Evaluation

Perez Jenny

Department of Computer Science
Earlham College

May 9, 2025

Overview

Evaluation

Process of measuring how well a model performs on a given, task:

- Object detection
- Image classification
- segmentation

Involves using specific metric and techniques to asses accuracy, robustness and generalization.

Metrics for Classification

Classifiction models assign labels to images.Evaluating these models requires specific classification metrics to measure performance.

- Accuracy = Correct Preditions / Total Predictions – We used it for balanced datasets
- F1-Score = 2 x (Presicion x Recall) / (Precision + Recall) – Imbalanced dataset

Metrics for Object Detection

In computer vision, evaluating object detection models requires specific metrics to measure how well the model identifies and localizes objects. The most common evaluation metrics for object detection are:

- Intersection over Union (IoU): Measures the overlap between the predicted bounding box and the ground-truth bounding box. IoU = Area of Overlap/Area of Union
- Mean Average Precision (mAP): is computed for each class and then averaged across all classes.

Metrics for segmentation

You are trying to classify every pixel of an image into a category (like "cat", "dog", "background", etc.). Once your model makes predictions, you must evaluate how well it did. There are formulas or measurements to check this:

- Pixel Accuracy = Number of correct pixels / Total number of pixels (Measures the percentage of correctly classified pixels)
- Intersection over Union (IoU) = Area of Overlap / Area of Union (Measures the overlap between the predicted segmentation and the ground truth)

Confusion matrix

Is a table that helps you visualize how well your model is doing by comparing its predictions against the ground truth (the real labels). It is very common in classification, but also plays a role in segmentation, where you're classifying every pixel instead of whole images.

Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

cross-validation

is a technique used to evaluate how well your model will generalize to new, unseen data. In simple words: Instead of training your model only once on a training set and testing on a separate test set, you split your data into multiple parts (called folds) and train/test your model multiple times on different splits.

- To reduce the risk of overfitting.
- To get a more reliable and stable estimation of your model's performance.
- To make sure that your model performs well not just on a lucky split of the data.

Visualization

Visualization means generating graphical outputs that help you:

- Understand your data.
- Inspect the predictions your model is making
- Analyze your model's errors.
- Communicate results to others.

In segmentation is bout comparing the input image, The ground truth mask (the correct segmentation) and The predicted mask (what your model generated).

Error analysis

Error analysis is the process of studying and understanding the mistakes your model is making. It's not just about knowing how much error you have (like "the IoU is 0.75"), but also where, why, and how the errors happen.

- False Positive (FP): Model predicts a pixel as belonging to a class when it actually doesn't.
- False Negative (FN): Model misses pixels that actually belong to the class.
- Misclassification: Model predicts the wrong class (in multi-class segmentation).