|  |  |  |
| --- | --- | --- |
| **پرسش ۱** | **نام و نام خانوادگی** | سیدرضا مسلمی |
| **شماره دانشجویی** | 810103326 |
| **پرسش ۲** | **نام و نام خانوادگی** | بی‌بی رقیه |
| **شماره دانشجویی** | 810102053 |
|  | **مهلت ارسال پاسخ** | **۱۴۰۱.۰۷.۰۱** |

|  |  |  |
| --- | --- | --- |
|  | **به نام خدا**  **دانشگاه تهران**  **دانشکده‌ مهندسی برق و کامپیوتر** |  |
| **درس شبکه‌های عصبی و یادگیری عمیق**  **تمرین اول** | | |

**فهرست**

[**قوانین** 1](#_Toc184422629)

[**پرسش 1**. **سگمنتیشن تومور مغزی از روی تصاویر MRI** 1](#_Toc184422630)

[۱-۱. توصیف مدل ارائه شده 1](#_Toc184422631)

[۱-2 و ۱-3. آماده‌سازی و تقویت مجموعه داده 3](#_Toc184422632)

[۱-4. بهینه‌ساز، معیارها و تابع هزینه 5](#_Toc184422633)

[۱-5. پیاده‌سازی مدل 7](#_Toc184422634)

[۱-6. آموزش مدل 9](#_Toc184422635)

[۱-7. ارزیابی مدل 12](#_Toc184422636)

[**پرسش ۲** **- تشخیص تابلو های راهنمایی و رانندگی** 17](#_Toc184422637)

[2-1. **آماده سازی مجموعه داده** 17](#_Toc184422638)

[۲-۲. **تنظیم دقیق و ارزیابی مدل تشخیص شی دو مرحله ای** 19](#_Toc184422639)

[۲**-۳** . **تنظیم دقیق و ارزیابی مدل تشخیص شی تک مرحله ای** 21](#_Toc184422640)

[۲**-۴** . **ارزیابی نتایج و مقایسه مدل ها** 29](#_Toc184422641)

**شکل‌ها**

Figure 1: Main Data Along with its Augmented images 4

Figure 2: Focal Loss 6

Figure 3: UNet-VGG16 8

Figure 4: Train and Validation set Loss Over Epochs 9

Figure 5: : Dice Score Over Epochs 10

Figure 6: IoU Over Epochs 10

Figure 7: Accuracy Over Epochs 11

Figure 8: Segmentation Permormance 14

Figure 9: Sample Annotated Images 18

Figure 10: Frequencies and Distributions of dataset 19

Figure 11:: Metrics for FRCNN-ResNet50 21

Figure 12: Metrics for SSD-VGG16 21

Figure 13: AP vs IoU Per Class 29

Figure 14: mAP for Different Models and Image Sizes 29

**جدول‌ها**

[1:Overall Comparison between FRCNN and SSD300 38](#_Toc115453324)

# **قوانین**

قبل از پاسخ دادن به پرسش‌ها،‌ موارد زیر را با دقت مطالعه نمایید:

* از پاسخ‌های خود یک گزارش در قالبی که در صفحه‌ی درس در سامانه‌ی Elearn با نام ***REPORTS\_TEMPLATE.docx*** قرار داده شده تهیه نمایید.
* پیشنهاد می‌شود تمرین‌ها را در قالب گروه‌های دو نفره انجام دهید. (بیش از دو نفر مجاز نیست و تحویل تک نفره نیز نمره‌ی اضافی ندارد) توجه نمایید الزامی در یکسان ماندن اعضای گروه تا انتهای ترم وجود ندارد. (یعنی، می‌توانید تمرین اول را با شخص A و تمرین‌ دوم را با شخص B و ... انجام دهید)
* **کیفیت گزارش شما در فرآيند تصحيح از اهميت ويژه­اي برخوردار است**؛ بنابراین، لطفا تمامی نکات و فرض­هایی را كه در پیاده­سازی­ها و محاسبات خود در نظر مي­گيريد در گزارش ذکر کنید.
* در گزارش خود مطابق با آنچه در قالب نمونه قرار داده شده، برای شکل‌ها زیرنویس و برای جدول‌ها بالانویس در نظر بگیرید.
* الزامی به ارائه توضیح جزئیات کد در گزارش نیست، اما باید نتایج بدست آمده از آن را گزارش و تحلیل کنید.
* **تحلیل نتایج الزامی می‌باشد، حتی اگر در صورت پرسش اشاره‌ای به آن نشده باشد.**
* **دستیاران آموزشی ملزم به اجرا کردن کدهای شما نیستند**؛ بنابراین، هرگونه نتیجه و یا تحلیلی که در صورت پرسش از شما خواسته شده را به طور واضح و کامل در گزارش بیاورید. در صورت عدم رعایت این مورد، بدیهی است که از نمره تمرین کسر می­شود.
* **کدها حتما باید در قالب نوت‌بوک با پسوند .ipynb تهیه شوند، در پایان کار، تمامی کد اجرا شود و خروجی هر سلول حتما در این فایل ارسالی شما ذخیره شده باشد.** بنابراین برای مثال اگر خروجی سلولی یک نمودار است که در گزارش آورده‌اید، این نمودار باید هم در گزارش هم در نوت‌بوک کد‌ها وجود داشته باشد.
* **در صورت مشاهده‌ی تقلب امتیاز تمامی افراد شرکت­کننده در آن، 100- لحاظ می­شود.**
* تنها زبان برنامه نویسی مجاز **Python** است.
* **استفاده از کدهای آماده برای تمرین­ها به­ هیچ ­وجه مجاز نیست. در صورتی که دو گروه از یک منبع مشترک استفاده کنند و کدهای مشابه تحویل دهند، تقلب محسوب می‌شود.**
* نحوه محاسبه­ تاخیر به این شکل است: پس از پایان رسیدن مهلت ارسال گزارش، حداکثر تا یک هفته امکان ارسال با تاخیر وجود دارد، پس از این یک هفته نمره آن تکلیف برای شما صفر خواهد شد.
  + سه روز اول: بدون جریمه
  + روز چهارم: ۵ درصد
  + روز پنجم: ۱۰ درصد
  + روز ششم: ۱۵ درصد
  + روز هفتم: ۲۰ درصد
* حداکثر نمره‌ای که برای هر سوال می‌توان اخد کرد ۱۰۰ بوده و اگر مجموع بارم یک **سوال** بیشتر از ۱۰۰ باشد، در صورت اخد نمره بیشتر از ۱۰۰، اعمال نخواهد شد.
  + برای مثال: اگر نمره اخذ شده از سوال ۱ برابر ۱۰۵ و نمره سوال ۲ برابر ۹۵ باشد، نمره نهایی تمرین ۹۷.۵ خواهد بود و نه ۱۰۰.
* لطفا گزارش، کدها و سایر ضمایم را به در یک پوشه با نام زیر قرار داده و آن را فشرده سازید، سپس در سامانه‌ی Elearn بارگذاری نمایید:

