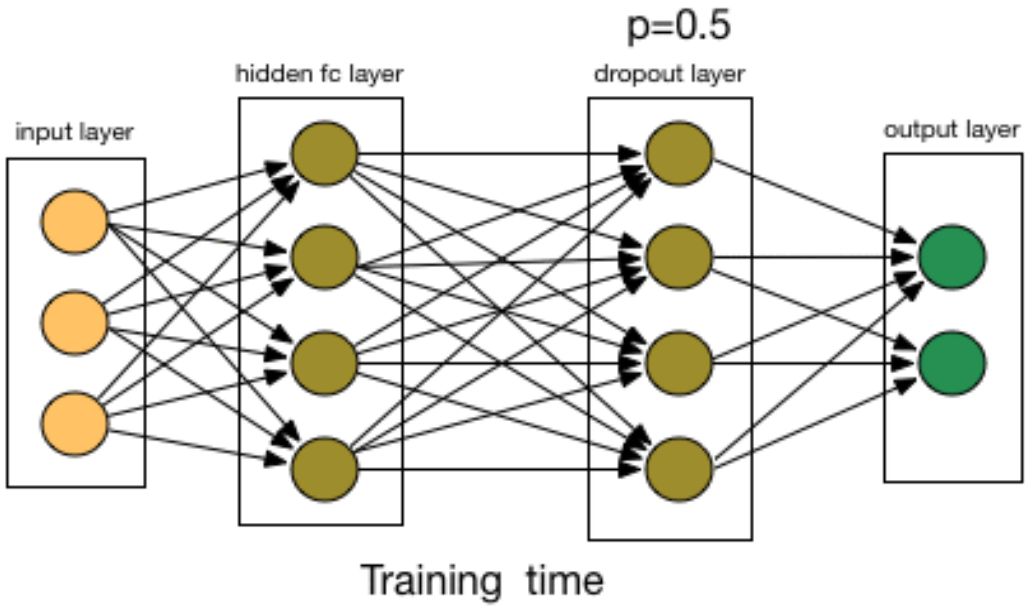
Maximum value from the portion of the image covered by the Kernel

Convolutional Neural Network Crash Course

Dropout Layer

Randomly selected neurons are removed during training to mitigate overfitting.

- Removal happens on the forward pass.
- Weight updates on the backward pass are no longer performed.



Regularized output: generalize well from the training data and make good predictions for unseen data

Convolutional Neural Network Crash Course

- **Flatten Layer**: Input from prior layer (2D) flattened to a one-dimensional column vector

- **Fully Connected/Dense Layer**: Features are connected together through weights and a non-linear activation function (ReLu) is applied on it.
  - Contains trainable weights
  - updated using backpropagation

- **Output Layer/SoftMax:** Uses probability distribution to predict input image's class.

Object Detection: Deep Learning

Region-based Convolutional Neural Networks (R-CNN) pioneered deep learning object detection models, with variations like Fast R-CNN and Faster R-CNN

Single Shot MultiBox Detectors (SSD) integrated object detection and classification in a single forward pass, accelerating detection tasks. However, its accuracy limitations were quickly superseded by You Only Look Once

You Only Look Once (YOLO) originally by Redmon (2016) represented a breakthrough in end-to-end real-time object detection, optimizing speed without compromising accuracy

Van den Berg, C. P., Rogers, A. C., & Burghardt, T. (2021). Advances in computer vision for wildlife monitoring and conservation. *Biological Conservation, 253*, 108844

Object Detection: YOLO

Newer YOLO versions eventually excelled beyond urban environments and can detect, classify, track etc. to help address wildlife monitoring challenges:

- complex backgrounds
- occlusion
- sizes
- shapes
- dynamic lighting
- tracking
- counting
- behaviours
- several other combined factors

However, it is non-trivial to adapt object detectors to work effectively across environment

Object Detection: Current State-Of-The-Art

YOLO-NAS and YOLOv 8, 9, 11 represent state-of-the-art object detection models, with various architecture and optimization strategies

