

# **Impact of Video Compression on Object Tracking Performance**

by

**Takehiro Tanaka**

Thesis Submitted in Partial Fulfillment of the  
Requirements for the Degree of  
Bachelor of Applied Science (Honours)

in the  
School of Engineering Science  
Faculty of Applied Sciences

**© Takehiro Tanaka 2021  
SIMON FRASER UNIVERSITY  
Summer 2021**

Copyright in this work rests with the author. Please ensure that any reproduction or re-use  
is done in accordance with the relevant national copyright legislation.

# APPROVAL

**Name:** Takehiro Tanaka

**Degree:** Bachelor of Applied Science (Honours)

**Title of Thesis:** Impact of Video Compression on Object Tracking Performance



---

Dr. Glenn Chapman, P.Eng  
Director, School of Engineering Science

**Examining Committee:**



---

Dr. Ivan V. Bajić, P.Eng (Supervisor and Chair)  
Professor, School of Engineering Science



---

Dr. Jie Liang, P.Eng  
Professor, School of Engineering Science



---

Dr. Mirza Faisal Beg, P.Eng  
Professor, School of Engineering Science

**Date Approved:** Aug. 11, 2021

---

## Abstract

Video compression is ubiquitous in virtually all visual processing pipelines nowadays. However, its impact on many visual analytics tasks is not fully understood. In this thesis, we study the impact of video compression on multiple object tracking (MOT). MOT problem is essentially the assignment of unique identifiers to the multiple objects. In our case, the multiple objects are initialized through object detection. To achieve multiple object tracking, we employed detection-based tracking, which requires a detector to detect the objects before the identifier (ID) assignments. We used You Only Look Once v3 (YOLOv3) for object detection and Simple Online Realtime Tracking (SORT) to perform identities association to the detected objects. We collected the data by applying High Efficiency Video Coding (HEVC) to uncompressed video sequences, varying quantization parameter (QP) and motion search range (MSR). We then ran the object tracker composed of YOLOv3 and SORT on decoded sequences. We focused our study on Multiple Object Tracking Accuracy (MOTA) and found that QP has significant impact on the MOTA score, but MSR does not. We also proposed the relationship between MOTA and QP. Analyzing the results from compressed and uncompressed video sequences, we understood that most performance metrics decrease as QP increases and the image quality drops. Increase of MOTA and Precision were observed in some video sequences because in certain cases, YOLOv3 detects wrong objects. We also found that occlusion makes a difference in ID switch (IDs). IDs decreases in the video with frequent occlusions as QP increases but increases in the video with fewer occlusions. Finally, the impact of detection on tracking performance was studied. MOTA and Mean Average Precision at Intersection over Union (IOU) threshold of 0.5 (mAP-50) are positively correlated, and we observed that the relationship between MOTA and mAP-50 is linear in some video sequences. However, we also observed non-linear cases such that the MOTA growth rate decreases as mAP-50 increases in some video sequences.

# Acknowledgments

I would like to thank my academic supervisor Prof. Ivan V. Bajić for the guidance in my thesis. I appreciate all his supervision and supports to progress my thesis. I would like to thank Hyomin Choi for his mentoring from ground truth preparation to the use of video compression in my thesis work. I would like to express my gratitude to Alon Harell for guidance in statistical analysis in object tracking. I would also like to thank Saeed Ranjbar Alvar, Tim Woinoski, and Mateen Ulhaq for their assistance in object detection and tracking. I also appreciate the support from Dr. Robert A. Cohen for the issues in using computing machines for my thesis work. Finally, I would like to thank Prof. Jie Liang and Prof. Mirza Faisal Beg, for their valuable feedback and engagement as committee members.

# Contents

<b>Approval</b>	<b>ii</b>
<b>Abstract</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>iv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Object Detection and Tracking . . . . .	1
1.2 Video Compression . . . . .	3
1.3 Thesis preview . . . . .	4
<b>2 Background</b>	<b>5</b>
2.1 YOLOv3 Object Detector . . . . .	5
2.2 Simple Online Realtime Tracking (SORT) . . . . .	6
2.3 High Efficiency Video Coding (HEVC) . . . . .	10
2.3.1 Quantization Parameter . . . . .	10
2.3.2 Motion Search Range . . . . .	10
2.4 Multiple Object Tracking Metrics . . . . .	11
2.4.1 Detection Performance Measure . . . . .	11
2.4.2 ID Measure . . . . .	12
2.4.3 Track Quality Measure . . . . .	13
2.4.4 CLEAR MOT Metrics . . . . .	14
2.5 Hypothesis . . . . .	14
2.6 Summary . . . . .	15
<b>3 Data and Methods</b>	<b>16</b>
3.1 Description of Datasets . . . . .	16
3.2 Optimizing Object Tracking . . . . .	19
3.3 Experiment Pipeline . . . . .	21
3.4 Summary . . . . .	23
<b>4 Results and Discussion</b>	<b>25</b>
4.1 Results on Multiple Video Sequences . . . . .	25

4.1.1	Visualization of Results across All Video Sequences . . . . .	25
4.1.2	Regression Analysis . . . . .	26
4.1.3	One-sided t-test . . . . .	33
4.2	Case Studies of Individual Sequences . . . . .	35
4.2.1	BasketballPass . . . . .	36
4.2.2	Johnny . . . . .	38
4.2.3	BlowingBubbles . . . . .	41
4.2.4	Cactus . . . . .	41
4.2.5	Discussion . . . . .	43
4.3	Impact of Detection on Tracking Performance . . . . .	45
4.4	Summary . . . . .	46
<b>5</b>	<b>Conclusion</b>	<b>48</b>
<b>References</b>		<b>50</b>
<b>A</b>	<b>Additional Information</b>	<b>55</b>
A.1	Justification of SORT Optimization . . . . .	55
A.2	Standard Deviation of Average Result . . . . .	56
A.3	Regression Analysis of All Metrics . . . . .	57
A.4	One-sided t-test for “person” object class . . . . .	58
A.5	Correlation of Performance Metrics . . . . .	59
<b>B</b>	<b>All Results of Individual Sequences</b>	<b>61</b>
B.1	BasketballDrive . . . . .	62
B.2	Cactus . . . . .	65
B.3	Kimono . . . . .	68
B.4	ParkScene . . . . .	71
B.5	BasketballDrill . . . . .	74
B.6	RaceHorsesC . . . . .	77
B.7	BasketballPass . . . . .	80
B.8	BlowingBubbles . . . . .	83
B.9	RaceHorsesD . . . . .	86
B.10	FourPeople . . . . .	89
B.11	Johnny . . . . .	92
B.12	KristenAndSara . . . . .	95

# List of Figures

Figure 1.1	The typical video compression architecture . . . . .	3
Figure 3.1	Semi-automated annotation process for ID assignment in the ground truth	17
Figure 3.2	Object tracking pipeline with YOLO v3 and SORT . . . . .	18
Figure 3.3	Optimizing YOLOv3 by maximizing F1 score on “all” object classes . . .	20
Figure 3.4	Optimizing SORT by maximizing MOTA score on “person” and “all” object classes . . . . .	22
Figure 3.5	Pipeline for the experiment . . . . .	24
Figure 4.1	Averaged result of MOTA score from all video samples at different QP . .	26
Figure 4.2	Visualization of the average performance results at different QP (x-axis) of all video samples . . . . .	27
Figure 4.3	Visualization of the averaged performance results at different MSR (x-axis) of all video samples . . . . .	28
Figure 4.4	Scatter plots before and after the QP transformation on the BasketballPass sequence . . . . .	29
Figure 4.5	Scatter plots before and after the MOTA transformation on all the video sequences . . . . .	30
Figure 4.6	Normal probability plot to test the adequacy of the regression model . . .	32
Figure 4.7	Scatter plot of average values of MOTA measurements and fitted model .	33
Figure 4.8	Scatter plots before and after the MOTA transformation for t-test . . . .	34
Figure 4.9	Visualization of the performance results on BasketballPass at different QP for the “person” object class . . . . .	36
Figure 4.10	Comparison of ground truth and tracking results on the BasketballPass sequence in frames 320 at different QP . . . . .	37
Figure 4.11	Comparison of BasketballPass frames 81, 95, 103 to explain IDs results .	38
Figure 4.12	Visualization of the performance results on Johnny at different QP for the “person” object class . . . . .	39
Figure 4.13	Comparison of ground truth and tracking results on the Johnny sequence in frame 400 at different QP . . . . .	40
Figure 4.14	Comparison of Johnny frames 282, 286, 290 to explain IDs results . . . .	40