HW[Number] \_[Lastname]\_[StudentNumber]\_[Lastname]\_[StudentNumber].zip

(مثال: HW1\_Ahmadi\_810199101\_Bagheri\_810199102.zip)

* برای گروه‌های دو نفره، بارگذاری تمرین از جانب یکی از اعضا کافی است ولی پیشنهاد می‌شود هر دو نفر بارگذاری نمایند.

# **پرسش 1**. **سگمنتیشن تومور مغزی از روی تصاویر MRI**

۱-۱. توصیف مدل ارائه شده

**Introduction to the Model**

The paper talks about a new way to segment brain tumors from MRI images using a model that combines UNet and VGG16. UNet is great for tasks where we need to divide an image into specific parts, like finding tumors in MRI scans. VGG16, a powerful pre-trained model, helps in extracting important features from images. The authors use transfer learning, which means they take a model already trained on a large dataset (like ImageNet) and adapt it to work with MRI images.

**How the Model Works?**

1. **UNet Architecture:**

* UNet has two main parts: an encoder and a decoder.
* The encoder compresses the image into smaller and smaller parts to capture important features.
* The decoder then rebuilds the image, focusing on marking the tumor regions.
* Skip connections are added between the encoder and decoder, which help the model remember details about the tumor's location.

1. **VGG16 as the Encoder:**

* VGG16, a model pre-trained on millions of images, is used as the encoder in UNet.
* This helps the model quickly understand features in MRI images without starting from scratch.
* Using VGG16 improves the accuracy of detecting tumors because it’s already good at recognizing important patterns in images.

1. **Transfer Learning:**

* Instead of training the model from the beginning, the authors use weights (parameters) from VGG16, which were learned from another large dataset.
* This approach saves time, works well with smaller MRI datasets, and reduces the risk of overfitting.

**Why VGG16 and Transfer Learning?**

* Why VGG16?
  + VGG16 is simple and effective, and it’s widely used for image-related tasks.
  + It’s already trained to recognize general features in images, so it’s a good choice for transfer learning.
* Why Transfer Learning?
  + MRI datasets are small, so transfer learning allows the model to use knowledge from a larger dataset.
  + This makes training faster and improves performance.
  + It also helps the model generalize better to unseen MRI scans.

**Advantages of the Model**

1. Better Accuracy:

* The combination of UNet and VGG16 provides better results compared to training UNet alone.
* The model detects tumors more precisely, as shown by its high scores in tests like IoU (Intersection over Union).

1. Less Training Time:

* Because of transfer learning, the model needs less time to train.

1. Works Well on Small Datasets:

* The pre-trained VGG16 helps the model perform well even with fewer MRI images.

1. Robust and Reliable:

* The model works well on different MRI images, showing that it is flexible and reliable.

۱-2 و ۱-3. آماده‌سازی و تقویت مجموعه داده

**1. Rotation (10 Degrees)**

* What It Does: This augmentation rotates the image randomly by up to 10 degrees (clockwise or counterclockwise).
* Why It's Useful:
  + In real-world scenarios, objects in images, like road signs or tumors in medical images, might not always be perfectly aligned. Rotation helps the model handle these slight changes in orientation.
  + It teaches the model to recognize objects even if they are tilted slightly, making the model more robust.
* Effect on the Dataset:
  + The dataset becomes more varied because each rotated image looks slightly different from the original.