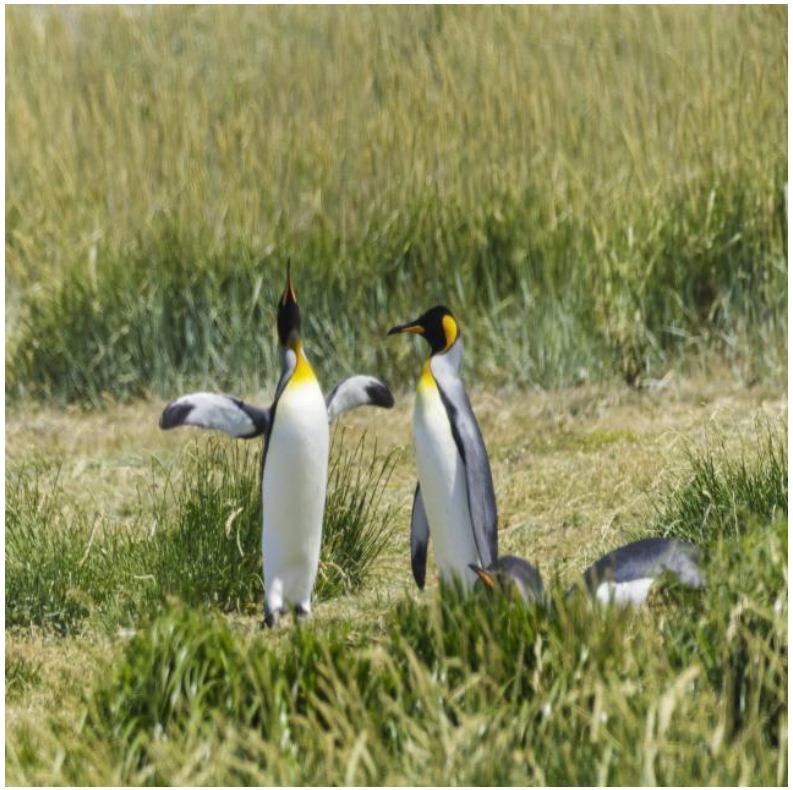
Optimize for real-time high-precision object detection in complex environments

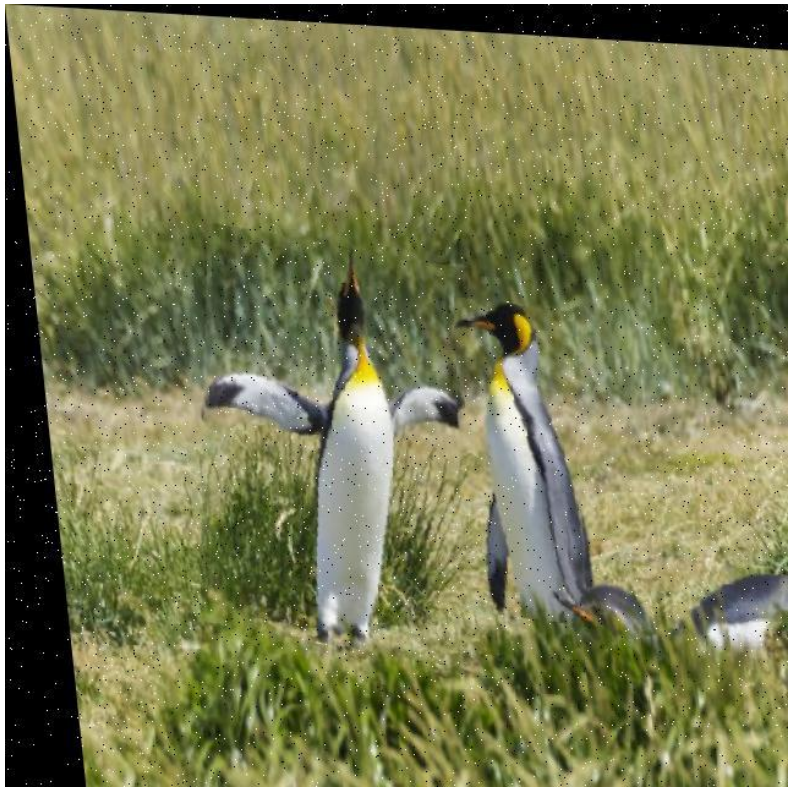
YOLO-NAS automatically chooses the DL architecture but suffers from low precision