Figure 4.15	Visualization of the performance results on BlowingBubbles at different QP for the “person” object class . . . . .	42
Figure 4.16	Comparison of ground truth and tracking results on the BlowingBubbles sequence in frame 465 at different QP . . . . .	43
Figure 4.17	Visualization of the performance results on Cactus at different QP for the “potted plant” object class . . . . .	44
Figure 4.18	Comparison of ground truth and tracking results on the Cactus sequence in frame 32 at different QP . . . . .	45
Figure 4.19	Scatter plot of MOTA vs. mAP-50 across all video sequences . . . . .	46
Figure 4.20	Linear and non-linear cases of scatter plots of MOTA and mAP-50 . . . . .	47
Figure A.1	Pearson correlation heatmap of performance metrics for “all” object classes	60
Figure B.1	Visualization of performance results on BasketballDrive at different QP . .	62
Figure B.2	Visualization of performance results on BasketballDrive at different MSR . .	63
Figure B.3	Visualization of performance results on Cactus at different QP . . . . .	65
Figure B.4	Visualization of performance results on Cactus at different MSR . . . . .	66
Figure B.5	Visualization of performance results on Kimono at different QP . . . . .	68
Figure B.6	Visualization of performance results on Kimono at different MSR . . . . .	69
Figure B.7	Visualization of performance results on ParkScene at different QP . . . . .	71
Figure B.8	Visualization of performance results on ParkScene at different MSR . . . .	72
Figure B.9	Visualization of performance results on BasketballDrill at different QP . .	74
Figure B.10	Visualization of performance results on BasketballDrill at different MSR . .	75
Figure B.11	Visualization of performance results on RaceHorsesC at different QP . . .	77
Figure B.12	Visualization of performance results on RaceHorsesC at different MSR . .	78
Figure B.13	Visualization of performance results on BasketballPass at different QP . .	80
Figure B.14	Visualization of performance results on BasketballPass at different MSR . .	81
Figure B.15	Visualization of performance results on BlowingBubbles at different QP . .	83
Figure B.16	Visualization of performance results on BlowingBubbles at different MSR . .	84
Figure B.17	Visualization of performance results on RaceHorsesD at different QP . . . .	86
Figure B.18	Visualization of performance results on RaceHorsesD at different MSR . .	87
Figure B.19	Visualization of performance results on FourPeople at different QP . . . . .	89
Figure B.20	Visualization of performance results on FourPeople at different MSR . . . .	90
Figure B.21	Visualization of performance results on Johnny at different QP . . . . .	92
Figure B.22	Visualization of performance results on Johnny at different MSR . . . . .	93
Figure B.23	Visualization of performance results on KristenAndSara at different QP . .	95
Figure B.24	Visualization of performance results on KristenAndSara at different MSR . .	96

# List of Tables

Table 2.1	Comparison of YOLOv3 software with the original implementation . . . . .	6
Table 3.1	List of video sequences . . . . .	17
Table 3.2	List of object class IDs adapted from [44]. . . . .	18
Table 3.3	Selected optimized parameters for YOLOv3 and SORT . . . . .	21
Table 3.4	Selected QP and MSR range for the experiment . . . . .	23
Table 4.1	Average performance results across all video samples . . . . .	29
Table 4.2	Regression analysis result of the MOTA score for “all” object classes across all video sequences . . . . .	31
Table 4.3	One-sided t-test for “all” object classes across all the video sequences . . . . .	35
Table 4.4	Performance results on BasketballPass . . . . .	37
Table 4.5	Performance results on Johnny . . . . .	38
Table 4.6	Performance results on BlowingBubbles . . . . .	41
Table 4.7	Performance results on Cactus . . . . .	42
Table A.1	Justification of SORT Optimization on “person” and “all” object classes. . . . .	55
Table A.2	Standard deviation of performance results of all video samples . . . . .	56
Table A.3	Regression analysis result of all metrics for “all” object classes across all video sequences . . . . .	57
Table A.4	One-sided t-test for “person” object class across all the video sequences . . . . .	58
Table B.1	Performance results on BasketballDrive . . . . .	64
Table B.2	Performance results on Cactus . . . . .	67
Table B.3	Performance results on Kimono . . . . .	70
Table B.4	Performance results on ParkScene . . . . .	73
Table B.5	Performance results on BasketballDrill . . . . .	76
Table B.6	Performance results on RaceHorsesC . . . . .	79
Table B.7	Performance results on BasketballPass . . . . .	82
Table B.8	Performance results on BlowingBubbles . . . . .	85
Table B.9	Performance results on RaceHorsesD . . . . .	88
Table B.10	Performance results on FourPeople . . . . .	91
Table B.11	Performance results on Johnny . . . . .	94



# Chapter 1

## Introduction

Object detection is a computer vision task to recognize the object instances from digital images. Specifically, the goal of object detection is to detect the different object classes such as “person”, “cup”, and “clock”. Object detection is a building block for various applications such as “smart vehicle”, that informs the driver if one is running in the right direction at a proper speed, or a system to detect pedestrians to prevent accidents [1]. Object detection can further be developed to object tracking, which is another computer vision task. Object detection is able to detect the objects for different classes, but this task alone has no unique identification of objects within the same class in a sequence of frames. For example, the system can detect multiple persons, but there is no identification of each person across the frames, and hence object tracking is supposed to assign each unique identifier (ID) to each object instance.

Video compression is ubiquitous in most visual processing pipelines. Compression of video is necessary because it would be impractical to transmit it or save it in a storage device without compression. For example,  $1920 \times 1080$  pixels of full-color 90 min video with 30 frames per second (FPS) takes 1007.8 GB uncompressed, as calculated below:

$$1920 \times 1080 \text{ px} \cdot 3 \frac{\text{B}}{\text{px}} \cdot 30 \frac{\text{frame}}{\text{sec}} \cdot 60 \frac{\text{sec}}{\text{min}} \cdot 90 \text{ min} \cdot 10^{-9} \frac{\text{GB}}{\text{B}} = 1007.8 \text{ GB}, \quad (1.1)$$

where px is pixel and B is byte. However, the size 1007.8 GB is too large to transmit. Since video compression is ubiquitous, any video we access is pre-compressed, and this raises the question of how much video compression impacts the object tracking performance. To our knowledge, the effect of video compression on object tracking has not been studied in detail. Hence, this project aims to analyze the effect of video compression on object tracking performance by assessing the various metrics used for multiple object tracking.

### 1.1 Object Detection and Tracking

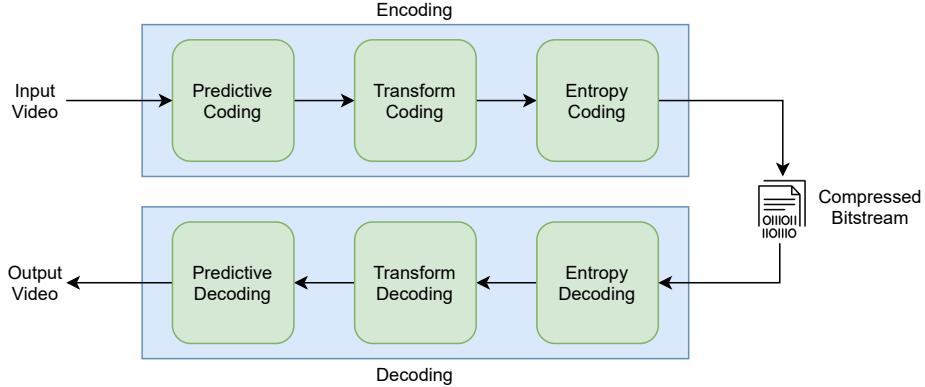
Object tracking problem can be categorized into Single Object Tracking (SOT) and Multiple Object Tracking (MOT). SOT tracks a single object over the sequence of digital images, while MOT tracks

multiple objects. The following introduction to MOT has been explained well by Luo *et al.* [2]. SOT tracks a single object with a focus on appearance and motion models. MOT has common challenges with the single object case, but tracking multiple objects involves additional challenges to be solved. In the multiple objects case, the tracking system needs to determine the number of objects that could vary over the frames and preserve their identities. In this thesis work, we do not want to restrict tracking to the single object case, but aim to analyze the effect of compression on tracking multiple objects, which is more general.