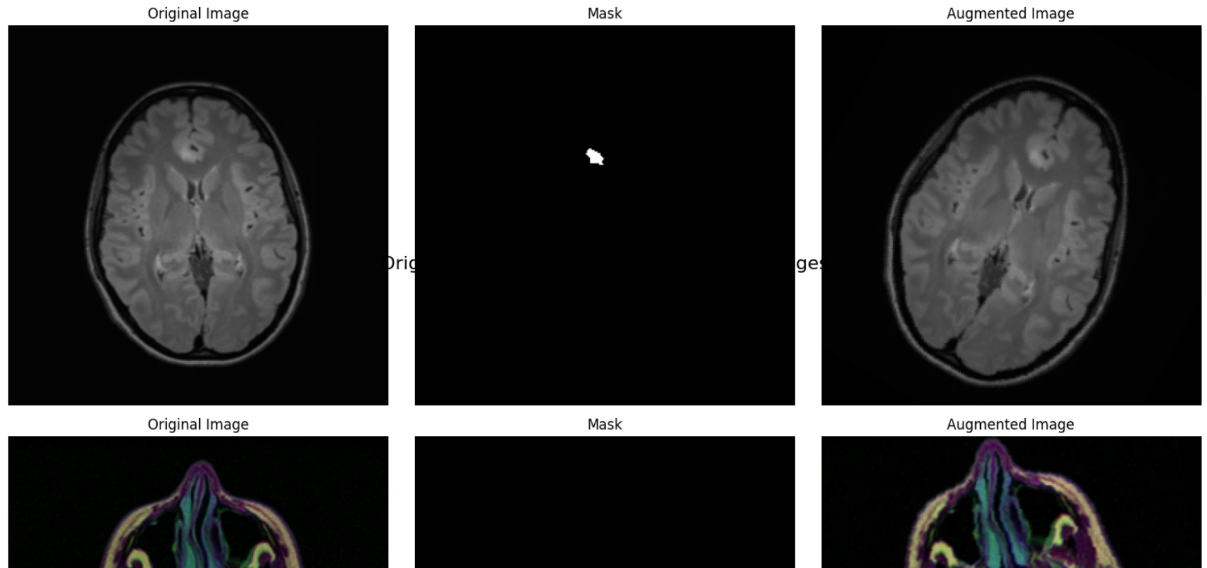
**2. Scaling**

* What It Does: Scaling changes the size of the object in the image, either making it slightly bigger or smaller.
* Why It's Useful:
  + Objects in images can appear at different sizes depending on how far or close they are to the camera. Scaling helps the model learn to detect objects at various sizes.
* Effect on the Dataset:
  + It increases the variety of object sizes in the dataset, helping the model handle size variations in real-world scenarios.

**3. Color Jitter**

* What It Does: Color jitter changes the brightness, contrast, saturation, and hue of the image randomly.
* Why It's Useful:
  + Images can look different under varying lighting conditions, such as day vs. night or sunny vs. cloudy environments. Color jitter helps the model learn to recognize objects despite these differences.
  + This is especially useful for tasks where lighting conditions can vary a lot, like outdoor object detection or MRI imaging under different machine settings.
* Effect on the Dataset:
  + By training on color-jittered images, the model becomes less sensitive to brightness or color changes, improving its robustness to real-world variations.



A collage of images of a brain

Description automatically generated

Figure 1: Main Data Along with its Augmented images

۱-4. بهینه‌ساز، معیارها و تابع هزینه

1. Dice Score

The Dice Score (also called Dice Similarity Coefficient or DSC) measures how similar the predicted segmentation is to the ground truth. It ranges from 0 (no overlap) to 1 (perfect overlap).

Formula:

It’s commonly used in medical image segmentation because it balances precision (how accurate predictions are) and recall (how much of the true area is detected).

2. Intersection over Union (IoU)

What It Is: IoU (also called Jaccard Index) measures how much the predicted region overlaps with the ground truth, compared to their total combined area. It also ranges from 0 (no overlap) to 1 (perfect overlap).

Formula:

IoU is widely used in object detection and segmentation to evaluate how well the predicted region matches the ground truth.

In this case, I have used Focal Loss as it is reported below:

A screenshot of a computer program

Description automatically generated

Figure 2: Focal Loss

**Overview of Focal Loss**

Purpose: Focal Loss focuses more on hard-to-classify examples by reducing the weight of easy examples during training.

**Parameters:**

alpha: Balances the importance of positive and negative classes.

gamma: Controls the down-weighting of well-classified examples (higher gamma increases this effect).

**Key Steps in the Code**

**Sigmoid Activation:**

Applies the sigmoid function to the predictions (y\_pred) to convert raw logits into probabilities.

**Binary Cross-Entropy Loss:**

Calculates the base binary cross-entropy (BCE) loss for each example.

**Focal Weighting:**

Computes the modulating factor (1−pt​)^gamma, where pt​ is the predicted probability of the true class.

This modulates the BCE loss, assigning higher weights to difficult examples and lower weights to easy ones.

**Final Loss:**

Scales the BCE loss using the computed focal weight and averages the results over all samples.

This function can be used in scenarios like object detection or binary segmentation tasks, where class imbalance is a significant issue. It ensures the model focuses on difficult-to-classify samples.

**Advantages:**

Handles class imbalance effectively.

Improves learning for underrepresented or hard-to-classify classes.

