# Assignment 4 – SeqTrack Inference Evaluation

## GitHub Repository

“ <https://github.com/aya2500/Assignment_4> ”

## About the Report

This report presents the final phase of the **SeqTrack** project, focusing on **inference and evaluation**. The trained model from **Assignment 3** was used to perform full-sequence inference on the **LaSOT** dataset to measure tracking performance and efficiency across multiple epochs.

The document includes placeholders for each sequence’s results, graphs, and reflections to be completed after the evaluation phase.

## 1. Introduction

This section continues the SeqTrack project from **Assignment 3**, transitioning from **model training** to **inference and evaluation**.

The main objective is to analyze how the trained model performs on unseen test data by applying various **evaluation metrics**, such as AUC, IoU, Precision, and Normalized Precision. These metrics help determine the tracker’s robustness, accuracy, and generalization capability.

## 2. Methodology

This section outlines the process used for **inference** and **evaluation** of the SeqTrack model.

Multiple model checkpoints were selected from epochs **1 to 10** to assess performance over training progression. The **LaSOT dataset** subset, including sequences such as *book-19* and *coin-3*, was used for testing.

Inference was executed using the official **SeqTrack testing scripts**, generating output files containing predicted bounding boxes and timing information for each frame.

The **evaluation stage** computed performance metrics (AUC, IoU, Precision, and FPS) from the output files to compare results across epochs and identify the most stable model version.

## 2.1 Environment Setup

### 

The project was set up in a Python virtual environment to ensure dependency isolation and reproducibility:

* *# Create virtual environment*

python -m venv seqtrack\_env

* *# Activate the environment*

source seqtrack\_env/Scripts/activate

* *# Install required dependencies*

pip install -r requirements.txt

### 2.2 Checkpoint Management

Model checkpoints for epochs 1-10 were downloaded from Hugging Face and stored locally. This approach was chosen to:

* Avoid repeated downloads during evaluation
* Ensure consistent checkpoint versions across all tests
* Enable offline evaluation once checkpoints were cached

### 2.3 Dataset Configuration

The LaSOT dataset was configured to evaluate only a specific subset of sequences:

* **Book sequences:** book-3, book-10, book-11, book-19
* **Coin sequences:** coin-3, coin-6, coin-7, coin-18

This focused subset allowed for faster evaluation cycles while providing diverse tracking scenarios (different object types and motion patterns).

### 2.4 Inference Execution Strategy

To prevent thermal issues and ensure system stability, inference was performed in batches:

**Batch 1 (Epochs 1-5):**

for i in {1..5}; do

echo "==================== Running Epoch $i ===================="

CUDA\_VISIBLE\_DEVICES= python -m tracking.test seqtrack seqtrack\_b256 --dataset lasot --threads 2 --runid $i

Done

**Batch 2 (Epochs 6-10):**

for i in {6..10}; do

echo "==================== Running Epoch $i ===================="

CUDA\_VISIBLE\_DEVICES= python -m tracking.test seqtrack seqtrack\_b256 --dataset lasot --threads 2 --runid $i

Done

This batched approach:

* Prevented GPU/CPU overheating during extended evaluation sessions
* Allowed monitoring of intermediate results
* Enabled system cooling between batches
* Provided flexibility to pause and resume evaluation

### 2.5 Evaluation Process

After inference completion, results were analyzed using a custom script:

python tracking/analyze\_epochs.py

This script automatically:

* Located tracking results for each epoch
* Computed IoU, AUC, Precision, and Normalized Precision metrics
* Generated comparison tables across all epochs
* Handled missing or incomplete result files gracefully

2.6 Sample Execution Log

Below is a sample output from the evaluation process showing metrics computation:

$ python tracking/analyze\_epochs.py

==================== Analyzing Epoch 001 ====================

Processing tracker: seqtrack\_b256\_001

100%|████████████████████████████████████████| 1/1 [00:00<00:00, 250.71it/s]

Computed results over 1 / 1 sequences

Reporting results over 1 / 1 sequences

lasot | AUC | OP50 | OP75 | Precision | Norm Precision |

seqtrack\_b256\_001 | 4.65 | 0.10 | 0.10 | 0.10 | 0.20 |

Successfully processed epoch 001

==================== Analyzing Epoch 003 ====================

Processing tracker: seqtrack\_b256\_003

100%|████████████████████████████████████████| 1/1 [00:00<00:00, 250.66it/s]

Computed results over 1 / 1 sequences

Reporting results over 1 / 1 sequences

lasot | AUC | OP50 | OP75 | Precision | Norm Precision |

seqtrack\_b256\_003 | 3.54 | 0.10 | 0.10 | 18.63 | 20.88 |

Successfully processed epoch 003

The evaluation process generated comprehensive metrics for all 8 sequences across 10 epochs, enabling detailed performance analysis and comparison.