Luo *et al.* [2] also explained the general categories of MOT. MOT initialization methods can be grouped into Detection-Based Tracking and Detection-Free Tracking. Detection-Based Tracking involves an object detector to perform detection as initialization to the tracker for every image in a sequence. The object tracker will then assign a unique identity to each detected object, in other words, performing data association of unique ID based on detected objects. Detection-Free Tracking does not involve an object detector but requires other initialization methods such as inference that exploits rich appearance and motion information [3]. MOT processing methods can be group into online tracking and offline tracking. Online tracking performs data association of unique identities based on current and past observations, while offline tracking does so by future, current, and past observations as a batch. The type of output from MOT can also be stochastic or deterministic. The output of stochastic tracking varies every time the tracker is run while for the deterministic case, the output does not change over the tracker runs.

For this thesis, we are using Detection-Based Tracking, and hence we will need an object detector as part of tracking. Object detection has been actively researched for the past 20 years, and there are two stages the research has gone through according to Zou *et al.* [4]. The first stage is traditional detection, which involves hand-crafted methods while the second stage involves deep learning-based methods. Deep learning-based methods can be further broken down into two categories: one-stage detection and two-stage detection. For two-stage detectors, the first stage generates the region of interest as a bounding box detection proposal and each region proposal is fed into a convolutional neural network (CNN). Classification and regression is performed on each proposal in the second stage. Various two-stage deep learning-based methods have been developed such as Region Based Convolutional Neural Network (R-CNN) [5], Spatial Pyramid Pooling network (SPP-net) [6], Fast R-CNN [7], Faster R-CNN [8], Feature Pyramid Network (FPN) [9], and Mask R-CNN [10]. As for one-stage detectors, the full image is fed into the single neural network and each bounding box as a region of interest is predicted by the network. The examples for the one-stage detector could be You Only Look Once (YOLO) [11], Single Shot Multibox Detector (SSD) [12], and RetinaNet [13].

Luo *et al.* [2] listed the examples of components used in multiple object tracker: appearance model that computes the affinity between the observations based on the visual cues and statistical measures, motion model that captures dynamic behavior of objects such as linear and non-linear motion, interaction model that considers individual interaction in the social environment, exclusion model that utilizes constraint between objects or trajectories, occlusion handling, and finally inference



**Figure 1.1:** The typical video compression architecture adapted from [17].

approach that utilizes probability distribution of object states or approach based on the deterministic optimization.

Out of all aforementioned possible ways to detect and track objects, as listed above, we employed the YOLOv3 object detector [14] with Simple Online Realtime Tracking (SORT) [15]. This solution is detection-based and online, with deterministic output, and the details are explained in Chapter 2.

## 1.2 Video Compression

The early form of video compression was described by Ray Davis Kell in 1929, as the difficulty of transmitting successive images of video can be avoided by only sending the difference between the successive images, though it was not actually used; however, it became the foundation for the video compression standards today [16]. Early video compression was analog, but digital video processing has been developed and is widely used today. Zhang *et al.* [17] explained the concept of typical video compression nowadays as following. The video compression consists of the encoder compressing the images into the compressed form, which can be stored or transmitted to another location, and the decoder to decompress the images. This process of coding and decoding is also called a codec. Typical video compression standards nowadays comprise predictive coding, transform coding, and entropy coding, as shown in Figure 1.1. Predictive coding is the component that reduces the inter-frame temporal redundancy and intra-frame spatial redundancy in a video by motion estimation (ME), motion compensation (MC), and spatial prediction techniques. Transform coding is the component where the quantized transform coefficients are generated through discrete cosine transform (DCT) to help reduce the spatial redundancy. Entropy coding is the component where compressed bitstreams are generated.

There are two types of video coding: lossless coding and lossy coding. Lossless coding compresses the images and obtains the reconstructed images after decompression without any loss of information. Lossy coding, however, compresses the images by removing the less important information, which will sacrifice the image quality to the level the human visual system can tolerate.

Lossy compression is more widely used today since it allows a much smaller compressed size and is more efficient than the lossless compression.

Since the first video compression standard H.120 developed in 1984, various video compression standards have been developed, such as MPEG and H.26X series [17]. They all followed the architecture shown in Figure 1.1. The organization of Moving Picture Experts Group (MPEG) in the International Standards Organization (ISO) and the International Electrotechnical Commission (IEC) are developing the MPEG series such as MPEG-1, MPEG-2, and MPEG-4.

The Video Coding Expert Group (VCEG) of the International Telecommunication Union (ITU-T) is developing H.26X series starting from the first standard H.120. They subsequently developed H.261, H.262, H.263, H.264 (AVC), and H.265 (HEVC). H.264, so-called Advanced Video Coding (AVC), was released in 2003. H.264/AVC is the most widely used standard nowadays and supports up to 4k resolution of video. H.265, so-called High Efficiency Video Coding (HEVC), was developed based on H.264/AVC structure and is a more recent standard that has been released in 2013. H.265/HEVC supports up to 8k resolution of the video but is not yet widely supported. There has been more recent development of standards such as, for example, Versatile Video Coding (VVC) and AOMedia Video 1 (AV1). These newly developed standards, after the predecessor of H.265/HEVC, achieve better coding performance and will allow high quality and efficiency, virtual reality (VR) system, and 360-degree video applications. However, these standards and H.265/HEVC are still yet to be supported due to the current hardware's lack of computational power. As a possible future generation of video compression standards after VVC and AV1, there has been research in compression technologies utilizing machine learning for video coding, hardware acceleration, and parallel computing. Out of these video compression standards, we have adopted H.265/HEVC for the experiments in this thesis.

### 1.3 Thesis preview

We started by illustrating the motivation for this research and providing history and background information for each object tracking and video compression concept in Chapter 1. We will explain the more detailed background information on the adopted methods in Chapter 2. Chapter 3 shows the methodologies and experimental procedures. Chapter 4 presents the results from the experiments and data analysis. Finally, we will conclude this thesis with the highlighted insights and future work in Chapter 5. The additional information is included in the Appendix.

# Chapter 2

## Background

In order to assess the multiple object tracking performance on compressed sequences, we have to build an object tracking system. Our objective in this thesis is not to build the best possible state-of-the-art object tracking system but to analyze the effect of video compression on tracking performance. Therefore, we have chosen a simple and effective object tracker design. YOLOv3 [14] was chosen as an object detector and SORT [15] was chosen as a tracking mechanism to perform data associations of object identities. Our object tracking system is in the category of detection-based online tracking with deterministic output. Due to the simplicity in design, this tracker does not solve the problems of long-term occlusions and object re-identifications.

In this chapter, we will be reviewing the background of YOLOv3, SORT, HEVC video compression standard, and MOT metrics for performance evaluation.

### 2.1 YOLOv3 Object Detector

You Only Look Once (YOLO) [11] was taken into consideration for the choice of our object detection component in the object tracking system. The following conceptual background of YOLO was described by Redmon *et al.* [11]. YOLO is a convolutional neural network-based model, and it divides the input image into  $S \times S$  grids. Each cell in the grid predicts the bounding box information and the object class probability. YOLO is much faster in speed and more suitable for real-time applications but gives somewhat lower accuracy than other state-of-the-art object detectors. YOLOv2 [18] has been developed as an improvement to its predecessor YOLO. YOLOv2 primarily focuses on improving recall and localization while preserving classification accuracy and speed. Anchor boxes have been implemented to improve recall, and YOLOv2 with anchor boxes can predict a lot more bounding boxes than YOLO. YOLOv2 estimates a bounding box by an offset from a candidate anchor box instead of estimating the coordinates directly as in the original YOLO. As an incremental improvement to YOLOv2, YOLOv3 [14] has been developed. The highlighted improvements are as follows; the logistic regression has been implemented to predict the objectness score. If the bounding box overlaps the ground truth more than any other anchor boxes, the objectness score will be 1. YOLOv3 uses an independent logistic classifier instead of softmax implemented

**Table 2.1:** Comparison of YOLOv3 software with the original implementation.

-	Resolution	mAP	mAP-50
YOLOv3, Darknet53	608x608	33.0	57.9
YOLOv3, Ultralytics	640x640	43.3	63.0

in YOLOv2 to allow multi-label class predictions. Finally, YOLOv3 predicts bounding boxes and the corresponding object classes at 3 different scales across the network, similar to Feature Pyramid Network [9]. With these improvements, YOLOv3 achieves localization of smaller objects while maintaining speed and accuracy.