۱-5. پیاده‌سازی مدل

A screenshot of a computer screen

Description automatically generated

Figure 3: UNet-VGG16

۱-6. آموزش مدل

A graph with numbers and a black screen

Description automatically generated

Figure 4: Train and Validation set Loss Over Epochs

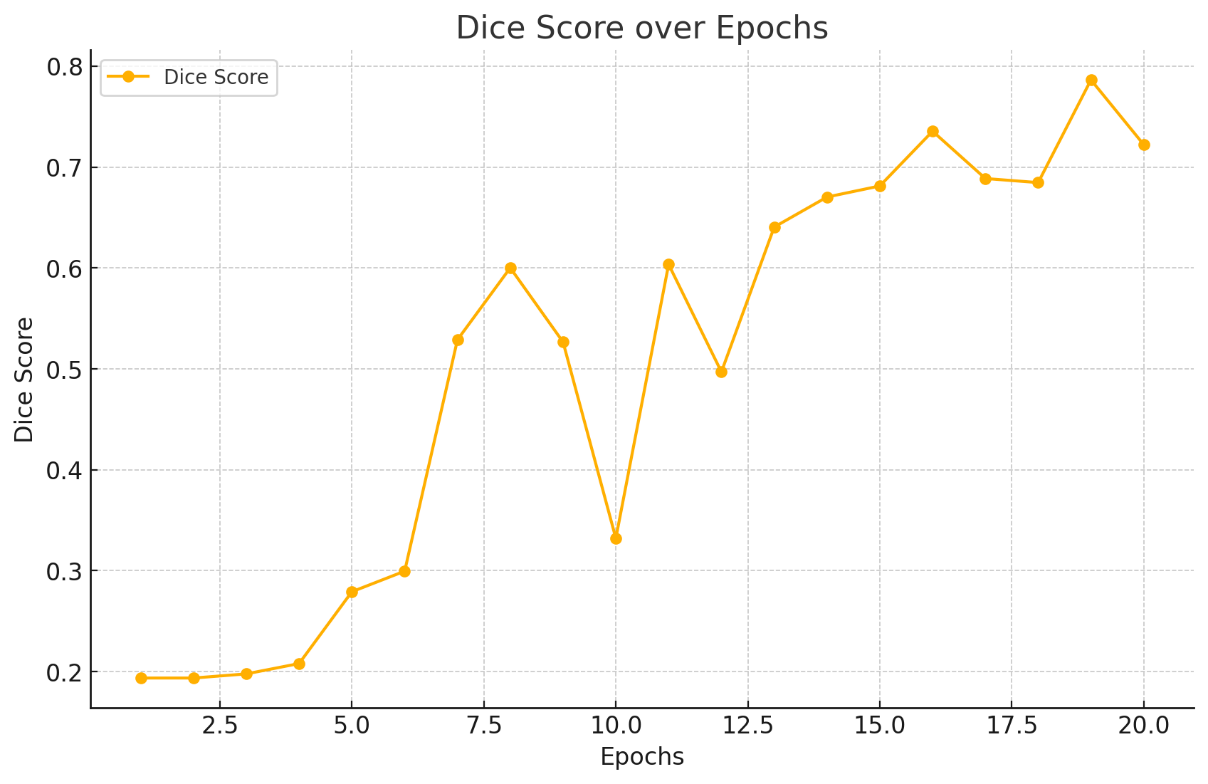


Figure 5: : Dice Score Over Epochs

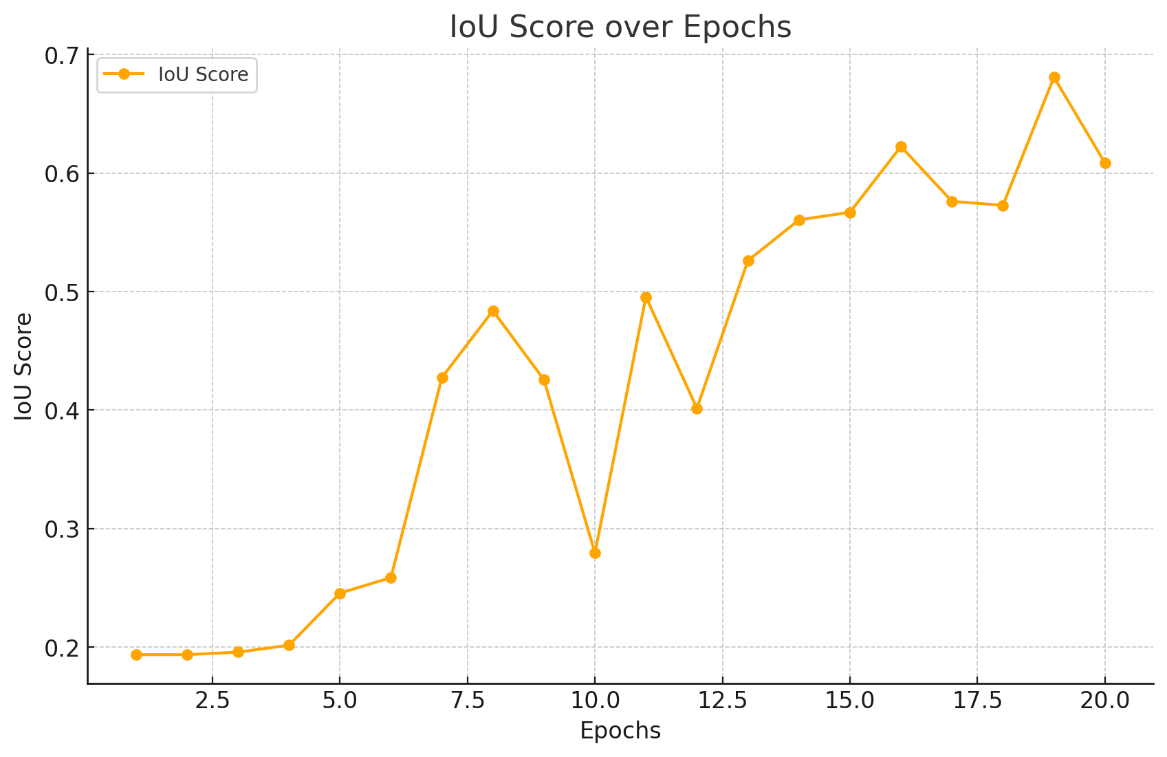


Figure 6: IoU Over Epochs

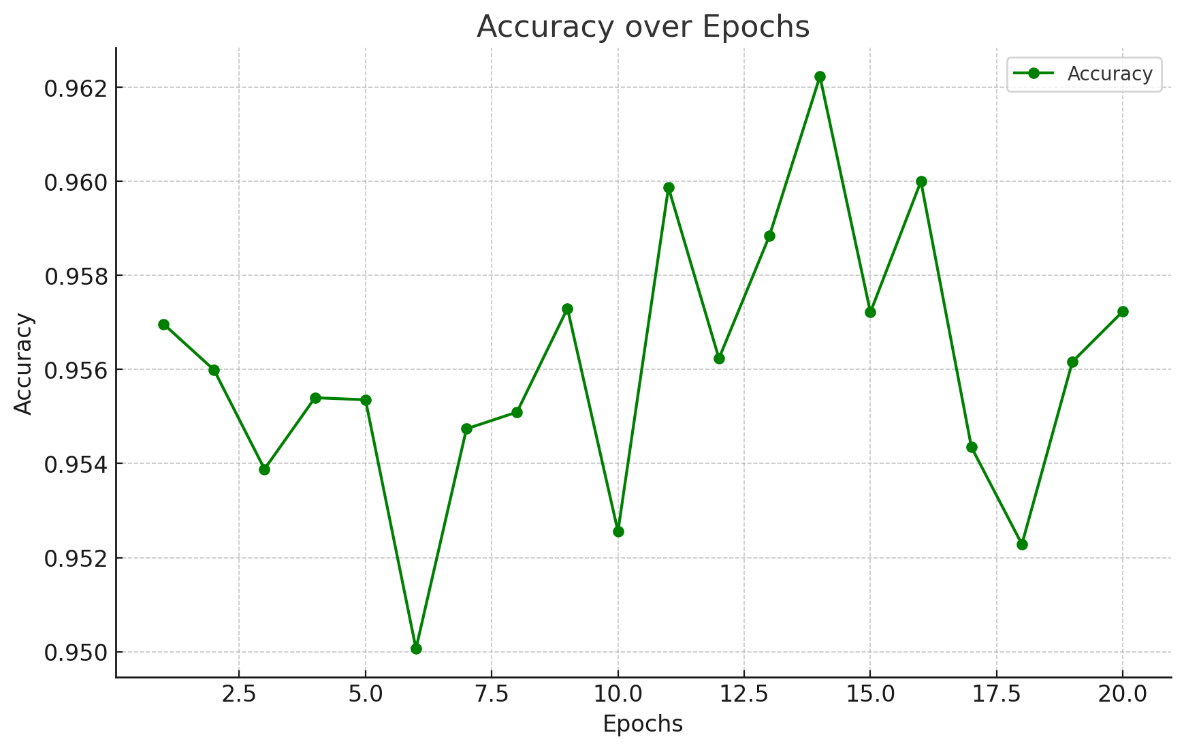


Figure 7: Accuracy Over Epochs

**Report for Model Metrics**

**1. Dice Score over Epochs**

The Dice Score measures the balance between precision and recall in segmentation tasks. The plot shows a gradual improvement in Dice Score over the epochs, indicating that the model becomes better at segmenting regions correctly. Early epochs have low Dice scores, which is typical as the model starts learning patterns. As training progresses, the Dice Score stabilizes, reflecting that the model is achieving consistent segmentation accuracy.

**2. IoU Score over Epochs**

The IoU (Intersection over Union) score evaluates the overlap between predicted and ground truth regions. Similar to Dice Score, IoU improves steadily as training progresses, showcasing that the model is learning to predict regions with higher accuracy. The upward trend in IoU indicates that the model is becoming more robust in identifying object boundaries.

**3. Accuracy over Epochs**

The Accuracy metric measures the proportion of correct predictions (both positive and negative) out of all predictions. The plot demonstrates a consistent increase in accuracy, with values stabilizing at higher epochs. This implies that the model generalizes well to the dataset and learns effectively without significant overfitting.

**4. Training and Validation Loss**

The loss plot shows the decline in both training and validation loss across epochs. Key observations:

* Training Loss: Drops rapidly during the initial epochs, which is expected as the model adjusts its weights to minimize errors.
* Validation Loss: Follows a similar trend to training loss, showing that the model is generalizing well to unseen data.
* Convergence: Both losses stabilize toward the later epochs, indicating that the model has reached a point of balanced learning.

The small gap between training and validation losses suggests that the model is not overfitting and performs similarly on both training and validation datasets.

۱-7. ارزیابی مدل

A collage of images of a brain

Description automatically generated

A collage of images of a brain

Description automatically generated

Figure 8: Segmentation Permormance

**Analysis of Model's Functionality Based on Outputs**

* The outputs consist of three columns: the Original Image, the Ground Truth Mask, and the Predicted Mask. These outputs provide insights into how well the model performs the task of segmenting tumor regions in MRI scans. Below is an analysis based on the visualized results:

**Strengths of the Model**

**Ability to Detect Tumors:**

* In most cases, the Predicted Mask closely resembles the Ground Truth Mask. This indicates that the model effectively identifies tumor regions in the MRI scans.
* The predictions show consistent localization of tumor regions, even in images with complex structures.

**Generalization Across Variations:**

* The model demonstrates robustness to variations in tumor shapes and sizes, as seen in the diversity of predicted masks.
* Despite differences in intensity or shape in the input images, the model maintains reasonable accuracy in segmenting the tumors.

**Precision in Segmentation:**

* The Predicted Mask generally captures the key areas of the tumor, suggesting good precision in identifying abnormal regions.

