Case Study #1 — Instant Segmentation in Medical Imaging

Rhichard Koh (100842848)

Faculty of Science, Engineering & Information Technology, Durham College

Advanced Neural Networks: Deep Learning for Spatial Analysis (COSC 41001)

Dr. Noopa Jagadeesh

November 6, 2024

Fundamental Principles and Challenges of Medical Image Diagnosis

Medical image diagnosis involves analyzing complex imaging data, such as X-rays, MRIs, and CT scans, to detect abnormalities and assist in medical decision-making. This field presents unique challenges compared to other domains due to the nature of medical images, which require high precision and expertise for accurate interpretation. One of the key challenges is the variability in image quality and the subtle differences between normal and abnormal tissues, which demand highly sophisticated models.

Another significant challenge is class imbalance, where certain medical conditions are much less common than others, leading to a scarcity of positive examples. This imbalance can result in biased models that perform well on common cases but fail to identify rare conditions. Additionally, the small size of available training datasets is a common issue, as collecting and annotating medical images is resource-intensive and requires expert knowledge. Furthermore, medical diagnosis often involves multitask learning, such as detecting multiple types of anomalies or segmenting different anatomical regions, adding another layer of complexity.

To address these challenges, several strategies can be used. For class imbalance, techniques such as data augmentation, synthetic data generation, and cost-sensitive learning can help improve model performance on minority classes. To tackle the issue of small training sets, transfer learning is commonly used, where models pre-trained on large datasets are fine-tuned for specific medical tasks. Multitask learning can be approached by designing architectures capable of sharing features across tasks, thus improving generalization and efficiency.

Ethical Considerations and Implications of Using AI in Medical Image Diagnosis

The use of AI in medical image diagnosis raises significant ethical considerations. Privacy concerns are paramount, as medical images contain sensitive patient information. Ensuring data security and adhering to regulations like HIPAA are essential to protect patient privacy. Additionally, biases in training data can lead to unequal outcomes, where certain demographics may receive less accurate diagnoses. Addressing these biases requires careful data curation and fairness checks throughout the model development process.

Another ethical concern is the potential consequences of incorrect diagnoses. Misdiagnoses can have severe implications for patient health, leading to incorrect treatments or delayed care. Therefore, human oversight is crucial in AI-assisted medical diagnosis to ensure that AI predictions are reviewed by qualified medical professionals. The role of AI should be to assist, not replace, healthcare providers, enhancing their ability to make informed decisions rather than fully automating the diagnostic process.

Building the Model

For the segmentation model used in this project, a deep learning architecture such as YOLOv11 was selected. YOLOv11 is particularly well-suited for medical image segmentation due to its integrated object detection and segmentation capabilities. Its architecture allows for precise localization of features, which is crucial for accurately segmenting small or complex structures within medical images with a fast speed as well.

To address the challenge of class imbalance, data augmentation techniques were employed, including rotations, flips, and intensity variations, to artificially increase the number of samples from underrepresented classes. Transfer learning was also used by initializing the model with weights from a network pre-trained on a large medical imaging dataset (yolo11n-seg.pt), which helped improve performance given the limited size of the training set. The data augmentation technique that was used after hyperparameter tuning was only changing the scale of the images. I believe the reason for this is because when the image is rotated or sheared the corresponding label for that image will no longer match with the augmented image. The final option is to just manually get more data for the imbalanced classes to prevent the model from overfitting.

The training process involved optimizing the model using the box_loss, seg_loss, cls_loss, and dfl_loss loss function and an adam optimizer, which is effective for instance segmentation tasks with imbalanced data. Hyperparameter tuning was conducted to find the optimal learning rate and batch size, and early stopping was used to prevent overfitting. Fine-tuning the model on a validation set helped further improve accuracy and generalization.

The model was implemented using the Ultralytics YOLO framework, with segmentation capabilities. The training and validation were carried out using an annotated dataset of medical images, and the model was trained to segment regions of interest with high precision. The use of pre-trained weights, combined with augmentation and careful fine-tuning, allowed for effective training despite the challenges posed by a relatively small dataset.

Evaluation Metrics and Performance

To evaluate the model's performance, several metrics were used, including Dice coefficient, Intersection over Union (IoU), precision, recall, and Hausdorff distance. The Dice coefficient and IoU were chosen to measure the overlap between predicted and ground truth masks, which is critical for segmentation tasks. Precision and recall were used to assess the model's ability to correctly identify relevant regions, while the Hausdorff distance provided insight into the spatial accuracy of the segmentation boundaries.