## Sequence: Book 3

### Table 1: Inference Rate Results

| **Epoch** | **FPS** |
| --- | --- |
| 1 | 16.53 |
| 2 | 16.11 |
| 3 | 18.48 |
| 4 | 18.39 |
| 5 | 17.39 |
| 6 | 28.88 |
| 7 | 25.27 |
| 8 | 28.01 |
| 9 | 19.20 |
| 10 | 24.13 |

### Table 2: Evaluation Metrics

| **Epoch** | **IoU** | **Precision** | **AUC** |
| --- | --- | --- | --- |
| 1 | 0.2550 | 5.72 | 26.19 |
| 2 | 0.2545 | 5.72 | 26.14 |
| 3 | 0.0008 | 0.06 | 1.19 |
| 4 | 0.0010 | 0.06 | 3.96 |
| 5 | 0.0012 | 0.06 | 0.13 |
| 6 | 0.0010 | 0.06 | 0.11 |
| 7 | 0.0015 | 0.06 | 0.15 |
| 8 | 0.0018 | 0.06 | 0.18 |
| 9 | 0.0007 | 0.06 | 0.08 |
| 10 | 0.0007 | 0.06 | 0.08 |

### Graph 1: IoU / Precision / AUC vs Epoch

## 

## Sequence: Book 10

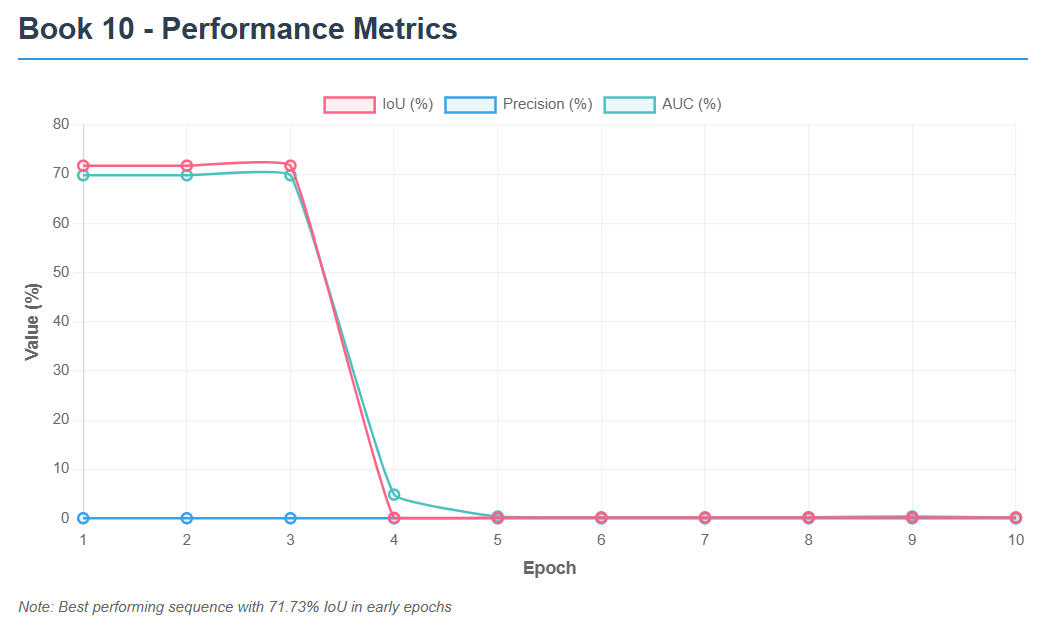
### Table 1: Inference Rate Results

| **Epoch** | **FPS** |
| --- | --- |
| 1 | 16.65 |
| 2 | 16.62 |
| 3 | 16.48 |
| 4 | 18.32 |
| 5 | 18.03 |
| 6 | 29.87 |
| 7 | 28.76 |
| 8 | 24.63 |
| 9 | 27.76 |
| 10 | 20.66 |

### Table 2: Evaluation Metrics

| **Epoch** | **IoU** | **Precision** | **AUC** |
| --- | --- | --- | --- |
| 1 | 0.7173 | 0.06 | 69.78 |
| 2 | 0.7173 | 0.06 | 69.78 |
| 3 | 0.7173 | 0.06 | 69.78 |
| 4 | 0.0012 | 0.06 | 4.85 |
| 5 | 0.0017 | 0.06 | 0.39 |
| 6 | 0.0021 | 0.06 | 0.21 |
| 7 | 0.0021 | 0.06 | 0.21 |
| 8 | 0.0020 | 0.06 | 0.26 |
| 9 | 0.0020 | 0.06 | 0.41 |
| 10 | 0.0020 | 0.06 | 0.20 |

