Advancements in Fish Segmentation: A Mask R-CNN Approach

Utilizing Point Rend Mask Head for Enhanced Precision

Traditional Approaches

1. **Manual Annotation** - Traditional methods often relied on manual annotation, where experts manually delineated fish boundaries in images.
    a. Pros: Provides high-quality annotations tailored to specific research needs.
    b. Cons: Time-consuming, labor-intensive, and not scalable to large datasets. Subject to human error and inter-annotator variability.

2. **Basic Computer Vision Techniques** - Basic computer vision techniques such as thresholding, edge detection, and morphological operations were commonly used for fish segmentation.
    a. Pros: Simplicity and ease of implementation.
    b. Cons: Time-consuming, labor-intensive, and not scalable to large datasets. Subject to human error and inter-annotator variability.

3. **Semi-Automated Algorithms GrabCut** - GrabCut's main goal is to separate foreground objects from the background of an image, but it also requires human interaction to refine the object and background regions.

Fishial Segmentation Dataset as of 10/2/2023

The latest version V5 of the segmentation training dataset includes:

1. 59789 Total train images
2. 6846 Total validation images

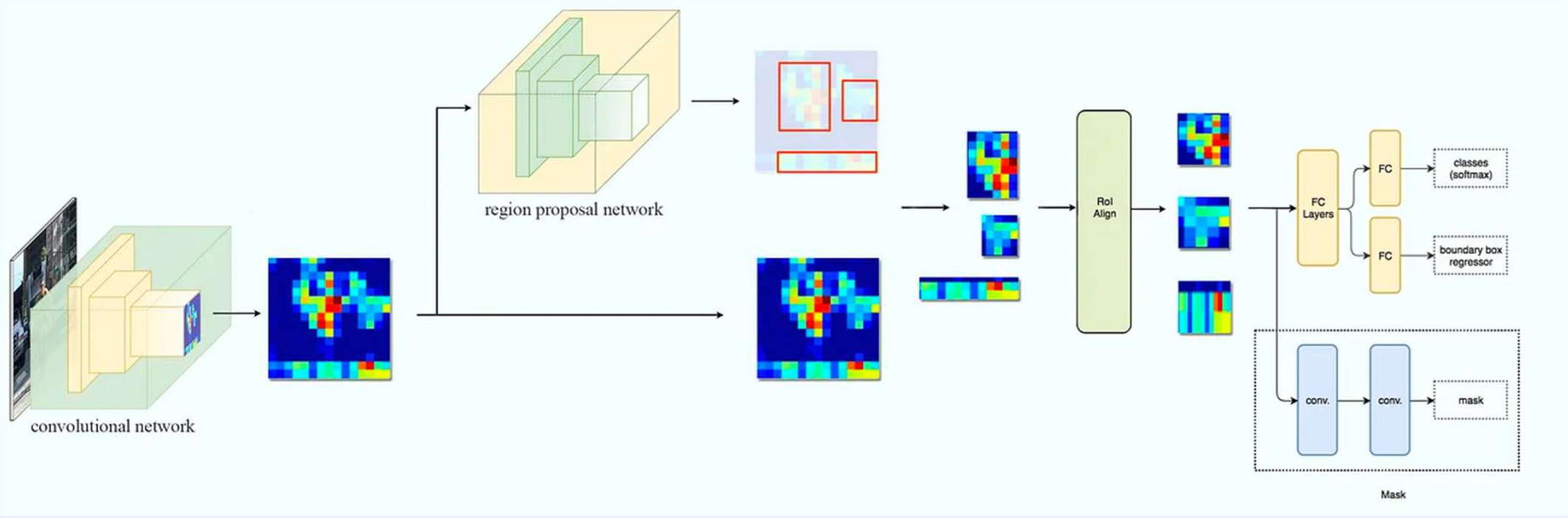
The latest version of the classification and training dataset includes:

1. 49426 Total train images.
2. 14156 Total validation images.
3. 289 Number of different classes.
4. 535 Maximum number of images for one class (Sphyraena barracuda).
5. 44 Minimum number of images for one class (Opisthonema oglinum)