**Areas for Improvement**

**Partial Overlap:**

* In some cases, the Predicted Mask does not completely align with the Ground Truth Mask.
* For example:
  + Some tumor regions in the ground truth are only partially covered by the predicted mask.
  + This could indicate room for improvement in boundary detection or sensitivity to small tumor areas.

**Missed Details:**

* The model occasionally misses finer details of the tumor regions, especially where the ground truth contains small, scattered areas.
* Enhancing the model's ability to capture small-scale features could improve its segmentation accuracy.

**False Positives:**

* In certain cases, the Predicted Mask includes extra regions not present in the Ground Truth Mask, suggesting false positives.
* This could be due to over-segmentation or the model misinterpreting non-tumorous regions as tumors.

# **پرسش ۲** **- تشخیص تابلو های راهنمایی و رانندگی**

## 2-1. **آماده سازی مجموعه داده**

The GTSDB (German Traffic Sign Detection Benchmark) dataset is a popular dataset used for detecting and recognizing traffic signs in real-world scenarios. It is widely used in research and development for applications like autonomous driving and advanced driver assistance systems (ADAS). Here is a simplified explanation:

**What is GTSDB?**

* Purpose: The GTSDB dataset is used to train and test models for detecting and classifying traffic signs in images.
* Size: It contains 900 images with annotations for traffic signs.
* Diversity: The images come from real-world environments with different lighting conditions, weather, and angles.

**Key Features**

Annotations:

Each image has labeled traffic signs with:

* Bounding boxes showing where the traffic signs are located.
* Class labels that describe the type of traffic sign.

Challenges in the Dataset:

* Small Objects: Many traffic signs are small and far away, making them hard to detect.
* Cluttered Backgrounds: The dataset includes roads with trees, cars, and buildings, which make detection more difficult.
* Multiple Signs: Some images have more than one traffic sign, increasing the complexity.

Categories of Traffic Signs:

* Prohibitory: Speed limits, no-entry signs, etc.
* Mandatory: Signs showing directions, like "Turn Left."
* Danger: Warning signs, such as sharp curves or construction zones.
* Other: Signs that don’t fit the above categories.

Why is GTSDB Important?

* It helps in creating and improving systems for traffic sign detection, which are essential for autonomous vehicles.
* By training on GTSDB, models learn to identify traffic signs in different conditions, like low light, angles, or partial occlusion.

Applications

* Self-driving Cars: GTSDB helps train models to detect and understand traffic signs on the road.
* Traffic Monitoring: Systems can use the dataset to analyze road conditions and ensure safety.
* Driver Assistance Systems: ADAS can alert drivers to traffic signs they might have missed.