We have chosen the YOLOv3 object detector developed by Ultralytics [19] with the provided pre-trained weights. The developer Jocher *et al.* pre-trained the network with the backbone of Darknet53, and the model has 261 layers with 61,922,845 parameters. The COCO dataset was used for training with 118,287 training images, 5,000 validation images, and 20,288 out of 40,670 test images. This network with the pre-trained weights can detect up to 80 object classes available from the COCO dataset [20]. To compare the chosen YOLOv3 detector from Ultralytics with the original implementation of YOLOv3 by Redmon *et al.*, the comparison of Mean Average Precision (mAP) on the COCO dataset has been shown in Table 2.1. The mAP value from the original implementation of Darknet53 is reported by the author Redmon *et al.*, and the value from the Ultralytics implementation is reported by Jocher *et al.* [19]. mAP is a detection performance metric, and it averages Average Precision (AP) values from different object classes. In the COCO Detection Challenge, mAP also averages AP values at 10 intersection over union (IOU) thresholds from 0.5 to 0.95 [20] [21]. mAP-50 is a detection performance metric at IOU threshold of 0.5 in particular, and its detail definition is explained in Section 2.4. The mAP and mAP-50 are reported based on Microsoft COCO dataset. As Table 2.1 shows, our chosen detector from Ultralytics achieves better detection performance than the original implementation of Darknet53.

## 2.2 Simple Online Realtime Tracking (SORT)

The detection-based tracking requires an object detector to initialize the tracker’s state by detecting the objects, then a tracking component will track those objects. Multiple object tracking problem can be viewed as a data association problem in assigning the optimal unique identity to each object. Simple Online Realtime Tracking (SORT) was selected for the tracking component, which is a detection-based online tracker with a deterministic output. As Bewley *et al.* describe [15], based on the detected objects, SORT utilizes the constant linear motion model to estimate the displacements of the objects across the frames. The tracker’s state at time  $k$  can be represented as a state vector:

$$\mathbf{x}_k = [u, v, s, r, \dot{u}, \dot{v}, \dot{s}]^T, \quad (2.1)$$

where  $u$  and  $v$  are the pixel coordinates in the horizontal and vertical direction;  $s$  and  $r$  are the area and aspect ratio of the bounding box;  $\dot{u}$  and  $\dot{v}$  represent the velocity;  $\dot{s}$  is the rate of change of the area. The aspect ratio  $r$  is assumed to be constant, and hence the state vector does not include  $\dot{r}$ . SORT utilizes the Kalman filter [22] (the original publication by Kalman *et al.*) to estimate the state vector optimally. The use of Kalman filter in the multiple object tracking is explained well by [23] and adapted from this explanation, we described the framework as following. Also, we adapted the following equations according to the source code of SORT [24] and the library source code of the Kalman filter [25].

Kalman filter framework can be implemented recursively to estimate the object's state in the next frame based on the previous state as well as its error. Recall that we utilize YOLOv3 for object detection, so the measurements, detected bounding boxes in our case, are taken at every frame. The measurement vector can be represented as

$$\mathbf{z}_k = [u, v, s, r]^T \quad (2.2)$$

The Kalman filter framework allows us to estimate the apriori state vector  $\hat{\mathbf{x}}^-$  and apriori covariance matrix  $\mathbf{P}^-$  at time  $k$  based on the previous state and covariance at time  $k - 1$ .

$$\hat{\mathbf{x}}_k^- = \mathbf{A}\hat{\mathbf{x}}_{k-1}, \quad (2.3)$$

$$\mathbf{P}_k^- = \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^T + \mathbf{Q} \quad (2.4)$$

$\mathbf{A}$  is a transition matrix represented as

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.5)$$

The measurement vector at time  $k$  can be obtained from the state vector  $\mathbf{x}_k$  as

$$\mathbf{z}_k = \mathbf{H}\mathbf{x}_k, \quad (2.6)$$

where  $\mathbf{H}$  is a measurement matrix represented as

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \quad (2.7)$$

Given the apriori<sup>1</sup> estimates from Equation (2.3), (2.4) and the estimated measurement  $\mathbf{H}\hat{\mathbf{x}}_k^-$ , Kalman filter framework corrects the apriori estimates, updating them to have aposteriori<sup>2</sup> estimates using the current measurement  $\mathbf{z}_k$ . The equations for the aposteriori estimates will be

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}^T (\mathbf{H} \mathbf{P}_k^- \mathbf{H}^T + \mathbf{R})^{-1}, \quad (2.8)$$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k (\mathbf{z}_k - \mathbf{H}\hat{\mathbf{x}}_k^-), \quad (2.9)$$

$$\mathbf{P}_k = (1 - \mathbf{K}_k \mathbf{H}) \hat{\mathbf{P}}_k^-, \quad (2.10)$$

where  $\mathbf{K}_k$  is the Kalman gain at time  $k$ . The Kalman gain  $\mathbf{K}_k$ , the aposteriori estimate of state vector  $\mathbf{x}_k$ , and the aposteriori estimate of error covariance matrix  $\mathbf{P}_k$  are recursively updated by the estimated apriori state vector  $\hat{\mathbf{x}}_k^-$  and apriori error covariance matrix  $\mathbf{P}_k^-$ . Therefore, the Kalman filter framework essentially consists of the prediction steps that will estimate the bounding boxes as apriori estimates and the correction steps that will correct the predicted information as aposteriori estimates, and the process will be recursively back to the prediction steps to proceed with the next frame. For the initial conditions, when the object is detected, the velocity components are set to 0 for the state vector  $\mathbf{x}_0$ .

$$\mathbf{x}_0 = [u, v, s, r, 0, 0, 0]^T \quad (2.11)$$

<sup>1</sup>Apriori estimates are the estimated states and covariance before the correction.

<sup>2</sup>Aposteriori estimates are the estimated states and covariance after the correction.

For the covariance matrix  $\mathbf{P}_0$ , the velocity components are set to large variance values as  $10^4$  since no velocity is observed at the initial state.

$$\mathbf{P}_0 = \begin{bmatrix} 10 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 10 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 10^4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 10^4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 10^4 \end{bmatrix} \quad (2.12)$$

For the process noise covariance  $\mathbf{Q}$  and the measurement noise covariance  $\mathbf{R}$ , constant values are assigned.

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 10^{-2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 10^{-2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 10^{-4} \end{bmatrix} \quad (2.13)$$

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 10 \end{bmatrix} \quad (2.14)$$

These initial conditions and noise matrices used the default values by [15].

As for the data association, the assignment cost matrix, consisting of IOU values between each measured bounding box and predicted box by the Kalman filter, can be computed. The assignments of unique identities on a current frame can be solved optimally from this matrix via the Hungarian algorithm [26]. When the measured object overlaps less than the minimum IOU threshold value with the predicted bounding box, the assignment for the measured bounding box will be rejected. For the trajectories initialization, however, the untracked objects are found when the detected objects overlap less than the minimum IOU threshold with the predicted bounding boxes. Followed by the minimum number of detections, the trajectories with the unique identities are initialized. When the objects are not detected for a certain number of frames labeled as  $T_{\text{Lost}}$ , the trajectories are terminated.

SORT is focused on simplicity of design, which could serve as a baseline tracking method. This is because SORT essentially consists of a Kalman filter framework for predicting motion and a Hungarian algorithm for data association. However, it does not deal with long-term occlusions. Adding the object re-identification feature to deal with occlusions will weigh significant complexities on the tracker. SORT is simple yet fast, and maintains high accuracy with deterministic output.

Combined with a fast YOLOv3 object detector, SORT will make it easy for us to run the experiments since both components are fast enough and produce deterministic results.

## 2.3 High Efficiency Video Coding (HEVC)

H.265/HEVC is the video compression standard developed by the ITU-T Video Coding Expert Group in 2013. H.265/HEVC follows the same structure of its predecessor H.264/AVC, as shown in Figure 1.1 but achieves a better coding performance [17]. Zhang *et al.* [17] explains that the typical structure consists of predictive coding with intra-frame and inter-frame prediction, transform coding that generates the quantized transform coefficients from DCT, and entropy coding that generates a compressed bitstream. To obtain a de-compressed video, HEVC performs entropy decoding, transform decoding, and predictive decoding, in a reversed order to the encoding part. For the experiments, we have used HEVC test model (HM) version 16.20, and varied two compression parameters: quantization parameter (QP) and motion search range (MSR).