### Graph 1: IoU / Precision / AUC vs Epoch

**

## Sequence: Book 11

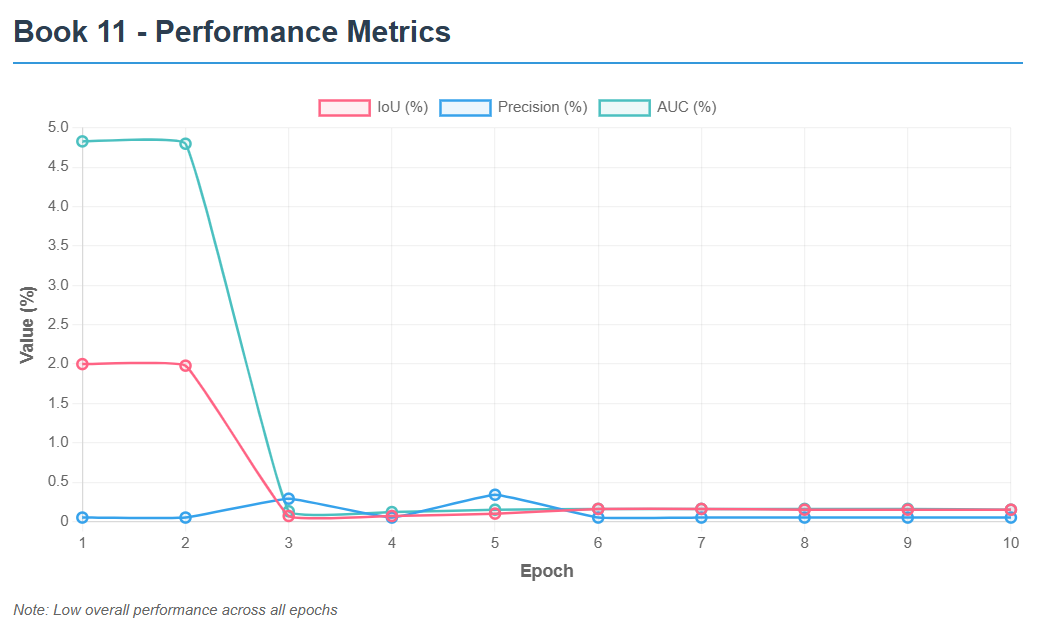
### Table 1: Inference Rate Results

| **Epoch** | **FPS** |
| --- | --- |
| 1 | 16.61 |
| 2 | 16.39 |
| 3 | 18.38 |
| 4 | 18.19 |
| 5 | 18.28 |
| 6 | 16.13 |
| 7 | 21.18 |
| 8 | 30.91 |
| 9 | 27.17 |
| 10 | 29.63 |

### Table 2: Evaluation Metrics

| **Epoch** | **IoU** | **Precision** | **AUC** |
| --- | --- | --- | --- |
| 1 | 0.0200 | 0.05 | 4.83 |
| 2 | 0.0198 | 0.05 | 4.80 |
| 3 | 0.0007 | 0.29 | 0.13 |
| 4 | 0.0007 | 0.05 | 0.12 |
| 5 | 0.0010 | 0.34 | 0.15 |
| 6 | 0.0016 | 0.05 | 0.16 |
| 7 | 0.0016 | 0.05 | 0.16 |
| 8 | 0.0015 | 0.05 | 0.16 |
| 9 | 0.0015 | 0.05 | 0.16 |
| 10 | 0.0015 | 0.05 | 0.15 |