Dataset

Example images in the classification and segmentation dataset

Introduction to Mask R-CNN

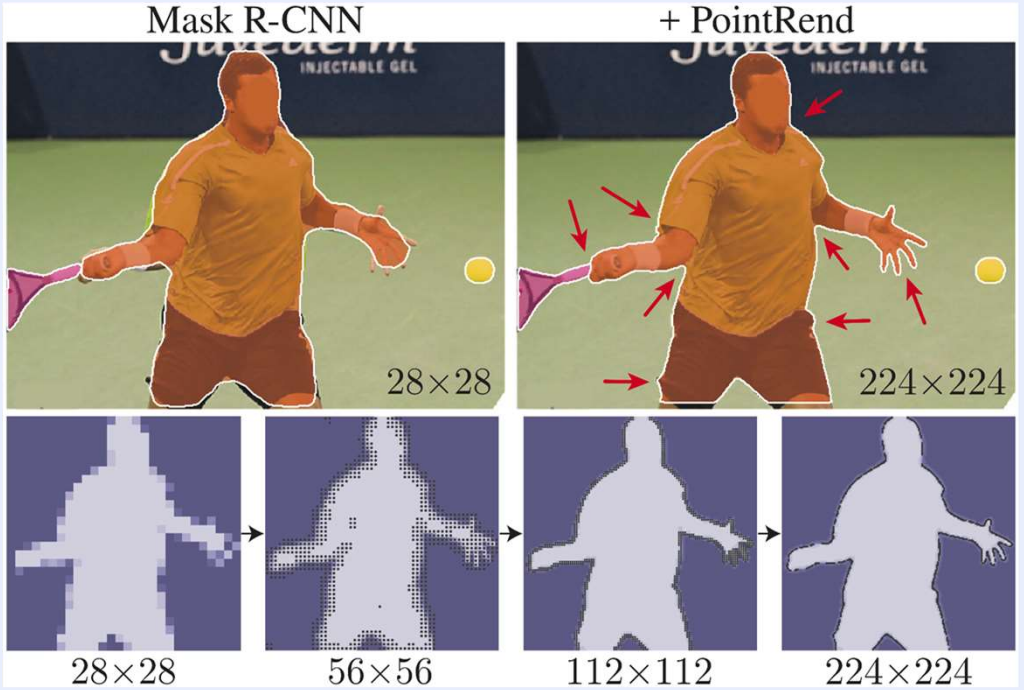


[src] https://jonathan-hui.medium.com/image-segmentation-with-mask-r-cnn-ebe6d793272

Architecture	Mask R-CNN
Backbone	Feature Pyramid Network Res Net 50
Mask Head	*Point Rend Mask Head*

Point Rend Mask Head

The standard implementation of **MaskRCNN** includes the **Mask Head** module, at the output of which we obtain a logit mask of **28x28** values, thus for large-sized images important information will be lost and the polygons will not be accurate, and increasing the logit mask entails a significant increase in the model's memory occupied and processing time. To do this, it was replaced with the **"Point Rend Mask Head"** module, which allows you to calculate an increased logit mask without resource-intensive operations.



Parameters of Model and Training Results

1. Input image size: 1024x1024 pixels
2. Training Configuration:
    a. AMP: True
    b. Optimizer: Adam
    c. Learning rate: 0.0028
    d. Batch size: 32
    e. Number of epochs: 100
3. Data Augmentation:
    a. Copy-Paste
    b. Horizontal flip
    c. Vertical flip
    d. Random rotation
    e. Random scale jitter