### 2.3.1 Quantization Parameter

Quantization parameter (QP) is a parameter used in transform coding. The value of QP determines the quantization step size by which we obtain the quantized transform matrix. QP ranges from 0 to 51, and an increase of 6 in QP will double the quantization step size [27] [28]. According to [29], QP has a significant impact on the bitrate. They showed that bitrate is inversely proportional to QP and is linearly proportional to the pixel rate. Note that they refer pixel rate as “frame rate multiplied by number of pixels in the frame” [29, p. 7]. This means that high QP will result in lower bitrate and hence the lower amplitude resolution. In other words, a high QP will cause amplitude resolution to be lower, while a low QP will sustain a high amplitude resolution.

### 2.3.2 Motion Search Range

Motion estimation is an inter-frame prediction technique that finds the best match for a block of pixels between the reference frame and the current frame while minimizing the rate-distortion cost or maximizing correlation. Motion estimation reduces temporal redundancy by obtaining a motion vector that points from the target candidate region in the current frame toward the region in the previous reference frame. The region where the block matching is performed is the search window, and its size is called the search range (SR). A high SR value means a larger search window and hence requires more memory and computation, but a low SR means a smaller search window and requires less memory [30] [31]. From this logic, we can interpret that large SR could cover fast motion in a video, while the low SR only covers the slower motion. In the experiment, we call this parameter motion search range (MSR), and we examine its impact on tracking accuracy.

## 2.4 Multiple Object Tracking Metrics

To evaluate the object detection and tracking performance quantitatively, we need metrics to quantify the performance with respect to the ground truth. Since there is no single metric that captures all aspects of the tracking performance, various metrics have been considered as listed below. The arrow symbol  $\uparrow$  indicates “the higher, the better”, while  $\downarrow$  indicates “the lower, the better”. We utilized the software from [32] to evaluate the performance metrics.

### 2.4.1 Detection Performance Measure

The following metrics of TP, FP, FN, Precision, Recall, F1, and mAP-50 measure the detection performance. TP, FP, and FN are based on the value of Intersection over Union (IOU) threshold, which is defined as the area of intersection of the detected bounding box and the ground truth bounding box, divided by the union of those boxes. For example, if we set the IOU threshold to 0.5 and obtain the IOU value of 0.8 at the target object, we count it as TP. If we obtain the IOU value of 0.2, we then count it as FP. If there is no detected bounding at the ground truth target, FN will be counted. The following definitions of detection performance measure are explained in [33] [34]. For mAP-50, we explained the definition adapted from [35] and [36].

- **TP ( $\uparrow$ )**: True Positive. The number of times the detector correctly detects a target.
- **FP ( $\downarrow$ )**: False Positive. The number of times the detector falsely detects a target.
- **FN ( $\downarrow$ )**: False Negative. This metric is the opposite of FP, i.e. the number of times the detector fails to detect a target.
- **Precision ( $\uparrow$ )**: The number of correct detection divided by the number of all detection made by the detector.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2.15)$$

- **Recall ( $\uparrow$ )**: The number of correct detections divided by the number of objects from the ground truth.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2.16)$$

- **F1 ( $\uparrow$ )**: F1 assesses the detection performance as the harmonic mean of precision and recall.

$$F1 = 2 \frac{\text{TP} \cdot \text{FN}}{\text{TP} + \text{FN}} \quad (2.17)$$

- **mAP-50 ( $\uparrow$ )**: Mean Average Precision (mAP) is the metric that assesses the detection performance and is the most popular metric used in the benchmarks [36]. mAP is the mean of Average Precision (AP) value for each object class. We adopted the mAP metric that evaluates the detection performance at IOU threshold of 0.5. We call it as mAP-50. mAP evaluated

at IOU threshold of 0.5 is the metric based on the PASCAL VOC challenge [37]. What mAP-50 differs from the metric F1 is that it evaluates the detection performance at multiple confidence thresholds from the detector while F1 evaluates only at one specific confidence threshold. With the given confidence threshold, the detector will only detect the objects with the object class probability greater than the threshold. To quantify the detection performance that accounts for different confidence thresholds, we would like to calculate the integration of Precision and Recall over different confidence thresholds, which will be the area under the curve of  $\text{Precision}(s) \times \text{Recall}(s)$ . This area under the curve will be the AP value. Note that Precision and Recall are now subject to the different confidence threshold  $s$ . However, the curve is often zig-zag, so the interpolation of Precision at different Recall will be necessary. By doing the interpolation, AP can be obtained via Riemann sums as following equations adapted from [35] [36].

$$\text{AP} = \sum_n (R_{n+1} - R_n) P_{\text{interp}}(R_{n+1}) \quad (2.18)$$

$$P_{\text{interp}}(R_{n+1}) = \max_{R(s) \geq R_n} P(R(s)) \quad (2.19)$$

We call Precision and Recall as  $P$  and  $R$  respectively for simplicity in the equations.  $P_{\text{interp}}$  is the interpolated Precision which takes the maximum Precision at given all available Recall values greater than or equal to  $R_n$ .  $R_n$  is the previous interpolated Recall value and  $R_{n+1}$  is the current interpolated Recall value. Based on this definition, AP can be computed iteratively and  $n$  indicates  $n$ -th iteration. This interpolation method is called all-point interpolation [35]. Evaluating AP value for each object class at IOU threshold of 0.5 and averaging them with the total number of object classes, we obtain mAP-50 as follows.

$$\text{mAP-50} = \frac{1}{C} \sum_{i=1}^C (\text{AP-50})_i \quad (2.20)$$

where  $C$  is the number of object classes and  $(\text{AP-50})_i$  is the AP value at  $i$ -th class with the IOU threshold of 0.5. We utilized the software [38] to compute mAP-50 using all-point interpolation and their corresponding literature is [39].

#### 2.4.2 ID Measure

IDP, IDR, IDF1 are measures of the tracker's ability to identify object trajectories. IDTP, IDFP, and IDFN are used in the definitions of IDP, IDR, and IDF1. Ristani *et al.* [33] listed the following definitions of ID performance measures.

- **IDTP ( $\uparrow$ )**: The number of correct identifications of the trajectories.
- **IDFP ( $\downarrow$ )**: The number of incorrect identifications of the trajectories.

- **IDFN** ( $\downarrow$ ): The number of times the tracker fails to identify the true trajectories.
- **IDP** ( $\uparrow$ ): Identification Precision. Similar to the definition of Precision, the number of correct identifications is divided by the number of all identifications.

$$\text{IDP} = \frac{\text{IDTP}}{\text{IDTP} + \text{IDFP}} \quad (2.21)$$

- **IDR** ( $\uparrow$ ): Identification Recall. The number of correct identifications is divided by the number of ground truth identifications.

$$\text{IDR} = \frac{\text{IDTP}}{\text{IDTP} + \text{IDFN}} \quad (2.22)$$

- **IDF1** ( $\uparrow$ ): IDF1 assesses the identification performance as the harmonic mean of IDP and IDR.

$$\text{IDF1} = 2 \cdot \frac{\text{IDP} \cdot \text{IDR}}{\text{IDTP} + \text{IDFN}} \quad (2.23)$$

- **IDs** ( $\downarrow$ ): Identity switch counts the number of times a different identity is assigned to a matched trajectory. This is also called a mismatch. IDs is only counted when the newly assigned identity  $i$  on the trajectory matches with the ground truth trajectory with the identity  $k$ , and the previously assigned trajectory with the identity  $j$  is not the same as  $i$  [34]. This metric is also sometimes denoted as IDSW in the literature [34].

### 2.4.3 Track Quality Measure

MT, PT, ML, and FM are metrics that measure the track quality. Note that these metrics do not account for ID assignments, so ID does not have to be the same on the trajectory. The following definitions of track quality metrics are explained by [34].