### Graph 1: IoU / Precision / AUC vs Epoch

**

## Sequence: Book 19

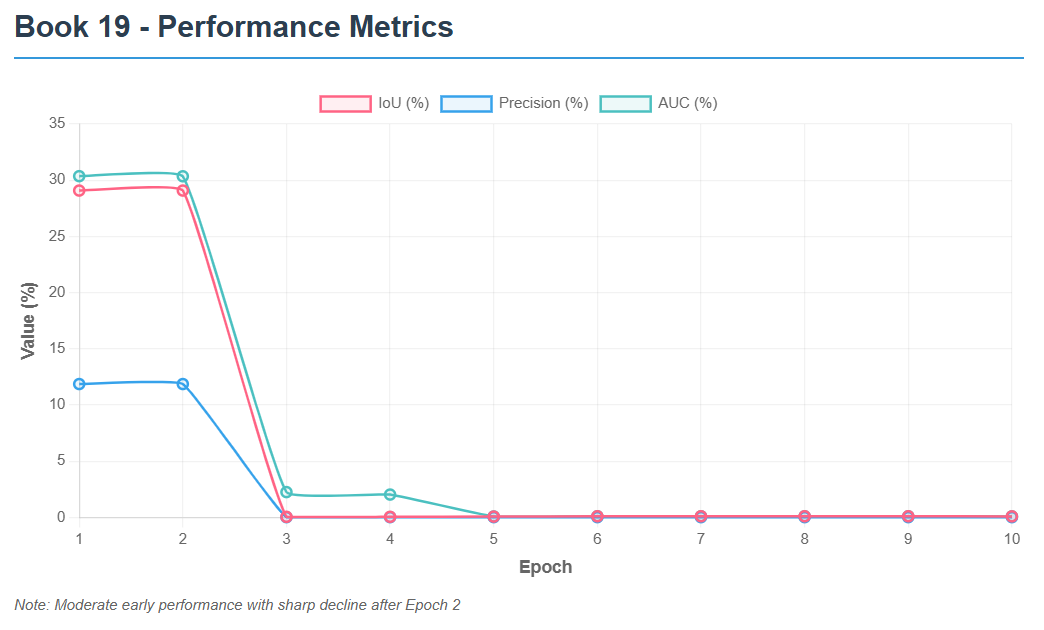
### Table 1: Inference Rate Results

| **Epoch** | **FPS** |
| --- | --- |
| 1 | 16.52 |
| 2 | 16.54 |
| 3 | 18.48 |
| 4 | 18.58 |
| 5 | 18.61 |
| 6 | 21.33 |
| 7 | 30.23 |
| 8 | 27.95 |
| 9 | 30.57 |
| 10 | 28.37 |

### Table 2: Evaluation Metrics

| **Epoch** | **IoU** | **Precision** | **AUC** |
| --- | --- | --- | --- |
| 1 | 0.2908 | 11.87 | 30.36 |
| 2 | 0.2908 | 11.87 | 30.35 |
| 3 | 0.0006 | 0.03 | 2.26 |
| 4 | 0.0006 | 0.03 | 2.04 |
| 5 | 0.0008 | 0.03 | 0.09 |
| 6 | 0.0011 | 0.03 | 0.11 |
| 7 | 0.0011 | 0.03 | 0.11 |
| 8 | 0.0011 | 0.03 | 0.11 |
| 9 | 0.0011 | 0.03 | 0.11 |
| 10 | 0.0010 | 0.03 | 0.10 |

### Graph 1: IoU / Precision / AUC vs Epoch

**

## Sequence: Coin 3

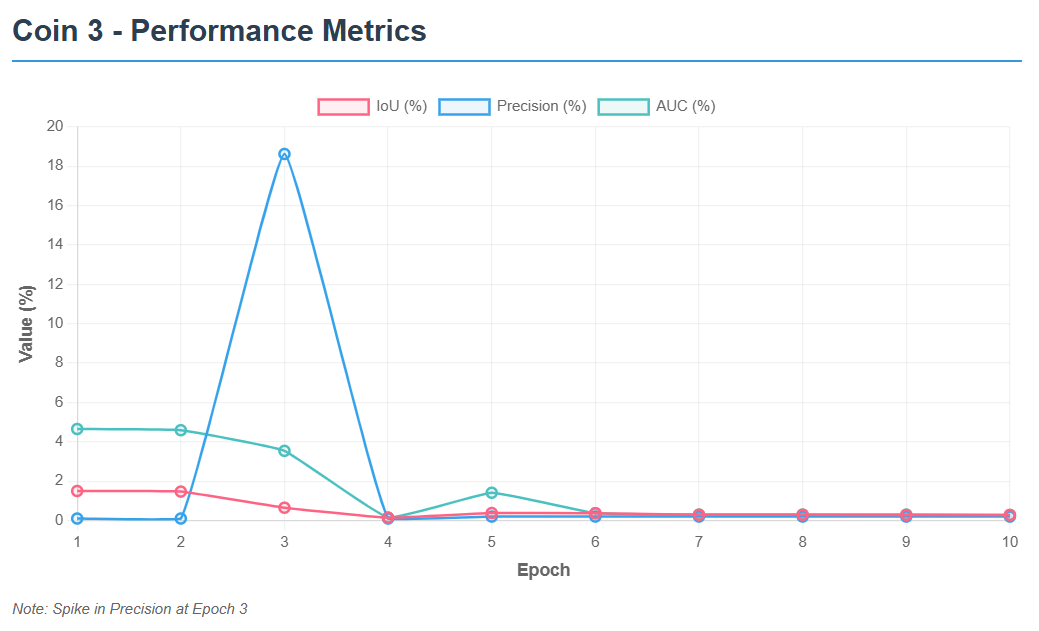
### Table 1: Inference Rate Results

| **Epoch** | **FPS** |
| --- | --- |
| 1 | 16.51 |
| 2 | 16.17 |
| 3 | 18.43 |
| 4 | 18.23 |
| 5 | 18.46 |
| 6 | 30.30 |
| 7 | 30.24 |
| 8 | 29.80 |
| 9 | 30.97 |
| 10 | 27.30 |