The evaluation was performed using the test set, where metrics were calculated for each image and then averaged to determine the model's overall performance. Based on these metrics, the model demonstrated strong performance in accurately segmenting the target regions, with high Dice and IoU scores indicating good overlap with the ground truth. Specifically, the model achieved a precision of 0.9008, recall of 0.9365, and an accuracy of 0.9889. The Dice coefficient and IoU scores were 0.9092 and 0.8455, respectively, indicating effective segmentation.

However, the Hausdorff distance of 5.1083 revealed some limitations in boundary precision, particularly in cases with complex or irregular shapes. The Mean Absolute Error (MAE) of 1.1293 and Pixel Accuracy of 0.9889 further provided insights into the model's strengths and areas for improvement. The model's strengths lie in its ability to generalize well to new images, thanks to the use of data augmentation and transfer learning. Its limitations include occasional misclassification of ambiguous regions, which underscores the importance of human oversight in clinical settings.

Overall, the evaluation metrics indicate that the model performs well in terms of segmentation quality, but further refinement is needed to handle complex cases more accurately. The continuous monitoring and fine-tuning of the model are essential to ensure its reliability and effectiveness in practical applications.

Best Model Parameters

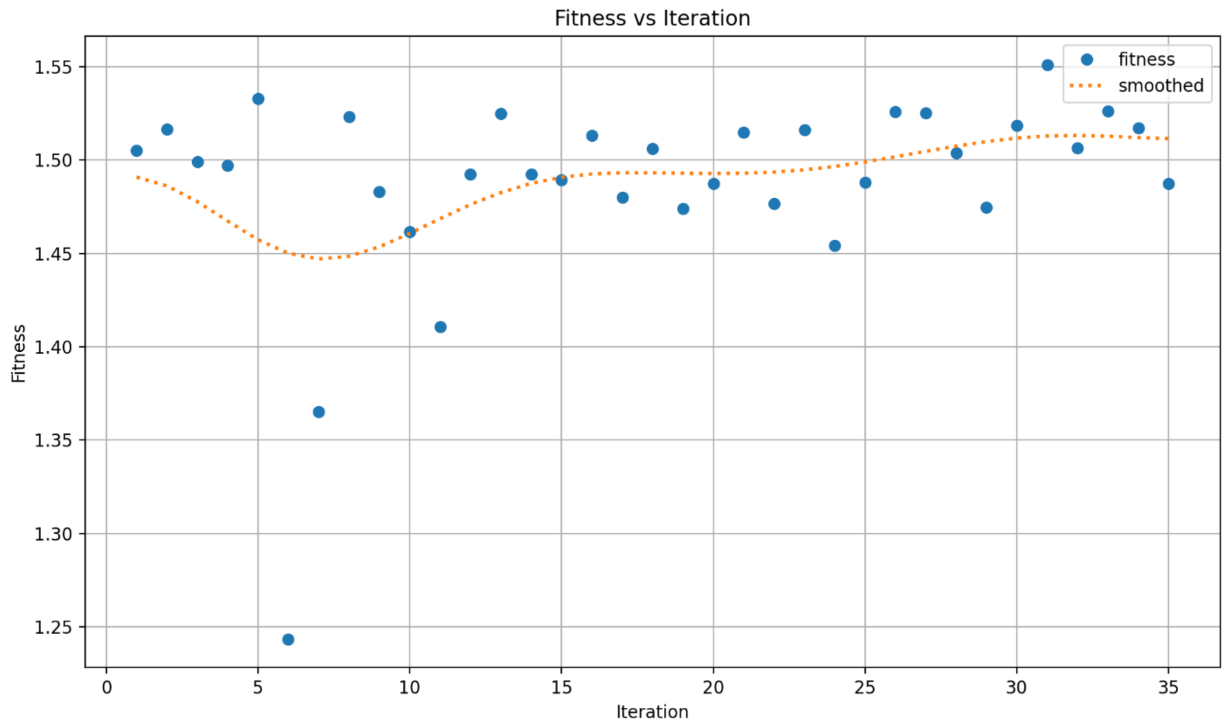
```
# 35/300 iterations complete ✅ (5321.74s)
# Results saved to runs\segment\tune
# Best fitness=1.55114 observed at iteration 31
# Best fitness metrics are {'metrics/precision(B)': 0.96714, 'metrics/recall(B)':
0.96825, 'metrics/mAP50(B)': 0.98874, 'metrics/mAP50-95(B)': 0.76163,
'metrics/precision(M)': 0.96714, 'metrics/recall(M)': 0.96825,
'metrics/mAP50(M)': 0.99027, 'metrics/mAP50-95(M)': 0.74197, 'val/box_loss':
0.58534, 'val/seg_loss': 1.06799, 'val/cls_loss': 0.76963, 'val/dfl_loss':
1.11082, 'fitness': 1.55114}
# Best fitness model is runs\segment\train32
# Best fitness hyperparameters are printed below.

lr0: 0.00892
lrf: 0.00951
momentum: 0.94102
weight_decay: 0.0005
warmup_epochs: 3.64238
```