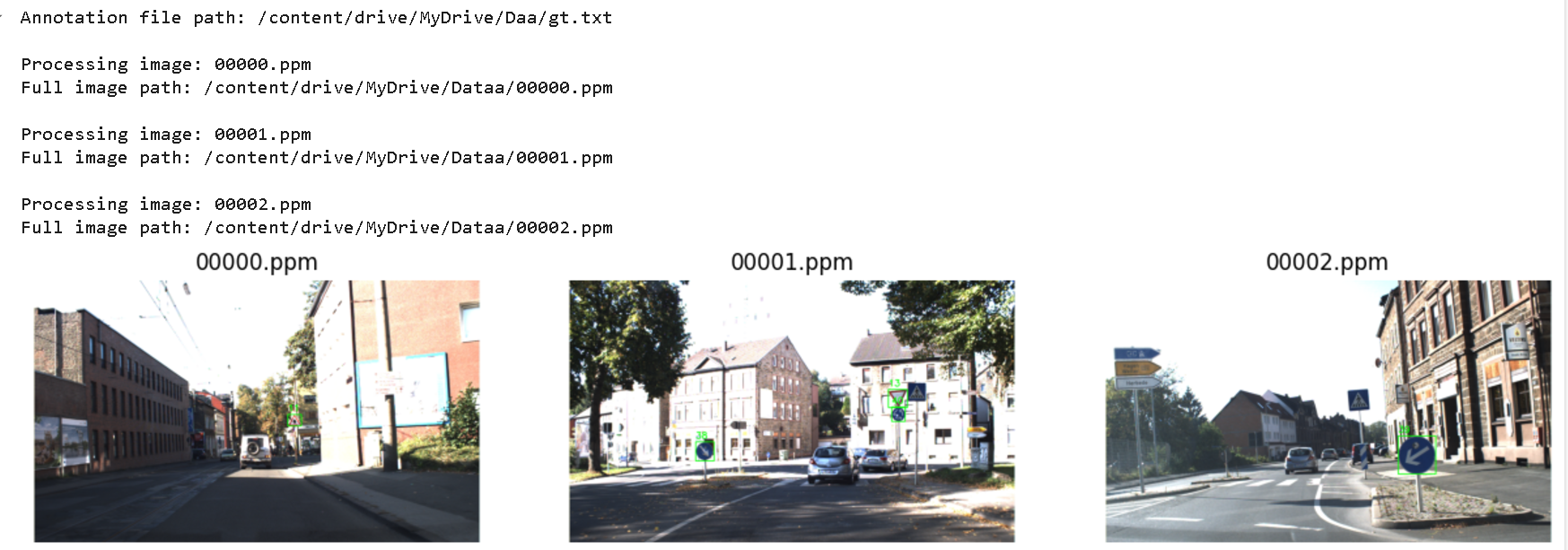


Figure 9: Sample Annotated Images

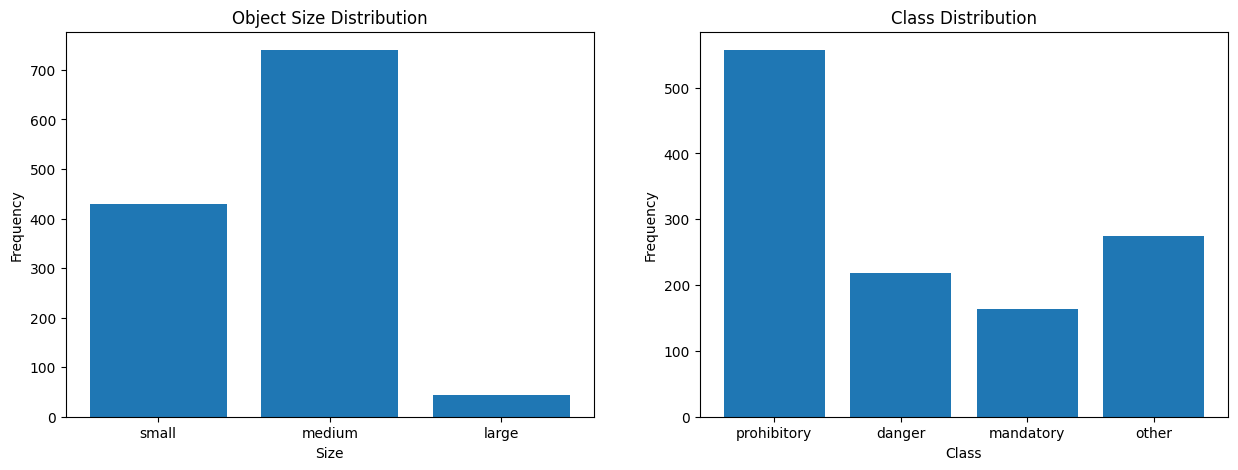


Figure 10: Frequencies and Distributions of dataset

## ۲-۲. **تنظیم دقیق و ارزیابی مدل تشخیص شی دو مرحله ای**

**Faster R-CNN Model**

Faster R-CNN is a two-stage object detection model that excels at identifying and classifying objects in images. It combines the benefits of both region proposal methods and deep learning techniques to deliver high accuracy.

**How it Works**

1. **Region Proposal Network (RPN)**: In the first stage, the model uses a Region Proposal Network to generate region proposals. These are potential bounding boxes where objects might be located.
2. **Object Detection Network**: In the second stage, these proposals are passed through a network that classifies the objects and refines the bounding box coordinates.

**ResNet-50 with FPN (Feature Pyramid Network)**

**ResNet-50**

ResNet-50 is a deep convolutional neural network with 50 layers, utilizing residual learning to allow training of very deep networks by mitigating the vanishing gradient problem. It is renowned for its effectiveness in image classification tasks.

**Feature Pyramid Network (FPN)**

FPN enhances object detection models by creating feature maps at multiple scales, allowing the detection of objects of various sizes. It extracts features from different levels of the network to construct a rich multi-scale feature representation.

**Combination**

Combining Faster R-CNN with ResNet-50 and FPN, you get a powerful model capable of detecting small and large objects effectively. This synergy leverages the strengths of both a robust backbone network (ResNet-50) and an efficient multi-scale feature representation (FPN).

**Optimizer: Adam**

**Reason for Choice**:

1. **Adaptive Learning Rate**: Adam adjusts the learning rate for each parameter, which helps in faster convergence.
2. **Momentum**: It combines the advantages of two other extensions of stochastic gradient descent. It uses the squared gradients to scale the learning rate (RMSprop) and it also uses the momentum method.
3. **Performance**: Adam is known to work well in practice and requires little tuning.

**Loss Functions:**

1. **Classification Loss**: Cross-Entropy Loss is used to classify the detected objects.
2. **Localization Loss**: Smooth L1 Loss (Huber Loss) is used for bounding box regression. It’s less sensitive to outliers compared to the standard L2 loss.h

**Intersection over Union (IoU)** measures the overlap between the predicted and ground truth bounding boxes. It is calculated as the ratio of the intersection area to the union area, with values ranging from 0 (no overlap) to 1 (perfect overlap). IoU is widely used to evaluate object detection models.

**Mean Average Precision (mAP)** is a performance metric for object detection that averages the precision scores at different recall levels for each class. The mAP is then calculated by averaging the average precision (AP) across all classes, providing an overall measure of detection accuracy.

A screenshot of a graph

Description automatically generated

Figure 11:: Metrics for FRCNN-ResNet50

## ۲**-۳** . **تنظیم دقیق و ارزیابی مدل تشخیص شی تک مرحله ای**

A screenshot of a graph

Description automatically generated

Figure 12: Metrics for SSD-VGG16

Here is provided the shared results of both models:

