DePondFi '24
Detection of Pond Fish Challenge Report

TEAM NAME: UNITECH

MEMBERS:
- Prenitha R - 21BCE1430
- Pardheev Krishna Tammineni - 21BCE1406

GITHUB LINK: click here
GOOGLE DRIVE LINK : click here

About

This report documents our approach and results for the DePondFi '24 challenge, which aims to detect fish key points from pond images.Due to the challenging conditions of underwater environments, the images are often obscured by haze, which makes direct analysis difficult. To address this, we employ a two-step process:

a. Image dehazing
b. Keypoint detection using PyTorch

Image Dehazing

We utilise two primary techniques: RGB stretching and LAB stretching.This is a fast and effective way compared to deep learning models.
RGB stretching adjusts the contrast and brightness of the image in the RGB colour space, enhancing the visibility of features while LAB stretching improves the colour and contrast of the image.
These techniques are chosen for their speed and efficiency compared to deep learning models for dehazing, which can be computationally intensive and time-consuming.

Speed and Efficiency: Traditional deep learning models for dehazing require significant computational resources and time for both training and inference. In contrast, RGB and LAB stretching are much faster as they involve simple mathematical operations rather than complex neural network computations.
Resource Requirements: DL models typically require powerful GPUs and substantial memory to operate efficiently. Our dehazing techniques can run on standard CPUs, making them more accessible and cost-effective.
Scalability: Given their low computational overhead, RGB and LAB stretching can easily handle large volumes of images, making them highly scalable for extensive datasets.

Ease of Implementation: These techniques are straightforward to implement and integrate into existing workflows, without the need for specialized deep learning frameworks or extensive tuning.

Key point Detection

With the dehazed images prepared, the next step is to detect keypoints on the fish. We train a YOLOv8n model specifically for this purpose. YOLO is a state-of-the-art object detection model, and the 'n' variant is optimised for efficiency and speed, making it suitable for our application.

The YOLOv8 Pose model operates with a top-down approach, where the detection and pose estimation tasks are learned simultaneously during training. However, the relationship between detection boxes and pose estimation is more nuanced than a direct sequential process.

The architecture is designed to leverage shared features for both detection and pose estimation, which is why the same input is used for both the detection and pose heads. During inference, the pose estimation is conditioned on the presence of detection boxes, but this relationship is implicit rather than explicit within the code.

The pose estimation relies on the detection results, but the connection between them is not as direct as taking detection boxes and then estimating poses within those boxes. Instead, the model learns to estimate poses in a way that is implicitly guided by the detection task during training. This is why poses are not estimated without corresponding detection boxes.

1. **Model Training**: The dehazed images are annotated with keypoints indicating important features on the fish, such as fins, eyes, and tail. These annotated images are used to train the YOLOv8n model, enabling it to learn the characteristics and positions of the keypoints.
2. **Inference**: Once trained, the model can be used to predict keypoints on new images, facilitating various downstream tasks such as fish behavior analysis, species identification, and health monitoring.

Key Points Labels:

0- Mouth
1- Eye
2- Top fin
3- Fish centre
4- Bottom centre
5- start point tail
6- top outline tail
7- mid outline tail
8- bottom outline tail