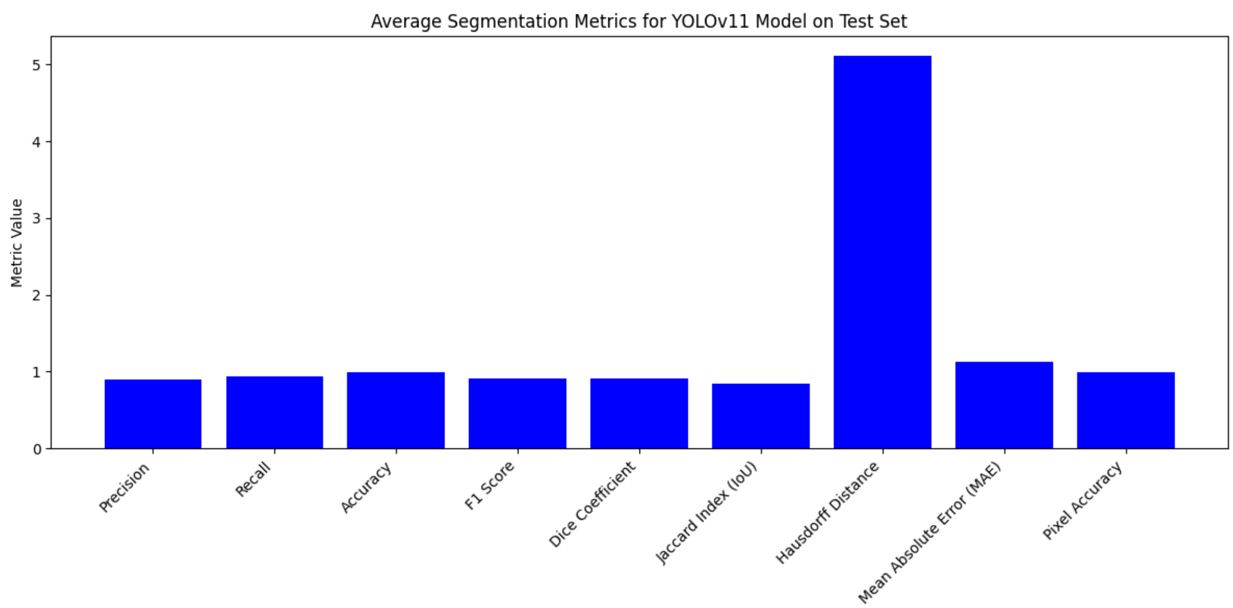
```
warmup_momentum: 0.76332
box: 5.12088
cls: 0.70699
dfl: 1.57928
hsv_h: 0.01536
hsv_s: 0.83299
hsv_v: 0.35813
degrees: 0.0
translate: 0.12168
scale: 0.53601
shear: 0.0
perspective: 0.0
flipud: 0.0
fliplr: 0.64594
bgr: 0.0
mosaic: 0.69792
mixup: 0.0
copy_paste: 0.0
================================================================================
task: segment
mode: train
model: yolo11n-seg.pt
data: BRAIN-TUMOR.v1i.yolov11/data.yaml
epochs: 10
time: null
patience: 100
batch: 16
imgsz: 640
save: true
save_period: -1
cache: true
device: 0
workers: 8
project: null
name: train32
exist_ok: false
pretrained: true
optimizer: AdamW
verbose: true
seed: 0
deterministic: true
single_cls: false
rect: false
cos_lr: false
close_mosaic: 10
```

```yaml
resume: false
amp: true
fraction: 1.0
profile: false
freeze: null
multi_scale: false
overlap_mask: true
mask_ratio: 4
dropout: 0.0
val: true
split: val
save_json: false
save_hybrid: false
conf: null
iou: 0.7
max_det: 300
half: false
dnn: false
plots: true
source: null
vid_stride: 1
stream_buffer: false
visualize: false
augment: true
agnostic_nms: false
classes: null
retina_masks: false
embed: null
show: false
save_frames: false
save_txt: false
save_conf: false
save_crop: false
show_labels: true
show_conf: true
show_boxes: true
line_width: null
format: torchscript
keras: false
optimize: false
int8: false
dynamic: false
simplify: true
opset: null
workspace: 4
```