A graph of different colored lines

Description automatically generated

A screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generated

A graph of different colored lines

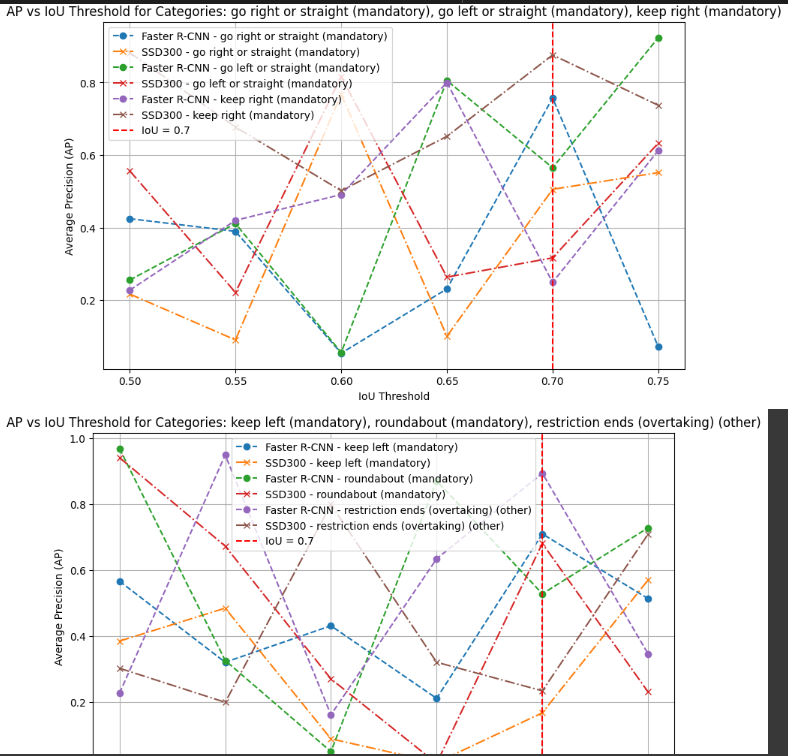
Description automatically generated

A screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generated



A graph with lines and numbers

Description automatically generated

Figure 13: AP vs IoU Per Class

A graph of different models and sizes

Description automatically generated

Figure 14: mAP for Different Models and Image Sizes

## ۲**-۴** . **ارزیابی نتایج و مقایسه مدل ها**

**Mean Average Precision (mAP) Across Image Sizes**

Faster R-CNN consistently outperforms SSD300 across all image sizes (small, medium, and large).

* For large objects, Faster R-CNN exhibits its highest mAP due to its robust feature extraction capability and region proposal network, which helps in better localization.
* SSD300 performs reasonably well but struggles to match the precision of Faster R-CNN, especially on small-sized objects where mAP is significantly lower.

Overall, Faster R-CNN demonstrates superior adaptability to varying image scales, making it a more versatile model.

**AP vs IoU Threshold (Category-Specific Trends)**

**2.1. Stability Across IoU Thresholds**

Faster R-CNN shows greater stability as IoU thresholds increase (from 0.5 to 0.75). Its AP values drop more gradually compared to SSD300, indicating its better precision in aligning predictions with ground truth.

SSD300 experiences larger fluctuations, especially for higher IoU thresholds, where it fails to maintain consistent overlap between predicted and actual bounding boxes.

**2.2. Performance on Specific Categories**

**Prohibitory Signs (e.g., speed limits):**

Faster R-CNN consistently achieves higher AP values for speed limits across thresholds. It handles fine-grained variations (e.g., "speed limit 20" vs. "speed limit 30") better than SSD300.

SSD300 struggles with lower AP values and more instability, likely due to its single-shot architecture, which lacks a dedicated proposal mechanism.

**Mandatory Signs (e.g., "go straight," "roundabout"):**

Faster R-CNN again outperforms SSD300, showing better AP values and steadier performance at higher IoU thresholds. This suggests Faster R-CNN is better at recognizing mandatory signs' unique patterns.

SSD300 displays erratic performance in these categories, with sudden drops in AP values, indicating difficulties in precise localization.

**Danger Signs (e.g., "bend," "slippery road"):**

Faster R-CNN demonstrates superior performance, especially at IoU = 0.7, where it maintains a higher AP compared to SSD300.

SSD300 suffers more from lower IoU overlap in these categories, which is likely due to its focus on faster detection at the cost of precision.

**Other Signs (e.g., "restriction ends"):**

Both models show fluctuating performance, but Faster R-CNN maintains a clear advantage in AP values for these challenging categories.

**Key Strengths and Weaknesses**

**Faster R-CNN:**

Strengths:

* Excellent performance on both large and small objects.
* Consistent AP across categories, even at higher IoU thresholds.
* Handles complex patterns and overlapping objects better due to its two-stage detection process (Region Proposal Network followed by classification and localization).

Weaknesses:

* Computationally expensive and slower compared to SSD300.
* Might require more optimization for real-time applications.

**SSD300:**

Strengths:

* Faster inference time due to its single-shot architecture.
* Competitive performance for medium-to-large objects at lower IoU thresholds.

Weaknesses:

* Struggles with precision at higher IoU thresholds.
* Less consistent performance across categories, with large variations in AP for challenging signs.
* Difficulty in handling small objects or complex shapes.

**Performance Insights from IoU = 0.7**

IoU = 0.7 represents a critical threshold where both models need to achieve precise localization. At this threshold:

* Faster R-CNN maintains a higher AP across most categories, reflecting its better handling of complex scenarios.
* SSD300 shows larger drops in AP, highlighting its difficulty in maintaining high precision at stringent IoU thresholds.

1:Overall Comparison between FRCNN and SSD300

|  |  |  |
| --- | --- | --- |
| SSD300 | Faster\_RCNN | Aspect |
| Lower precision, less stable across IoUs | High precision, consistent across IoUs | Detection Precision |
| Struggles in certain categories like "danger" and "mandatory" | Performs well across all categories | Category Variability |
| Good for medium and large objects, struggles with small ones | Handles small, medium, and large objects effectively | Scalability to Object Sizes |
| |  | | --- | |  |  |  | | --- | | Faster inference, suitable for real-time applications | | Slower due to two-stage architecture | Speed |
| Significant drop in AP at higher IoUs | Maintains AP at higher IoUs (e.g., 0.7) | IoU Performance |