YOLOv 8, 9, 11 have different advantages for wildlife detection and classification

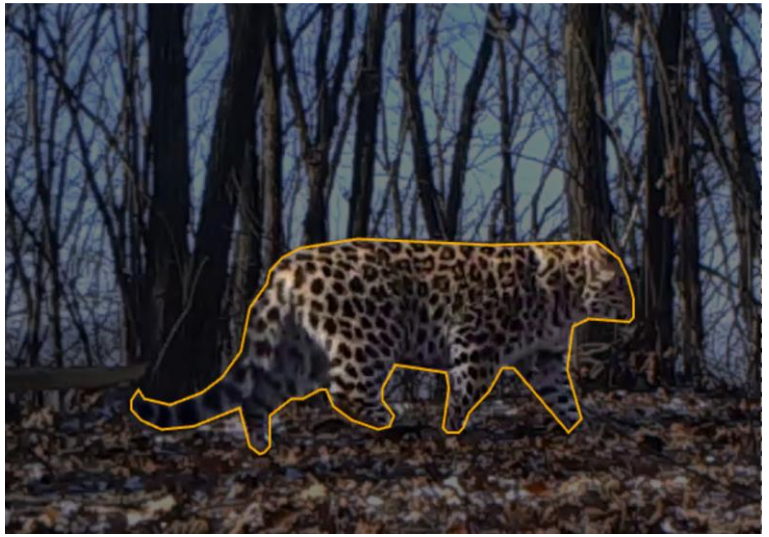
Data Annotation

Offline Augmentation

