**Task 4:-**

**Github link:** [**https://github.com/Aslamgithub123/assignment-4**](https://github.com/Aslamgithub123/assignment-4)

**huggingface link:**[**https://huggingface.co/datasets/HGuserx9/epoch/tree/main**](https://huggingface.co/datasets/HGuserx9/epoch/tree/main) **Inference tables and graphs:-**

Phase 1 metrics:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| phase | checkpoint | epoch | num\_samples | time\_total\_s | ms\_per\_frame | fps | iou\_mean | precision\_at\_20px | auc\_success |
| phase1 | checkpoint\_epoch\_1.pth | 1 | 39988 | 719.37759 | 17.989837 | 55.586941 | 0.637 | 0.884 | 0.743 |
| phase1 | checkpoint\_epoch\_2.pth | 2 | 39988 | 716.37818 | 17.914829 | 55.819679 | 0.604 | 0.831 | 0.623 |
| phase1 | checkpoint\_epoch\_3.pth | 3 | 39988 | 713.037014 | 17.831275 | 56.08124 | 0.549 | 0.745 | 0.368 |
| phase1 | checkpoint\_epoch\_4.pth | 4 | 39988 | 703.937886 | 17.603728 | 56.806148 | 0.532 | 0.719 | 0.557 |
| phase1 | checkpoint\_epoch\_5.pth | 5 | 39988 | 705.810791 | 17.650565 | 56.65541 | 0.535 | 0.724 | 0.486 |
| phase1 | checkpoint\_epoch\_6.pth | 6 | 39988 | 715.601678 | 17.895411 | 55.880249 | 0.635 | 0.805 | 0.653 |
| phase1 | checkpoint\_epoch\_7.pth | 7 | 39988 | 714.474872 | 17.867232 | 55.968378 | 0.628 | 0.726 | 0.635 |
| phase1 | checkpoint\_epoch\_8.pth | 8 | 39988 | 705.145198 | 17.63392 | 56.708888 | 0.57 | 0.943 | 0.483 |
| phase1 | checkpoint\_epoch\_9.pth | 9 | 39988 | 713.538382 | 17.843813 | 56.041835 | 0.585 | 0.807 | 0.593 |
| phase1 | checkpoint\_epoch\_10.pth | 10 | 39988 | 713.007634 | 17.83054 | 56.083551 | 0.601 | 0.779 | 0.572 |

A graph with blue lines and points

AI-generated content may be incorrect.

A graph with blue lines and dots

AI-generated content may be incorrect.

A graph with blue lines

AI-generated content may be incorrect.

**Phase 2 metrics:**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| phase | checkpoint | epoch | num\_samples | time\_total\_s | ms\_per\_frame | fps | iou\_mean | precision\_at\_20px | auc\_success |
| phase2 | checkpoint\_epoch\_1.pth | 1 | 39988 | 704.534468 | 17.618647 | 56.758046 | 0.556 | 0.757 | 0.453 |
| phase2 | checkpoint\_epoch\_2.pth | 2 | 39988 | 712.084258 | 17.807449 | 56.156276 | 0.649 | 0.902 | 0.563 |
| phase2 | checkpoint\_epoch\_3.pth | 3 | 39988 | 717.310433 | 17.938142 | 55.747133 | 0.602 | 0.828 | 0.434 |
| phase2 | checkpoint\_epoch\_4.pth | 4 | 39988 | 708.855227 | 17.726699 | 56.412083 | 0.526 | 0.71 | 0.643 |
| phase2 | checkpoint\_epoch\_5.pth | 5 | 39988 | 715.960696 | 17.904389 | 55.852228 | 0.574 | 0.784 | 0.532 |
| phase2 | checkpoint\_epoch\_6.pth | 6 | 39988 | 706.437068 | 17.666227 | 56.605184 | 0.608 | 0.911 | 0.611 |
| phase2 | checkpoint\_epoch\_7.pth | 7 | 39988 | 713.015317 | 17.830732 | 56.082947 | 0.555 | 0.906 | 0.622 |
| phase2 | checkpoint\_epoch\_8.pth | 8 | 39988 | 712.678721 | 17.822315 | 56.109435 | 0.563 | 0.723 | 0.599 |
| phase2 | checkpoint\_epoch\_9.pth | 9 | 39988 | 717.05227 | 17.931686 | 55.767204 | 0.545 | 0.709 | 0.657 |
| phase2 | checkpoint\_epoch\_10.pth | 10 | 39988 | 712.193067 | 17.81017 | 56.147696 | 0.647 | 0.714 | 0.671 |