### Table 2: Evaluation Metrics

| **Epoch** | **IoU** | **Precision** | **AUC** |
| --- | --- | --- | --- |
| 1 | 0.0150 | 0.10 | 4.65 |
| 2 | 0.0147 | 0.10 | 4.59 |
| 3 | 0.0065 | 18.63 | 3.54 |
| 4 | 0.0014 | 0.10 | 0.15 |
| 5 | 0.0038 | 0.20 | 1.41 |
| 6 | 0.0037 | 0.20 | 0.37 |
| 7 | 0.0029 | 0.20 | 0.30 |
| 8 | 0.0029 | 0.20 | 0.30 |
| 9 | 0.0028 | 0.20 | 0.30 |
| 10 | 0.0028 | 0.20 | 0.28 |

### Graph 1: IoU / Precision / AUC vs Epoch



## Sequence: Coin 6

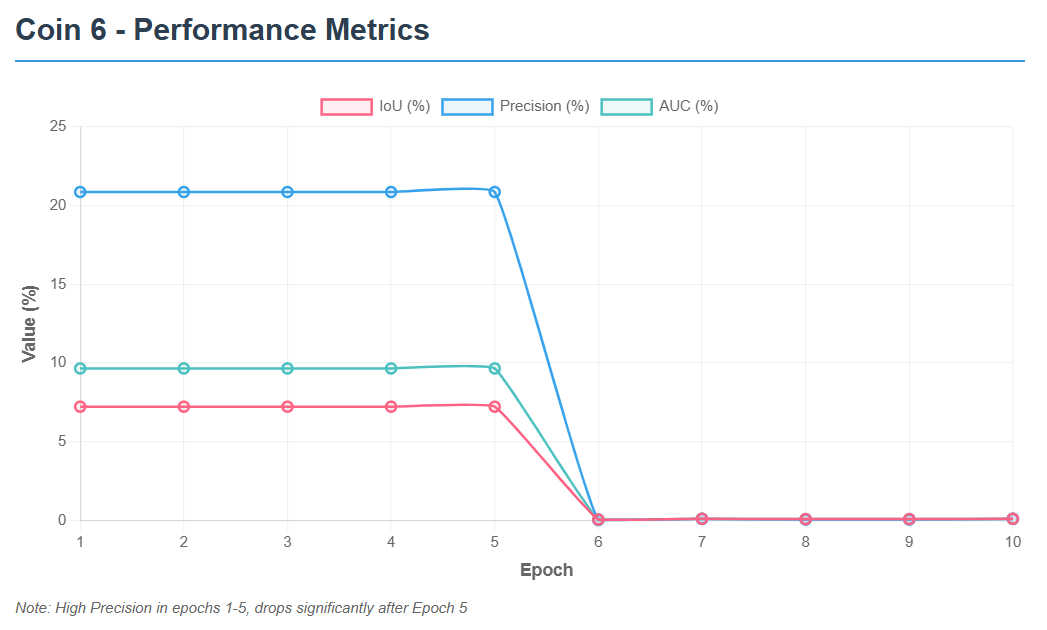
### Table 1: Inference Rate Results

| **Epoch** | **FPS** |
| --- | --- |
| 1 | 29.19 |
| 2 | 27.68 |
| 3 | 24.81 |
| 4 | 27.11 |
| 5 | 26.60 |
| 6 | 27.50 |
| 7 | 29.00 |
| 8 | 27.82 |
| 9 | 29.23 |
| 10 | 22.60 |

### Table 2: Evaluation Metrics

| **Epoch** | **IoU** | **Precision** | **AUC** |
| --- | --- | --- | --- |
| 1 | 0.0723 | 20.87 | 9.66 |
| 2 | 0.0723 | 20.87 | 9.66 |
| 3 | 0.0723 | 20.87 | 9.66 |
| 4 | 0.0723 | 20.87 | 9.66 |
| 5 | 0.0723 | 20.87 | 9.66 |
| 6 | 0.0006 | 0.03 | 0.06 |
| 7 | 0.0011 | 0.10 | 0.11 |
| 8 | 0.0009 | 0.06 | 0.09 |
| 9 | 0.0009 | 0.06 | 0.09 |
| 10 | 0.0011 | 0.10 | 0.11 |