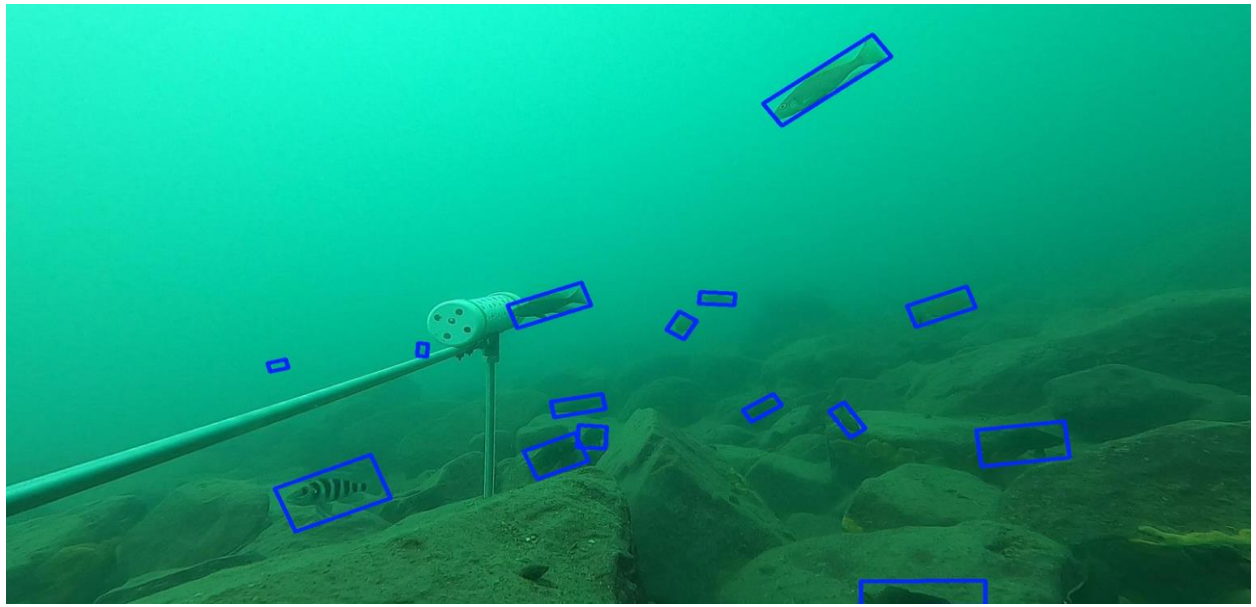


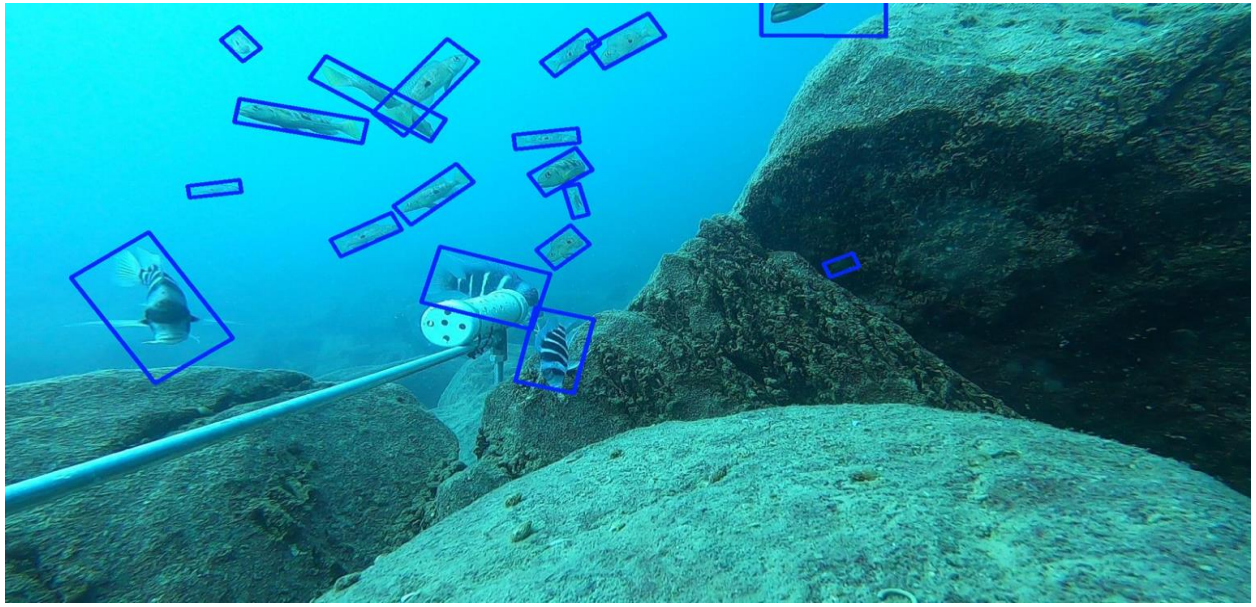


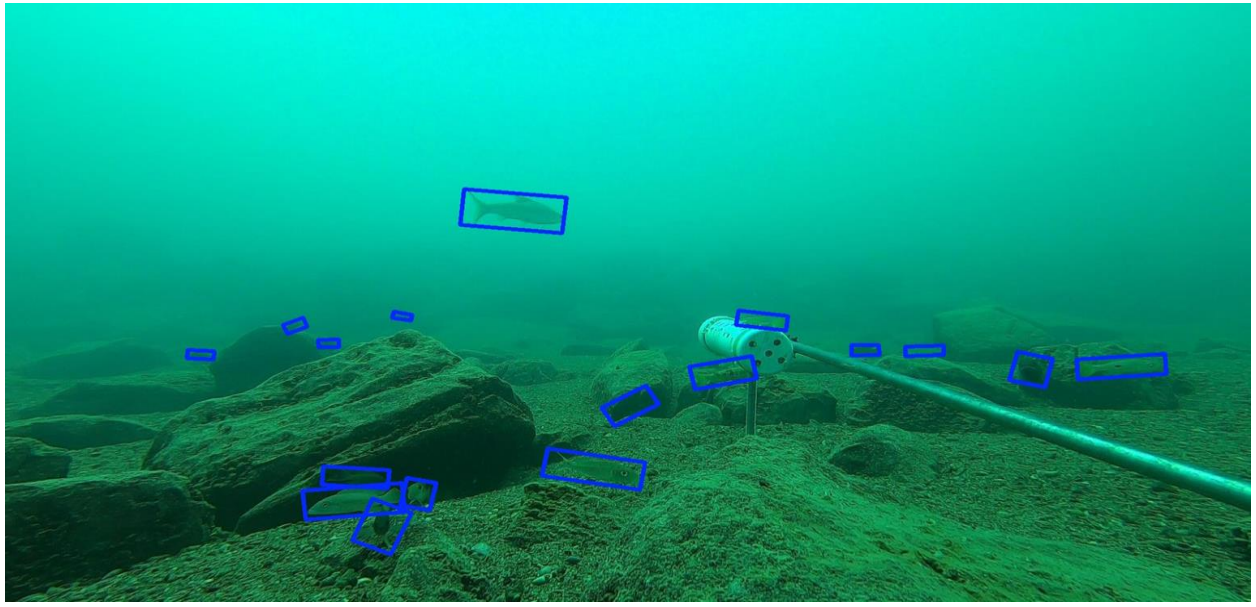
Wildlife can be used for Transfer Learning