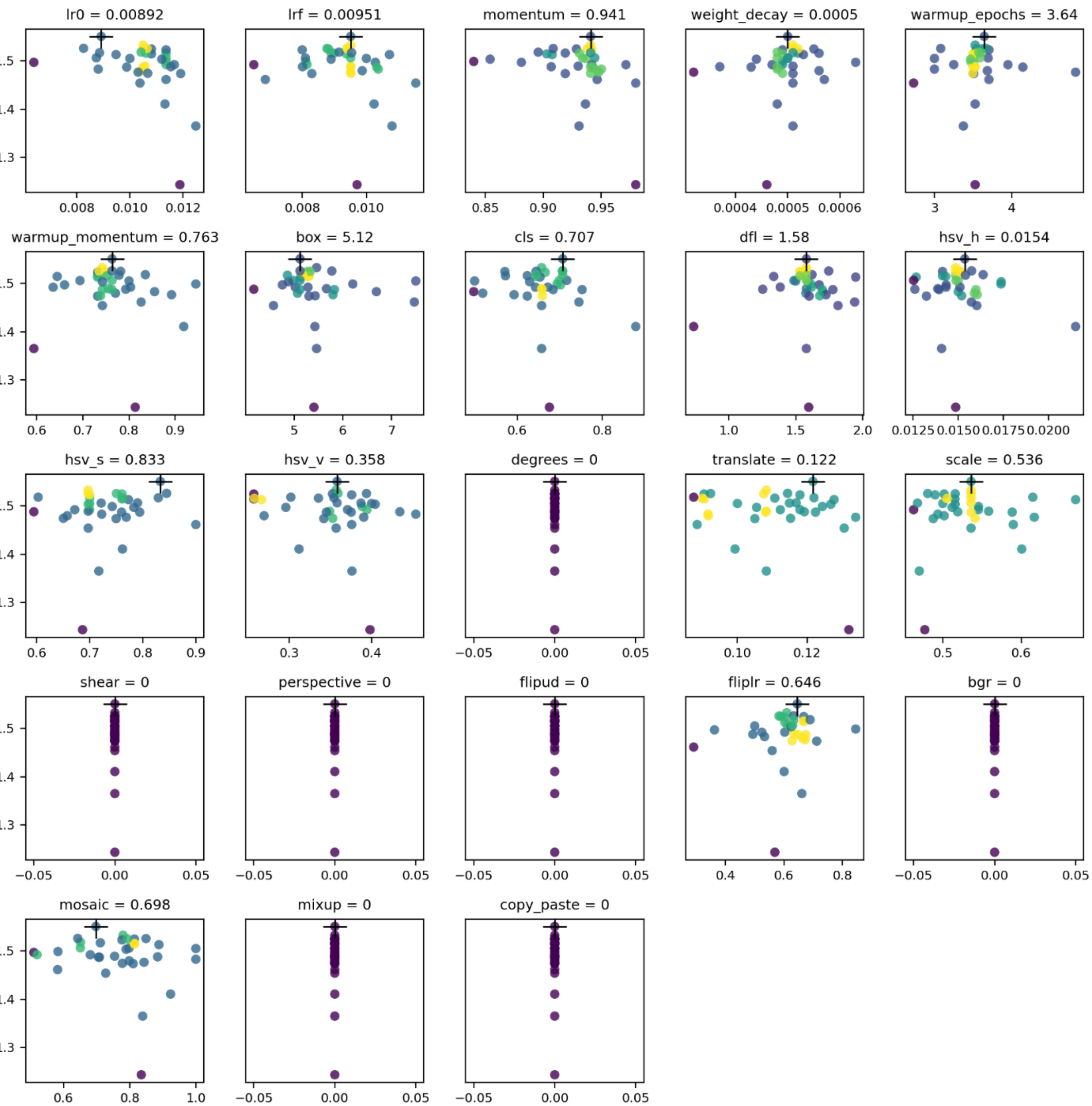
```yaml
nms: false
lr0: 0.00892
lrf: 0.00951
momentum: 0.94102
weight_decay: 0.0005
warmup_epochs: 3.64238
warmup_momentum: 0.76332
warmup_bias_lr: 0.1
box: 5.12088
cls: 0.70699
dfl: 1.57928
pose: 12.0
kobj: 1.0
label_smoothing: 0.0
nbs: 64
hsv_h: 0.01536
hsv_s: 0.83299
hsv_v: 0.35813
degrees: 0.0
translate: 0.12168
scale: 0.53601
shear: 0.0
perspective: 0.0
flipud: 0.0
fliplr: 0.64594
bgr: 0.0
mosaic: 0.69792
mixup: 0.0
copy_paste: 0.0
copy_paste_mode: flip
auto_augment: randaugment
erasing: 0.4
crop_fraction: 1.0
cfg: null
tracker: botsort.yaml
save_dir: runs\segment\train32
```