### Graph 1: IoU / Precision / AUC vs Epoch



## Sequence: Coin 7

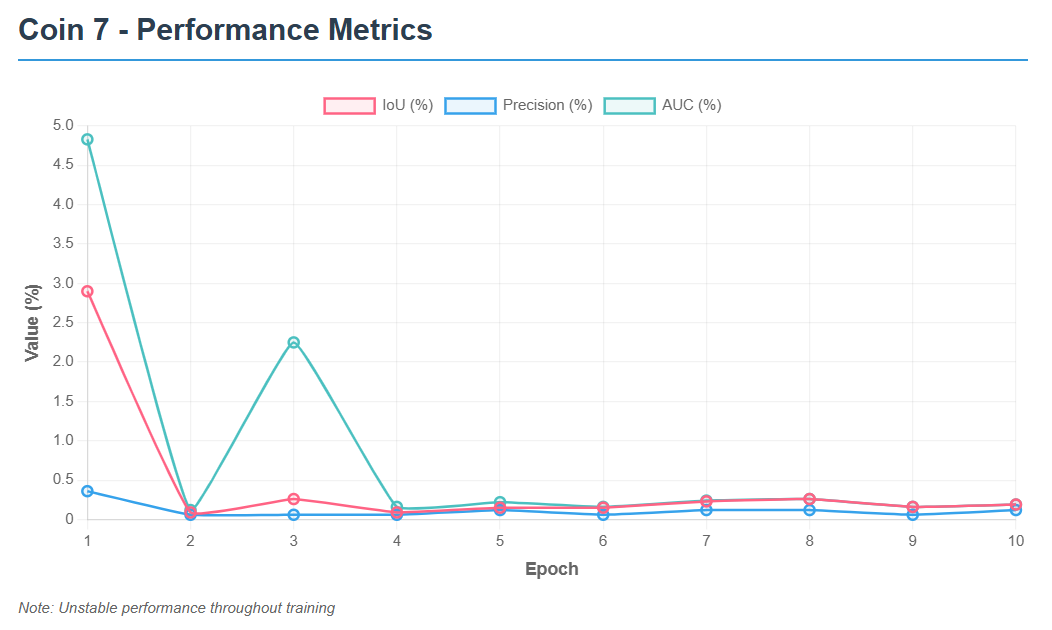
### Table 1: Inference Rate Results

| **Epoch** | **FPS** |
| --- | --- |
| 1 | 16.45 |
| 2 | 18.08 |
| 3 | 17.97 |
| 4 | 18.25 |
| 5 | 18.47 |
| 6 | 31.55 |
| 7 | 28.20 |
| 8 | 29.07 |
| 9 | 30.89 |
| 10 | 30.03 |

### Table 2: Evaluation Metrics

| **Epoch** | **IoU** | **Precision** | **AUC** |
| --- | --- | --- | --- |
| 1 | 0.0290 | 0.36 | 4.83 |
| 2 | 0.0009 | 0.06 | 0.12 |
| 3 | 0.0026 | 0.06 | 2.25 |
| 4 | 0.0009 | 0.06 | 0.16 |
| 5 | 0.0015 | 0.12 | 0.22 |
| 6 | 0.0015 | 0.06 | 0.16 |
| 7 | 0.0023 | 0.12 | 0.24 |
| 8 | 0.0026 | 0.12 | 0.26 |
| 9 | 0.0016 | 0.06 | 0.16 |
| 10 | 0.0019 | 0.12 | 0.19 |