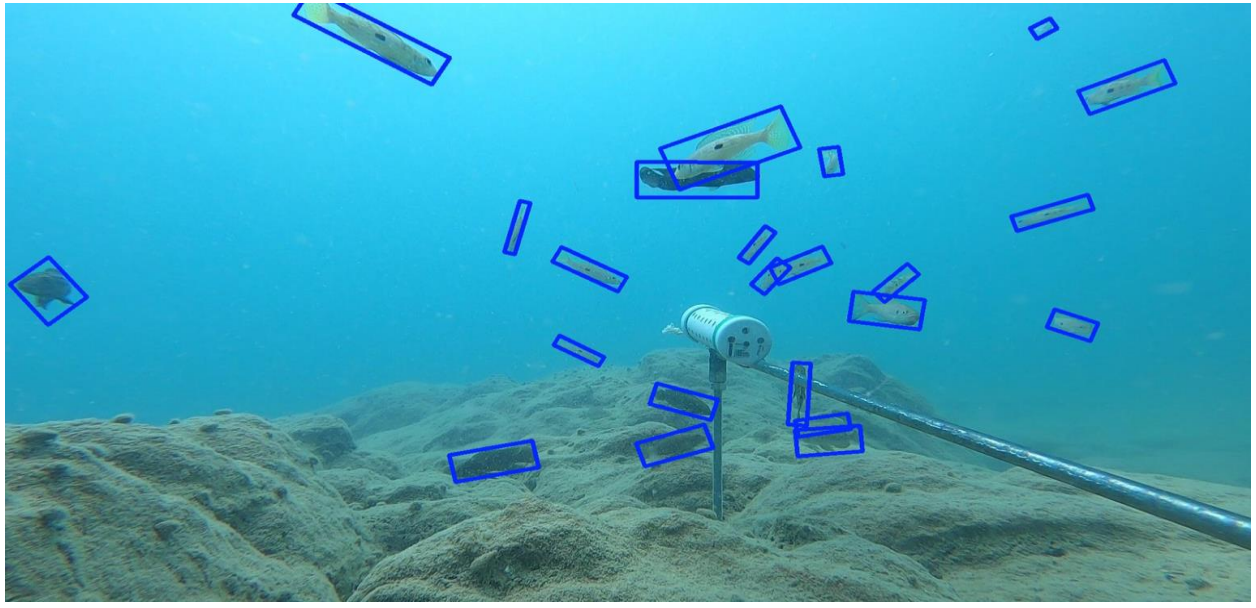


Wildlife can be used for Transfer Learning

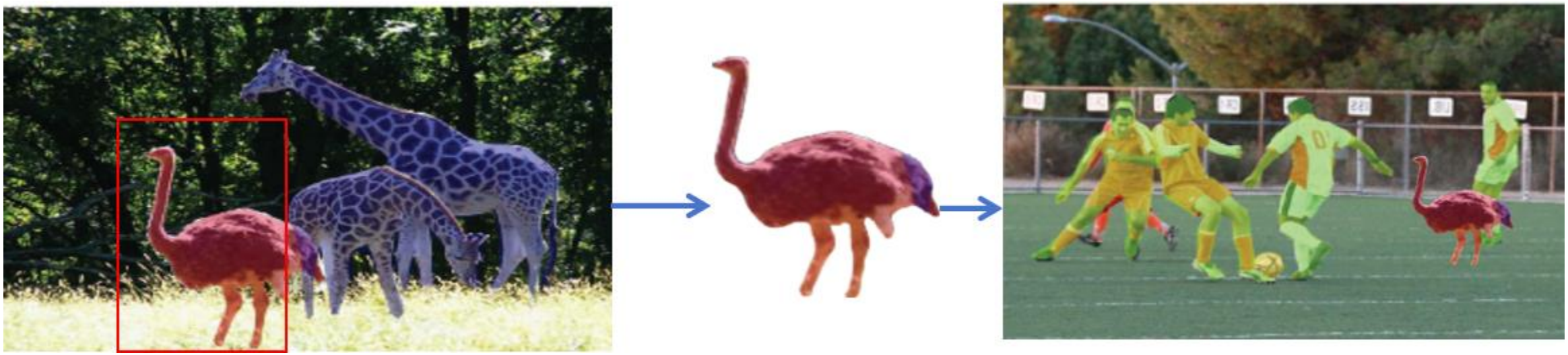








Transfer Learning from Copy-Paste Environments



Experimental Setup

- Public Marine Aerial Dataset – 600 Images
- 70:20:10 split for train, validation and test sets
- Nvidia RTX™ 4070
- Intel® Core™ i5-12400 CPU
- PyTorch

Default Transfer Learning Results

Table 1: Prediction with default COCO objects

Model	**mAP_{50}**	**mAP_{50-95}**	**F1**