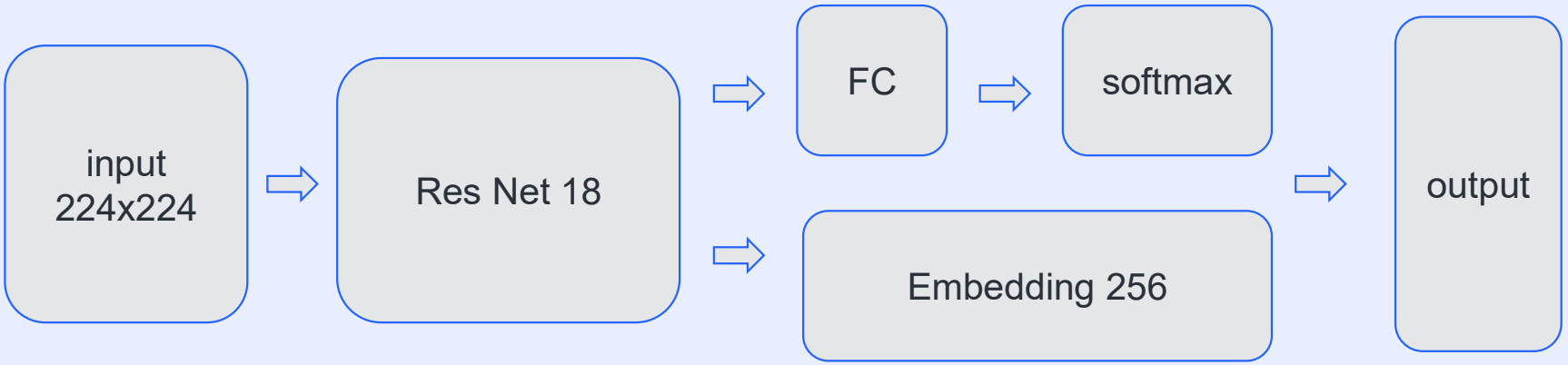
1. Loss Function:
    a. Combined loss: Sum of classification, bounding box regression, and mask segmentation losses
2. Point Rend Mask Head:
    a. Iterations: 4
    b. Number of points: 14x14 grid
3. Evaluation Metrics:
    - Mean Average Precision (mAP)

Mean Average Precision (mAP)

AP	AP50	AP75	APs	APm	APl
82.504	96.742	94.727	13.283	58.029	84.540

Fish Classification

Architecture:



The architecture of the current classification model version distinguishes itself from standard classification models by incorporating two parallel layers: **Fully Connected** and **Embedding Layer.** The training process unfolds in two stages: initially, the **Embedding network** and the **'Feature extractor' (ResNet-18)** are trained, followed by the freezing of the network and the exclusive training of the Fully Connected layer. This approach yields enhanced classification accuracy through improved ranking strategies, particularly beneficial in noisy data scenarios, ultimately contribute positively to the final outcome.

Training Classification Model

1. **Loss Function** (Embedding Network): Quadruplet Loss
2. **Loss Function** (Fully Connected Network): Cross Entropy Loss
3. **Batch Sampler**:
    a. classes_per_batch: 18
    b. samples_per_class: 8
4. **Learning rate:** 0.015
5. **Momentum**: 0.9
6. **Epoch:** 4000
7. **Warmup Steps:** 500
8. **Opt Level:** 'O2'
9. **Data Augumentation:**
    a. RandomAutocontrast
    b. RandomHorizontalFlip
    c. RandomVerticalFlip
    d. RandomErasing