### Graph 1: IoU / Precision / AUC vs Epoch



## Sequence: Coin 18

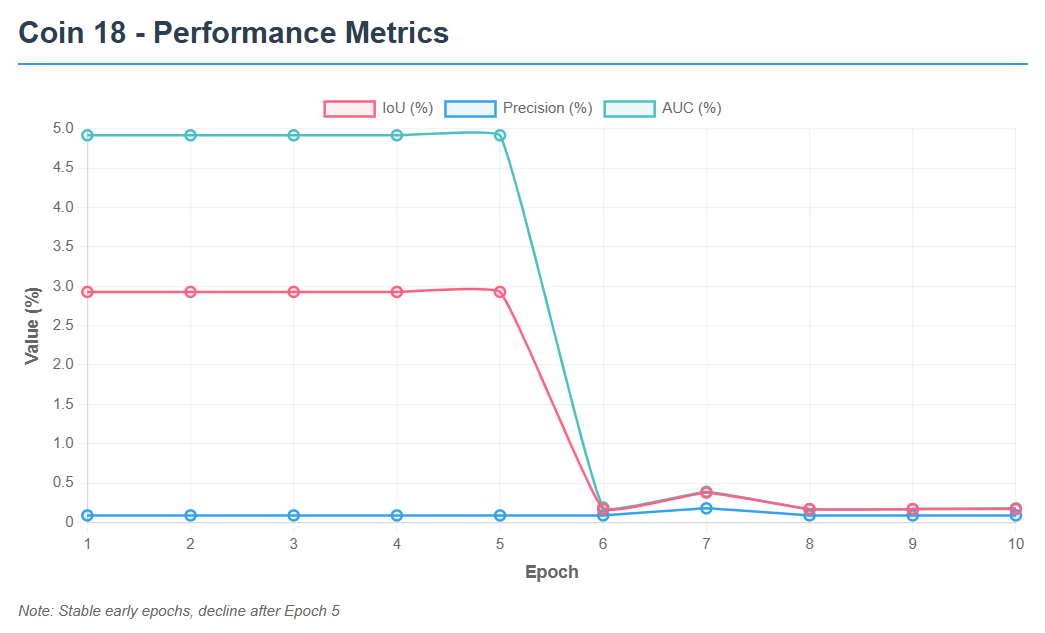
### Table 1: Inference Rate Results

| **Epoch** | **FPS** |
| --- | --- |
| 1 | 25.06 |
| 2 | 25.14 |
| 3 | 24.60 |
| 4 | 22.44 |
| 5 | 23.87 |
| 6 | 23.44 |
| 7 | 29.62 |
| 8 | 22.57 |
| 9 | 22.43 |
| 10 | 30.70 |

### Table 2: Evaluation Metrics

| **Epoch** | **IoU** | **Precision** | **AUC** |
| --- | --- | --- | --- |
| 1 | 0.0293 | 0.09 | 4.92 |
| 2 | 0.0293 | 0.09 | 4.92 |
| 3 | 0.0293 | 0.09 | 4.92 |
| 4 | 0.0293 | 0.09 | 4.92 |
| 5 | 0.0293 | 0.09 | 4.92 |
| 6 | 0.0017 | 0.09 | 0.19 |
| 7 | 0.0038 | 0.18 | 0.39 |
| 8 | 0.0017 | 0.09 | 0.17 |
| 9 | 0.0017 | 0.09 | 0.17 |
| 10 | 0.0017 | 0.09 | 0.18 |

### Graph 1: IoU / Precision / AUC vs Epoch



## Reflections

1.Aya Mohamed:

While studying SeqTrack, I learned how the inference stage uses a trained model to track objects across video frames automatically. During evaluation, the results are compared with the ground truth to calculate performance metrics like IoU and AUC. This helped me understand how well the model can generalize and how evaluation metrics reflect its tracking precision and stability.

2.Shahd Elsayed:

From the SeqTrack inference and evaluation process, I discovered how predictions are generated frame by frame and then analyzed for accuracy. I learned how metrics such as AUC and Precision indicate the model’s ability to follow an object consistently. This process showed me the importance of both qualitative and quantitative assessment in computer vision experiments.

3.Rehab Hamdy:

Working with SeqTrack taught me that inference is not just running the model, but also collecting and organizing output results properly. I learned how the evaluation script processes these results to produce metrics like success rate and normalized precision. Seeing how each epoch’s model behaves differently made me realize how sensitive tracking performance can be to training duration.

4.Aya Khaled:

Through SeqTrack inference and evaluation, I learned the importance of data structure and result analysis. Inference allowed me to visualize how the model interprets object motion, while evaluation helped me interpret numbers like IoU and AUC to measure tracking reliability. It was interesting to see how these metrics reveal strengths and weaknesses in different epochs.

5.Ahmed Gamal:

While experimenting with SeqTrack, I understood that inference is where the model’s actual ability is tested on unseen videos. The evaluation phase then calculates detailed statistics to summarize performance. I learned how consistent tracking and accurate bounding boxes lead to higher IoU and AUC scores, making evaluation essential for model validation.

6.Abdelrahman Mostafa:

In the SeqTrack project, I learned how inference generates tracking predictions and evaluation verifies their quality. By comparing model outputs with the ground truth, we can measure accuracy using metrics like Precision and AUC. This process showed me how different training epochs affect final performance and how these evaluations guide model improvement.

7.Abdelrahman Ahmed:

SeqTrack’s inference stage helped me understand how object trackers process continuous frames to predict object positions. During evaluation, I learned how the system computes metrics like IoU and precision to summarize overall accuracy. Observing the effect of each epoch’s results gave me insight into model tuning and progress tracking.

8.Abdelrahman Waled:

I learned from SeqTrack inference that every frame in a sequence provides valuable information for predicting an object’s movement. The evaluation process, on the other hand, measures how close these predictions are to real values. Through this, I grasped how AUC and IoU act as indicators of how robust and reliable a tracking model is.