- **GT**: It counts the number of ground truth trajectories.
- **MT** ( $\uparrow$ ): Mostly Tracked. It counts the number of trajectories where each trajectory is being tracked at least 80% of the time with respect to the entire time of the ground truth trajectory.
- **PT** ( $\downarrow$ ): Partially Tracked. It counts the number of trajectories where each trajectory is being tracked at least 20% of the time but less than 80% with respect to the entire time of the ground truth trajectory.
- **ML** ( $\downarrow$ ): Mostly Loss. It counts the number of trajectories where each trajectory is being tracked less than 20% of the time with respect to the entire time of the ground truth trajectory.
- **FM** ( $\downarrow$ ): Fragmentation counts the number of times the trajectories being tracked switched to untracked.

#### 2.4.4 CLEAR MOT Metrics

According to [40], classification of events, activities, and relationships (CLEAR) evaluations introduced two metrics of MOTA and MOTP in 2008. Although there is no single metric that is able to assess the all aspects of tracking performance, MOTA and MOTP emphasize the more general and overall performance of the tracker. Especially, MOTA is considered the most popular metric to assess the overall tracking performance [34] and according to [41], MOTA is the best measure that aligns with the human visual assessment of tracking accuracy.

- **MOTA ( $\uparrow$ ):** Multiple Object Tracking Accuracy. MOTA is a metric that combines three other metrics: FN, FP, and IDs. It scores the overall tracking performance in the context of MOT and the value could range in  $(-\infty, 100]$ . The score becomes negative when the overall error count exceeds the total number of ground truth trajectories.

$$\text{MOTA} = 1 - \frac{\sum_t (\text{FN}_t + \text{FP}_t + \text{IDs}_t)}{\sum_t \text{GT}_t} \quad (2.24)$$

where  $t$  is the frame index.

- **MOTP ( $\uparrow$  in %):** Multiple Object Tracking Precision. The numerator shows the total overlap of all target bounding boxes with respect to the ground truth bounding boxes for all the frames, and the denominator indicates the total number of matches for all the frames. This measure quantifies how well the objects are localized by the detector in MOT, but it does not give a good indication of the overall tracking performance.

$$\text{MOTP} = \frac{\sum_{t,i} d_{t,i}}{\sum_t c_t} \quad (2.25)$$

where  $d_{t,i}$  shows how much a target bounding box overlaps with the ground truth bounding box at target  $i$  at a frame index  $t$  while  $c_t$  denotes the matched objects with the correct ID at a frame index  $t$ . We report this score as a percentage rather than the value based on the original definition; the following conversion is made.

$$\text{MOTP}(\%) = (1 - \text{MOTP}) \cdot 100 \quad (2.26)$$

### 2.5 Hypothesis

We hypothesized that the object detection and tracking accuracy would be higher in uncompressed sequences than the compressed sequences. Applying lossy video compression to the uncompressed sequence will inevitably cause loss of information, so we expect to see the accuracy higher for the uncompressed video. Also, the higher the QP, we expect to see a decrease in tracking accuracy, since increasing QP will decrease bitrate and cause a larger loss of information. For MSR, the lower

the value, HEVC is less able to cover fast motion, so we expect to see a decrease in the tracking performance.

## 2.6 Summary

To summarize, in this chapter, we explained the background of the YOLOv3 object detector and SORT for multiple object tracking. We explained the basics of the HEVC codec and its configuration parameters to be varied for the experiment. We also defined several metrics that evaluate the object detection and tracking accuracy in the context of multiple object tracking. Finally, we stated the hypotheses for the detection and tracking accuracy on uncompressed and compressed sequences over different QP and MSR values. To test these hypotheses, we will conduct experiments to take measurements of accuracy scores for each metric on different QP and MSR values. The experiments are explained in the next chapter.

# Chapter 3

## Data and Methods

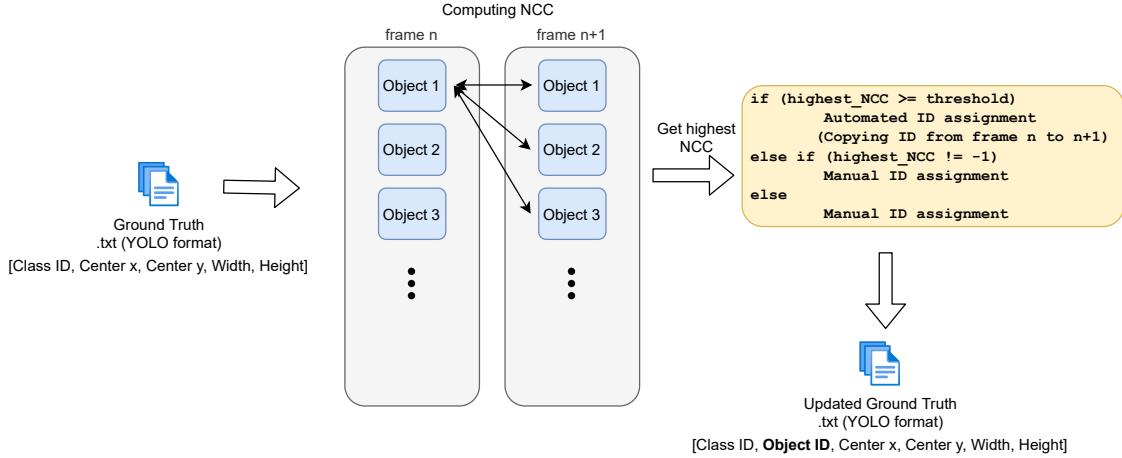
In this chapter, we will describe the dataset and experimental procedures employed to test our hypotheses. This thesis work started with preparing the ground truth data by annotating the objects in raw HEVC test sequences. We then set up the experiment by choosing suitable values of the parameters in the object tracking system. After the setup, we applied HEVC video compression to the uncompressed video sequences and ran the object tracking system with YOLOv3 and SORT to obtain the tracking results. We then compared the result with the ground truth to assess the tracking performance using MOT metrics.

### 3.1 Description of Datasets

To our knowledge, there is no public or open-source dataset of uncompressed video sequences with object tracking ground truth. Our group members prepared object detection annotations for the uncompressed HEVC v1 Common Test Conditions (CTC) sequences [42] in the YUV420 format. The sequences can be obtained from Joint Collaborative Team on Video Coding (JCT-VC). They annotated them to obtain object detection ground truth, using YOLOv3 [14] and YOLO Mark software [43] in a semi-automated annotation process [44]. The existing annotations prepared by our group members are suitable for object detection; however, for the purpose of analyzing the tracking performance, we further annotated the unique object identifier (ID) on the existing ground truths using Normalized Cross-Correlation (NCC) [45]. NCC value gives a similarity between two images. Figure 3.1 shows the semi-automated annotation procedure for assigning unique IDs.

Given the existing ground truth, each object annotation is in the YOLO format as [Class ID, Center x, Center y, Width, Height, Confidence]. Class ID indicates the identifier to the type of object class; for example, “person” as 0, “sports ball” as 32, and “chair” as 56, which are part of the 80 COCO object classes [20]. Center x and Center y are the center position of the bounding box in relative coordinates<sup>1</sup>; Width and Height are the corresponding dimensions of boxes in relative coordinates.

<sup>1</sup>A value in a relative coordinate is equivalent to a value in a pixel coordinate divided by the frame width horizontally and the frame height vertically.



**Figure 3.1:** Semi-automated annotation process for ID assignment in the ground truth.

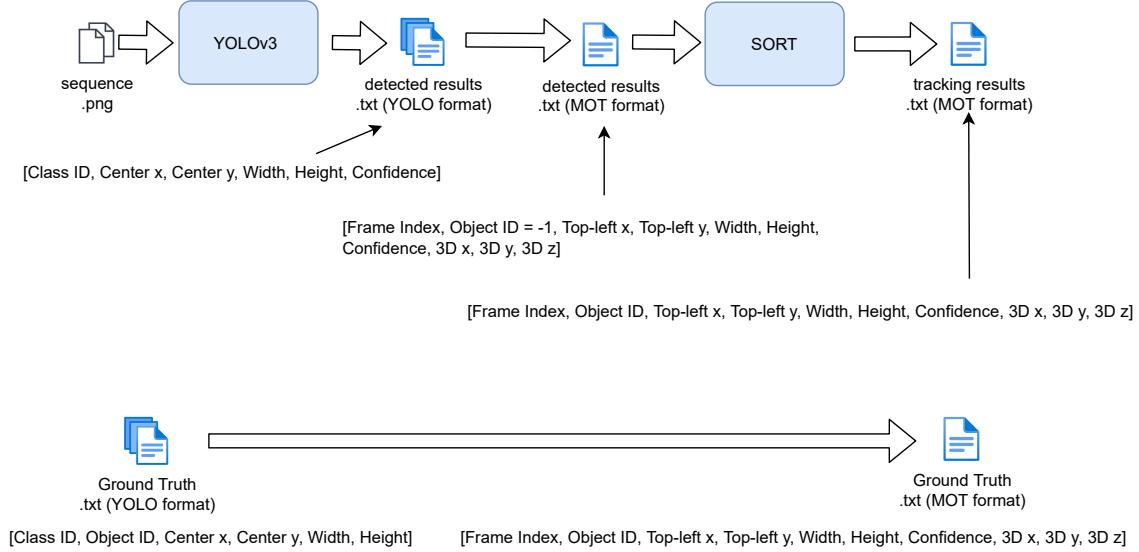
**Table 3.1:** List of video sequences adapted from [44].