YOLO8x	87.6	75.0	85.1
YOLO9e	85.7	75.7	84.3
YOLO11x	**88.1**	**77.9**	**85.9**

Transfer Learning Copy-Paste Results

Table 2: Prediction using transferred augmenting environments

Model	**mAP_{50}**	**mAP_{50-95}**	**F1**
YOLO8x	93.0	77.0	93.0
YOLO9e	**94.6**	79.3	93.9
YOLO11x	94.3	**80.1**	**94.1**

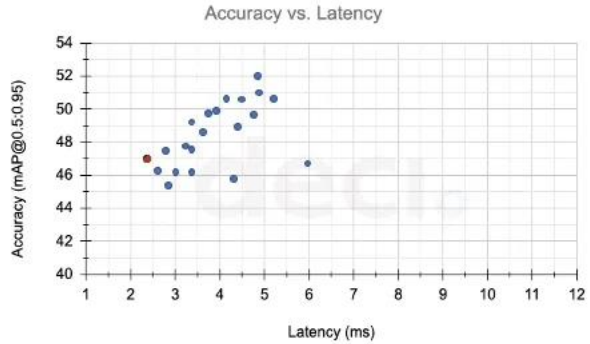
Unseen Predictions

Questions/Suggestions… Answers/Money?

Questions/Suggestions… Answers/Money?

YOLO-NAS

Neural Architecture Search



deci