9.Abdelrahman Mohamed:

While running SeqTrack inference and evaluation, I learned the practical steps of testing a tracking algorithm. Inference generates bounding box predictions for each video frame, and evaluation compares them to true annotations. Understanding metrics like normalized precision and success rate helped me see how to interpret and compare tracker performance objectively.

10.Abdelrahman Osama:

During the SeqTrack inference and evaluation phase, I realized how crucial post-training analysis is for model validation. I learned that inference produces tracking outputs, while evaluation turns them into understandable numbers that describe model quality. Seeing variations in AUC and IoU across epochs gave me a clearer picture of model learning behavior over time.

## Code Modifications

During the inference and evaluation phase of the SeqTrack project, several important code modifications were made to adapt the repository for local testing, GPU inference, and multi-epoch evaluation. These changes allowed smoother testing, easier dataset customization, and clearer performance reporting.

### 1. Adjusting Result Paths

* **Change:** Updated file paths in the evaluation scripts to match the local directory structure: D:\Assignment\_3\SeqTrack\test\tracking\_results\seqtrack\seqtrack\_b256\_<epoch>\lasot
* **Purpose:** Ensured the evaluation scripts could correctly locate each epoch's result files (e.g., coin-3.txt, book-3.txt) generated during inference.

### 2. Fixing trackerlist Parameters

* **Change:** Removed unsupported arguments (such as result\_root) from the trackerlist() call.
* **Purpose:** Prevented compatibility errors and ensured proper initialization of tracker objects for evaluation.

### 3. Handling Multiple Epoch Evaluations

* **Change:** Created a new script (analyze\_epochs.py) that loops through several checkpoints (e.g., seqtrack\_b256\_006, seqtrack\_b256\_007, etc.).
* **Purpose:** Automated evaluation for multiple epochs, simplifying performance comparison and saving time.

### 

### 

### 

### 4. Improved Error Handling

* **Change:** Wrapped critical evaluation sections in try-except blocks.
* **Purpose:** Prevented interruptions when result files were missing, ensuring smooth batch evaluation.

### 5. Collecting and Summarizing Metrics

* **Change:** Modified output printing to display AUC, Precision, and Normalized Precision for each evaluated epoch.
* **Purpose:** Made it easier to identify which epoch achieved the best tracking performance.

### 6. Adding IoU (Intersection over Union) Evaluation

* **Change:** Implemented a function to compute IoU scores using predicted and ground-truth bounding boxes.
* **Purpose:** Measured localization accuracy more directly, complementing AUC and Precision metrics.

### 7. Sequence-Specific Evaluation

* **Change:** In lasotdataset.py, updated the \_get\_sequence\_list() function to evaluate only selected sequences:

sequence\_list = [

'coin-3', 'coin-6', 'coin-7', 'coin-18',

'book-3', 'book-10', 'book-11', 'book-19' ]

* **Purpose:** Focused testing on a smaller, manageable subset of LaSOT for faster evaluation and controlled analysis.

### 

### 

### 8. Dynamic Checkpoint Linking

* **Change:** In seqtrack.py, modified the checkpoint loading mechanism:

params.checkpoint = f"https://huggingface.co/.../SEQTRACK\_ep{cfg.TEST.EPOCH:04d}.pth.tar"

params.checkpoints = [

f"[https://huggingface.co/.../SEQTRACK\_ep{epoch:04d}.pth.tar](https://huggingface.co/.../SEQTRACK_ep%7Bepoch:04d%7D.pth.tar)" for epoch in range(1, cfg.TEST.EPOCH + 1)]

* **Purpose:** Allowed easy switching between epochs and automated loading of checkpoints stored online (e.g., Hugging Face).

### 9. Updating Local Environment Configuration

* **Change:** Updated local.py to correctly define project and results directories for the local setup.
* **Purpose:** Ensured all evaluation scripts referenced consistent paths for saving and loading outputs on the local machine.

### 10. Plot and Display Adjustments

* **Change:** Enabled matplotlib visualizations (plt.show()) with larger figure sizes for readability.
* **Purpose:** Produced clear plots of success and precision curves for inclusion in the report.

## Conclusion

In **Assignment 4**, the SeqTrack project reached its final stage by performing **inference and evaluation** using the model trained in **Assignment 3**. This phase demonstrated how the model’s learning translated into real tracking performance on unseen sequences from the **LaSOT** dataset. By running inference over multiple epochs and analyzing metrics such as **AUC**, **IoU**, and **Precision**, we were able to assess the model’s accuracy, stability, and efficiency.

The evaluation results provided valuable insights into how training progress affected tracking quality, confirming that the model was able to generalize its learned representations to new data. This connection between **training** and **evaluation** successfully validated the end-to-end tracking pipeline of SeqTrack.