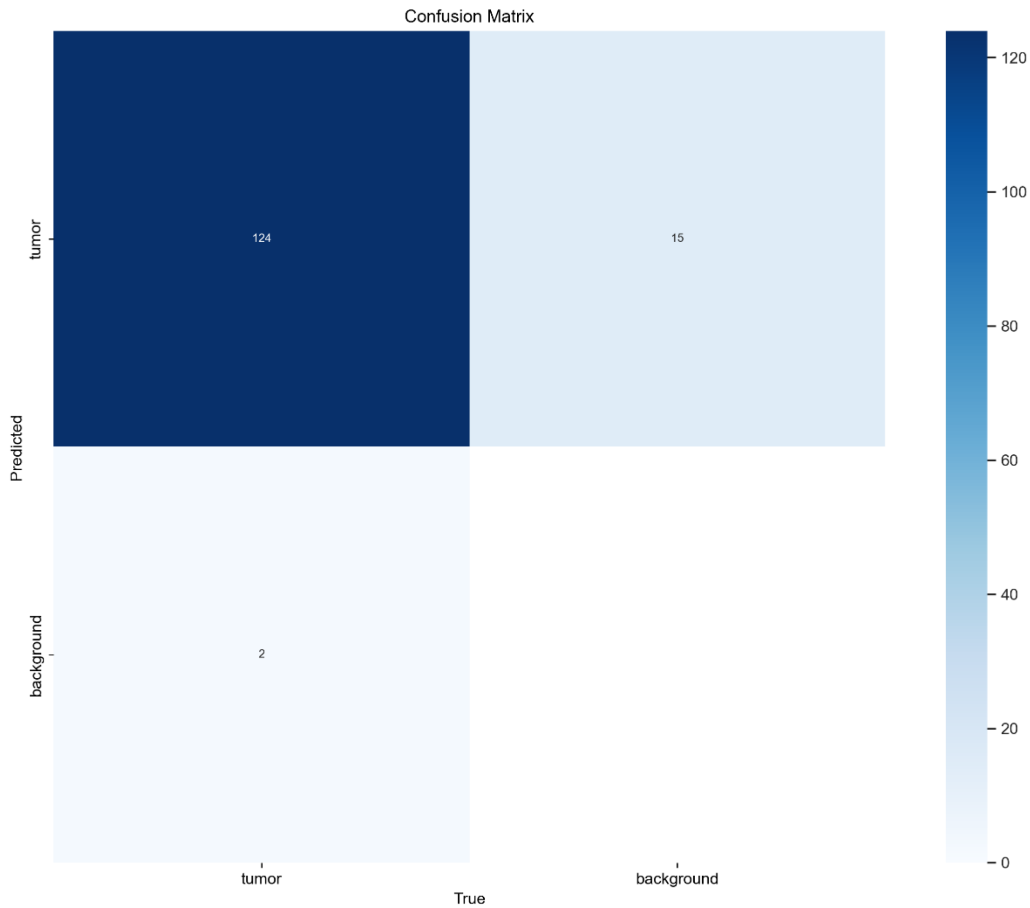


Fitness vs Iteration
fitness
smoothed
Fitness
Iteration
1.55
1.50
1.45
1.40
1.35
1.30
1.25
0
5
10
15
20
25
30
35



Average Segmentation Metrics for YOLOv11 Model on Test Set
Metric Value
0
1
2
3
4
5
Precision
Recall
Accuracy
F1 Score
Dice Coefficient
Jaccard Index (IoU)
Hausdorff Distance
Mean Absolute Error (MAE)
Pixel Accuracy


lr0 = 0.00892
lrf = 0.00951
momentum = 0.941
weight_decay = 0.0005
warmup_epochs = 3.64
warmup_momentum = 0.763
box = 5.12
cls = 0.707
dfl = 1.58
hsv_h = 0.0154
hsv_s = 0.833
hsv_v = 0.358
degrees = 0
translate = 0.122
scale = 0.536
shear = 0
perspective = 0
flipud = 0
fliplr = 0.646
bgr = 0
mosaic = 0.698
mixup = 0
copy_paste = 0



Confusion Matrix
124
15
2
Predicted
tumor
background
True
120
100
80
60
40
20
0