Sequence Class	Sequence Name	Frame Count	Resolution	Object Class IDs	Frame rate (Hz)	Bit depth
B	BasketballDrive	500	1920x1080	[0, 32, 56]	50	8
B	Cactus	500	1920x1080	[58]	50	8
B	Kimono	240	1920x1080	[0, 26]	24	8
B	ParkScene	240	1920x1080	[0, 1, 13]	24	8
C	BasketballDrill	500	832x480	[0, 32, 56]	50	8
C	PartyScene	500	832x480	[0, 41, 58, 74, 77]	50	8
C	RaceHorsesC <sup>2</sup>	300	832x480	[0, 17]	30	8
D	BasketballPass	500	416x240	[0, 32, 56]	50	8
D	BlowingBubbles	500	416x240	[0, 41, 77]	50	8
D	RaceHorsesD <sup>2</sup>	500	416x240	[0, 17]	30	8
E	KristenAndSara	600	1280x720	[0, 63, 67]	60	8
E	Johnny	600	1280x720	[0, 27, 63]	60	8
E	FourPeople	600	1280x720	[0, 41, 56, 58]	60	8

We compare every object in a given class in the previous frame  $n$  with every object of the same class in the current frame  $n + 1$  and compute NCC values for every possible pair between those image patches containing objects. As shown in the pseudocode in Figure 3.1, we take the pair of objects that corresponds to the highest NCC value, and if the value is greater than or equal to the threshold, we copy the ID of the object from frame  $n$  to the corresponding object in a pair, which gives the highest NCC, in frame  $n + 1$  automatically. Threshold is a value that tells how similar the two image patches of objects must be, and we chose 0.60 for most video sequences. If the highest NCC is less than the threshold, ID is manually assigned. When there is no pair of objects between the previous and current frame, the highest NCC is assigned -1 to signal that the pair of objects does not exist, and ID is manually assigned. This situation could happen when objects do not exist for the particular object class in the previous frame but objects exist in the current frame. Throughout this semi-automated process of annotation, object IDs are assigned in the second column of the ground truth files.

**Table 3.2:** List of object class IDs adapted from [44].

Class ID	Object class name	Class ID	Object class name
0	person	41	cup
1	bicycle	56	chair
13	bench	58	potted plant
17	horse	63	laptop
26	handbag	67	cell phone
27	tie	74	clock
32	sports ball	77	teddy bear



**Figure 3.2:** Object tracking pipeline with YOLO v3 and SORT.

Table 3.1 shows the 13 uncompressed HEVC v1 CTC video sequences for which we created ground truth annotations, out of 18 available video sequences from [44]. The sequence class (B, C, D, E) indicates the resolution ( $\text{Width} \times \text{Height}$ ). Each sequence has different number of object classes and each class ID is from the 80 COCO object classes [20]. Table 3.2, adapted from [44], shows the corresponding object class name for each class ID, and we only listed the object classes that we detect and track in the given sequences from Table 3.1. In the experiments, “all” refers to all object classes available in the ground truth.

The object tracking pipeline consists of the YOLOv3 detector and SORT, as shown in Figure 3.2. Input video sequences of PNG files are input to the YOLOv3 object detector. The output from YOLOv3 will be generated in the YOLO format as [Class ID, Center x, Center y, Width, Height, Confidence]. Note that Confidence is the object class probability, and this score tells how likely the object is to be in a particular class. These results are converted to the MOT format used in the MOT

<sup>2</sup> The source [44] uses the same sequence name “RaceHorses” for Class C and Class D, but we call it as “RaceHorsesC” and “RaceHorsesD” respectively to distinguish.

Challenge 2015 benchmark [46]. The object ID, the unique identifier to the object, for the detected result is initialized as -1. Applying SORT to this detected result, we obtain the tracking result in the MOT format with the assigned object ID as [Frame index, Object ID, Top-left x, Top-left y, Width, Height, Confidence, 3D x, 3D y, 3D z]. Top-left x and Top-left y are the positions of the bounding box at the top-left corner in pixel coordinates. Width and Height are dimensions of boxes in pixel coordinates. 3D x, 3D y, 3D z are the bounding box position in 3D, but we assigned -1 in our experiment since the 3D position is not applicable to our experiment. The ground truth is also converted from the YOLO format to the MOT format.

## 3.2 Optimizing Object Tracking

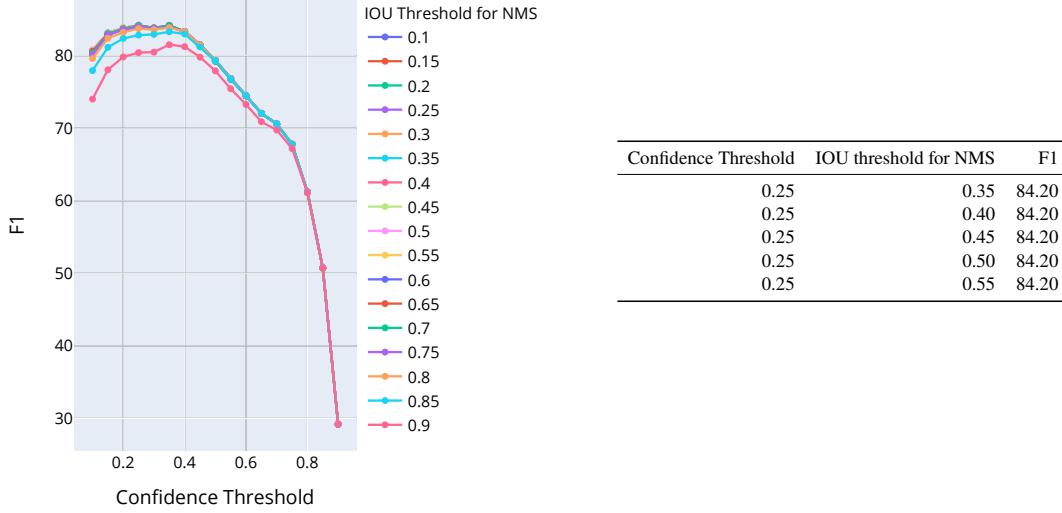
Our object tracking pipeline consists of YOLOv3 and SORT, and using the software from [19] and [24], there are 6 parameters to be set to run object tracking. We set the appropriate parameters by optimizing the detection performance of YOLOv3 and tracking performance of SORT. The following are the parameters that need to be set for YOLOv3.

- **Image size:** This is the image resolution to which input image will be resized prior to being input to the YOLOv3 model.
- **Confidence threshold:** This is the minimum confidence needed to declare a detected object. The higher the threshold, the less likely object detector will be to detect the target, but fewer mistakes will be made in detection.
- **IOU threshold for Non-Maximum Suppression (NMS):** NMS prevents multiple detections on the same target [11]. IOU threshold is used in applying NMS in our YOLOv3.

For image size, any input image will be resized to this resolution for prediction. Since the resolution of  $640 \times 640$  was the largest resolution for which the pre-trained network weights are available, we speculated that the detection performance would be best with this particular resolution. Therefore, we chose  $640 \times 640$  as the image size, which is also a default value provided by the software [19].

For the confidence threshold and IOU threshold for NMS, we apply grid search to determine suitable values. To tune these parameters, we run YOLOv3 and SORT on the sequence PartyScene. Subsequently, this sequence will not be used in testing tracking accuracy. Since both parameters range from 0 to 1, we run the detector with a step size of 0.05 and plot the detection F1 score over different confidence thresholds and IOU thresholds as shown in Figure 3.3.