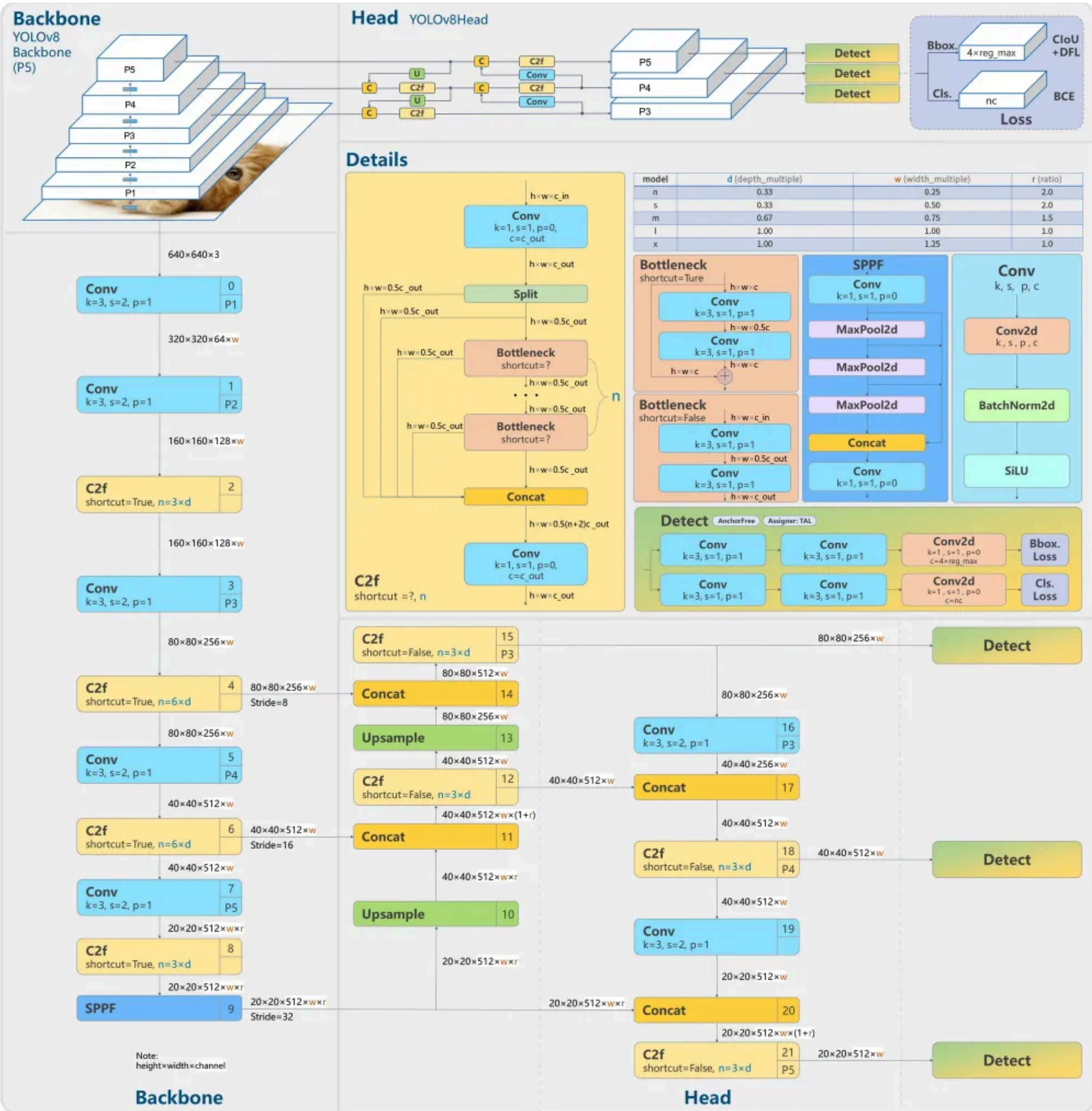
There are 3 essential blocks in the algorithm and everything will occur in these blocks, which are: Backbone, Neck and Head. The function of each block is described below.

- **Backbone**: The backbone, also known as the feature extractor, is responsible for extracting meaningful features from the input.
  - Captures simple patterns in the initial layers, such as edges and textures.
  - Can have multiple scales of representation as you go, capturing features from different levels of abstraction.
  - Will provide a rich, hierarchical representation of the input.
- **Neck**: The neck acts as a bridge between the backbone and the head, performing feature fusion operations and integrating contextual information. Basically the Neck assembles feature pyramids by aggregating feature maps obtained by the Backbone, in other words, the neck collects feature maps from different stages of the backbone.
  - Perform concatenation or fusion of features of different scales to ensure that the network can detect objects of different sizes.
  - Integrates contextual information to improve detection accuracy by considering the broader context of the scene.
  - Reduces the spatial resolution and dimensionality of resources to facilitate computation, a fact that increases speed but can also reduce the quality of the model.
- **Head**: The head is the final part of the network and is responsible for generating the network's outputs, such as bounding boxes and confidence scores for object detection.
  - Generates bounding boxes associated with possible objects in the image.
  - Assigns confidence scores to each bounding box to indicate how likely an object is present.
  - Sorts the objects in the bounding boxes according to their categories.

Convolution equation:

$$(X * Y)(z) = \int_{-\infty}^{\infty} p_X(x) \cdot p_Y(z - x)\, dx \tag{1}$$

Architecture diagram:



Conclusion

Our approach successfully detected fish key points in hazy pond images, contributing to the DePondFi '24 challenge. The results support intelligent aquaculture systems in fish identification and biomass estimation, aligning with the challenge's objectives. This report details the complete process from data preparation to visualisation, showcasing our methodology and results.