GIF demonstrating the distribution of fish embeddings on a 3D surface

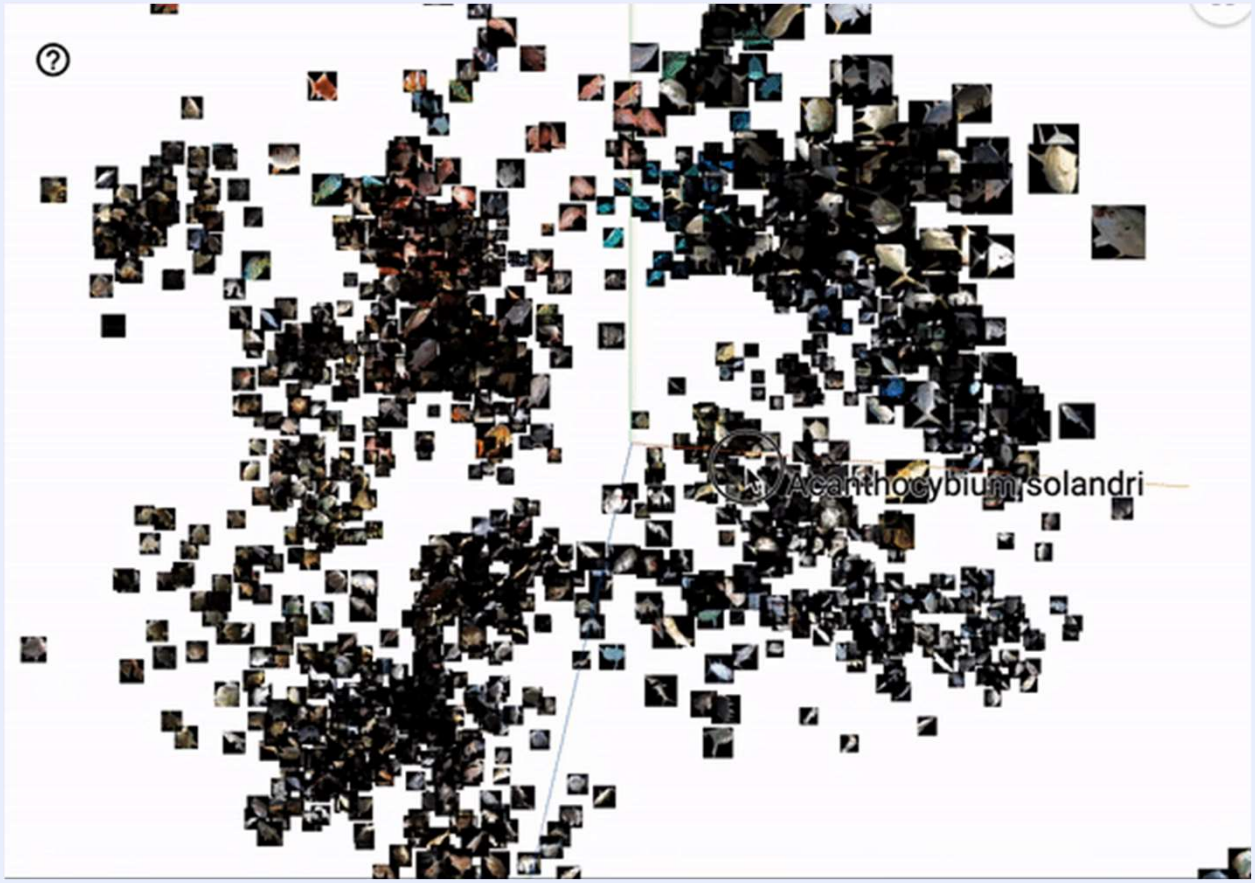
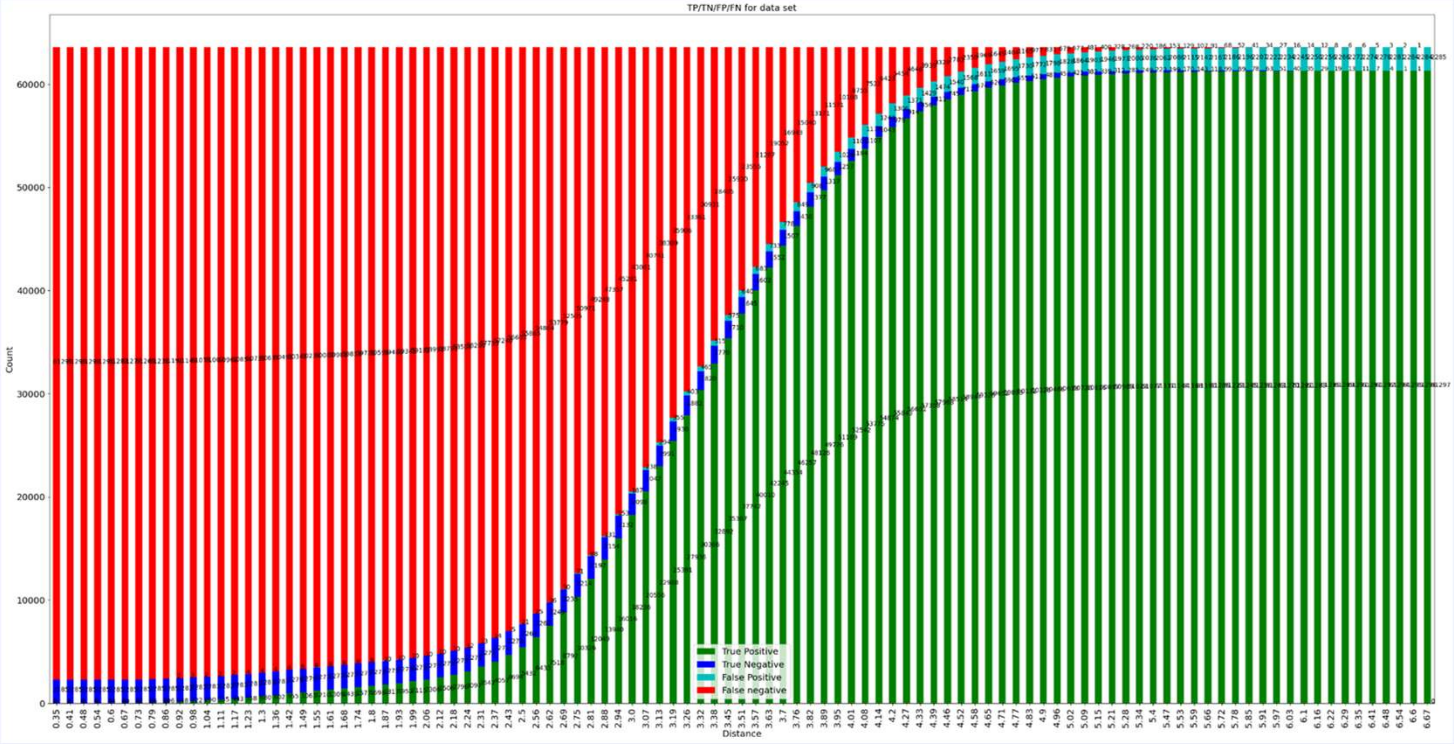