A graph with blue lines and dots

AI-generated content may be incorrect.

A graph with blue lines and dots

AI-generated content may be incorrect.

A graph with blue lines

AI-generated content may be incorrect.

**Filenames and line numbers (key modifications):-**

evaluate.py — line 135: Used class: LaSOTPairDataset for constructing (template, search, gt\_bbox) pairs

evaluate.py — line 78: Performance log setup with FileHandler and StreamHandler

evaluate.py — line 492: Per-checkpoint metrics logging: samples, ms/frame, FPS, IoU, Precision@20px, AUC

evaluate.py — line 240: List checkpoints from Hugging Face repo with optional prefix filtering

evaluate.py — line 253: Download checkpoint file from Hugging Face to local cache

evaluate.py — line 685: Assemble phase1 checkpoint list with prefix and fallbacks

evaluate.py — line 686: Assemble phase2 checkpoint list with prefix and fallbacks

evaluate.py — line 754: Local checkpoint paths produced for phase1 and phase2

**Reflections:-**

student 1:  
Working on SeqTrack inference clarified how IoU, Precision@20px, and AUC jointly describe tracking quality rather than relying on a single metric.​  
Plotting these against epoch made it easier to spot overfitting and stability across checkpoints in both phases.​

student 2:  
Throughput metrics like ms\_per\_frame and FPS helped relate model quality to deployment constraints, reminding that fast but consistent inference matters for trackers.​  
Comparing these values across epochs showed small efficiency shifts that could add up over long sequences.​

student 3:  
Automating checkpoint discovery and download from Hugging Face taught how to manage experiment artifacts reproducibly with prefixes and fallbacks.​  
Sorting checkpoints by epoch number ensured the tables and graphs reflected a coherent training timeline.​

student 4:  
Building the LaSOT pair dataset deepened understanding of template–search construction and the importance of robust ground-truth parsing across varied directory layouts.​  
Handling missing or misaligned annotations safely prevented metric inflation and misleading conclusions.​

student 5:  
Designing flexible bbox extraction logic showed how real models return outputs in different shapes and keys, and why resilient post-processing is essential.​  
Graceful handling of absent predictions avoided crashes and kept the evaluation loop informative.​

student 6:  
Generating a Word report with actual tables and embedded graphs highlighted the value of “push-button” reproducibility for grading and team reviews.​  
Including a repository link and a clear reflections section made the document actionable and complete.​

student 7:  
Structured logging to both console and file gave visibility into per-checkpoint performance and helped trace issues without rerunning full evaluations.​  
Concise logs of IoU, Precision@20px, AUC, and speed metrics created an auditable trail for comparisons.​

student 8:  
Device selection, transforms, and normalization emphasized that consistent preprocessing is as critical as the model itself for stable metrics.​  
Batching and DataLoader choices balanced throughput and correctness given the template–search pairing.​

student 9:  
Merging phase results by epoch and exporting combined CSVs enabled phase-to-phase consistency checks and quick downstream analyses.​  
This alignment step prevented accidental apples-to-oranges comparisons between different checkpoint sets.​

student 10:  
End-to-end automation—from HF checkpoint retrieval to plots and a finalized report—showed how evaluation engineering accelerates iteration and collaboration.​  
The process also surfaced next steps like refining output parsing and validating trends across additional classes beyond the current tables.