Figure showing the TP/FP/TN/TP for different threshold values



Training Results

Overall classification accuracy:

1. Average **Precision**: 0.9643
2. Average **Recall**: 0.9641
3. Average **F1-score**: 0.964

Table of metrics for each class

	precision	recall	f1-score	support
Opisthonema oglinum	1.0000	0.9773	0.9885	44.0000
Sphyrna mokarran	0.9375	0.9375	0.9375	48.0000
Hiodon alosoides	1.0000	0.9388	0.9684	49.0000
Neogobius melanostomus	0.9800	0.9608	0.9703	51.0000
Squalus acanthias	0.9574	0.8824	0.9184	51.0000
Leuciscus aspius	0.9623	0.9623	0.9623	53.0000
Xiphias gladius	0.9630	0.9811	0.9720	53.0000
Lutjanus cyanopterus	0.9464	0.9636	0.9550	55.0000
Acroteriobatus annulatus	1.0000	0.9828	0.9913	58.0000
Clupea harengus	0.9123	0.8966	0.9043	58.0000
Scaphirhynchus platorynchus	0.9825	0.9492	0.9655	59.0000
Arripis truttacea	0.9180	0.9333	0.9256	60.0000
Micropterus notius	0.9259	0.8333	0.8772	60.0000
Scomberomorus commerson	1.0000	0.9677	0.9836	62.0000
Carcharhinus obscurus	0.8030	0.8281	0.8154	64.0000
Pterois volitans	1.0000	1.0000	1.0000	68.0000

[Complete list]

Thank you!

Support@fishial.ai