The table and plot in this figure show that there are optimal F1 scores over different confidence and IOU threshold values. The table with the plot shows that the detector achieved the same F1 score over different IOU thresholds at a confidence threshold of 0.25. The default threshold values provided by the software [19] are a confidence threshold of 0.25 and IOU threshold of 0.45. Since the higher the IOU threshold we choose, it makes the detector more careful in detection (in other



**Figure 3.3:** Optimizing YOLOv3 by maximizing F1 score on “all” object classes.

words, we expect to see less FP and more FN), so we chose the confidence threshold of 0.25 and the IOU threshold for NMS to be 0.55.

For optimizing the tracking performance of SORT, the following input parameters were considered [15].

- **Max age:** Denoted  $T_{lost}$  as explained in Section 2.2, the value of maximum age determines the maximum number of frames for a trajectory to be alive while no objects are detected before its termination.
- **Min hits:** Minimum number of necessary detections before the trajectory creation and its assignment initialization.
- **IOU threshold:** Minimum IOU threshold for object matching. IOU less than this threshold indicates that the detected object bounding box does not overlap enough with the predicted bounding box, so identity will not be assigned, but an assignment occurs when IOU is higher than the threshold.

For the max age, we chose the value of 1 because Bewley *et al.* [15] justified this value with two reasons; the constant linear motion in the Kalman filter framework does not cover the true dynamics where non-linear motion exists, and SORT does not deal with object re-identification. The max age of 1 is also the default value provided by their software [24].

For the min hits and IOU threshold, we run SORT on the training sequence PartyScene for both parameters from 0.1 to 0.9 with a step size of 0.05 as a grid search. To optimize these parameters, we

**Table 3.3:** Selected optimized parameters for YOLOv3 and SORT.

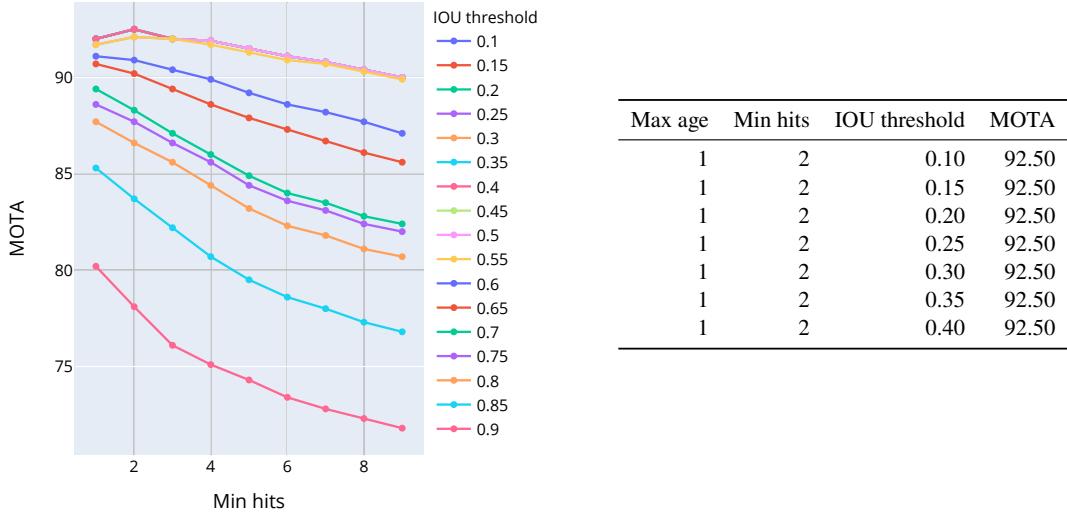
YOLOv3 parameters	Selected value
Image size	640x640
Confidence threshold	0.25
IOU threshold for NMS	0.55
SORT parameters	Selected value
Max age	1
Min hits	5
IOU threshold for matching	0.4

have tested both single-class of “person” tracking, and “all” object classes as multi-class tracking. We have chosen “person” for single-class tracking in this tuning since it is an important special case. For the “all” object classes tracking case, SORT will track all object classes available in the ground truth. Figure 3.4a shows the result of “person” class tracking. The evaluation based on MOTA over different min hits and IOU thresholds is plotted, and the highest scores are shown in the table. Although there is no single metric that has been agreed to evaluate the tracking performance, MOTA is the most popular metric in MOT and serves as a good indicator for the general tracking performance [34] [40]. Therefore, we optimized based on MOTA instead of solving multiple objective optimization problem for all the performance metrics. This result indicates that for “person” tracking, the MOTA score is optimal at min hits of 2 and the IOU threshold from 0.10 to 0.40. Figure 3.4b, which corresponds to tracking “all” object classes, shows that min hits of 9 and IOU threshold from 0.10 to 0.40 achieve the optimal performance. Note that the default values are min hits of 3 and IOU threshold of 0.3 provided by the software [24], and Bewley *et al.* [15] used these values to only track the “person” object class. Using these default values in our sequence PartyScene, we got MOTA score of 92.00 on the “person” object class and 67.60 on the “all” object classes, which are slightly lower than our optimal MOTA scores of 92.50 and 71.30 for both tracking cases respectively. Based on these results, we chose 0.4 for the IOU threshold since both tracking cases, single-class of “person” tracking and multi-class tracking, give the best MOTA score at the IOU threshold up to 0.4, and the higher the IOU threshold, the more careful the tracker is. For the min hits, we chose the value of 5, which is roughly halfway between 2 and 9, taking both single-class and multi-class tracking into consideration.

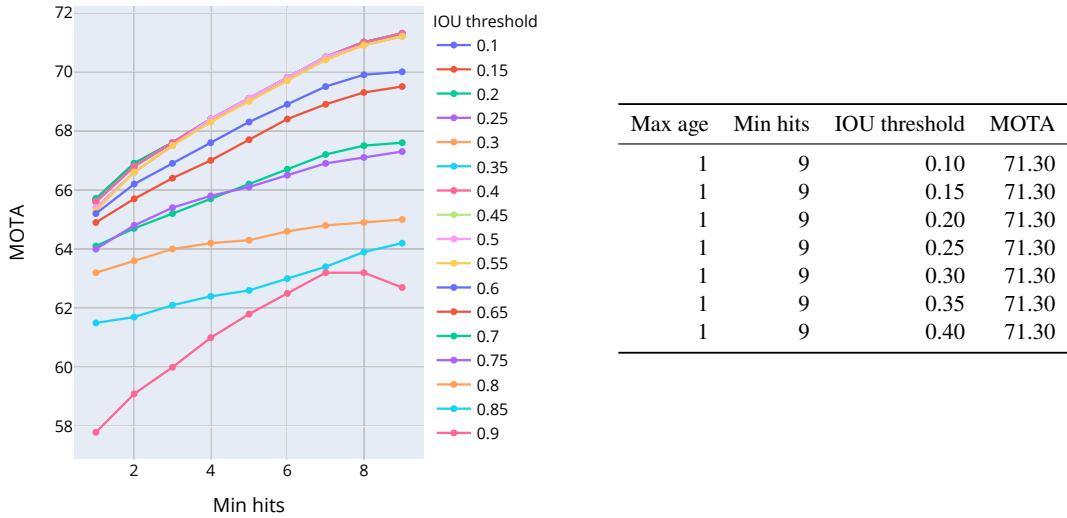
As we have selected the values of three parameters based on MOTA (max age of 1, min hits of 5, and IOU threshold of 0.4), the further justification for other metrics is given in the Appendix A.1. The selected values of parameters in YOLOv3 and SORT are summarized in Table 3.3.

### 3.3 Experiment Pipeline

Once we selected the parameters for YOLOv3 and SORT, we will run the tracking experiment to assess tracking accuracy. The experiment will involve running the object detector and tracker on



(a) Tuning parameters by maximizing MOTA on “person” object class.



(b) Tuning parameters by maximizing MOTA on “all” object classes.

**Figure 3.4:** Optimizing SORT by maximizing MOTA score on “person” and “all” object classes.

the uncompressed and compressed video sequences. For the case of compressed sequences, we apply HEVC video compression (HM16.20) at different quantization parameter (QP) and motion search range (MSR) values. Table 3.4 shows the range of QP and MSR chosen for the experiment.

**Table 3.4:** Selected QP and MSR range for the experiment.

HEVC parameter	Range of values
QP	[18, 22, 26, 30, 34, 38, 42, 46]
MSR	[8, 16, 32, 64]

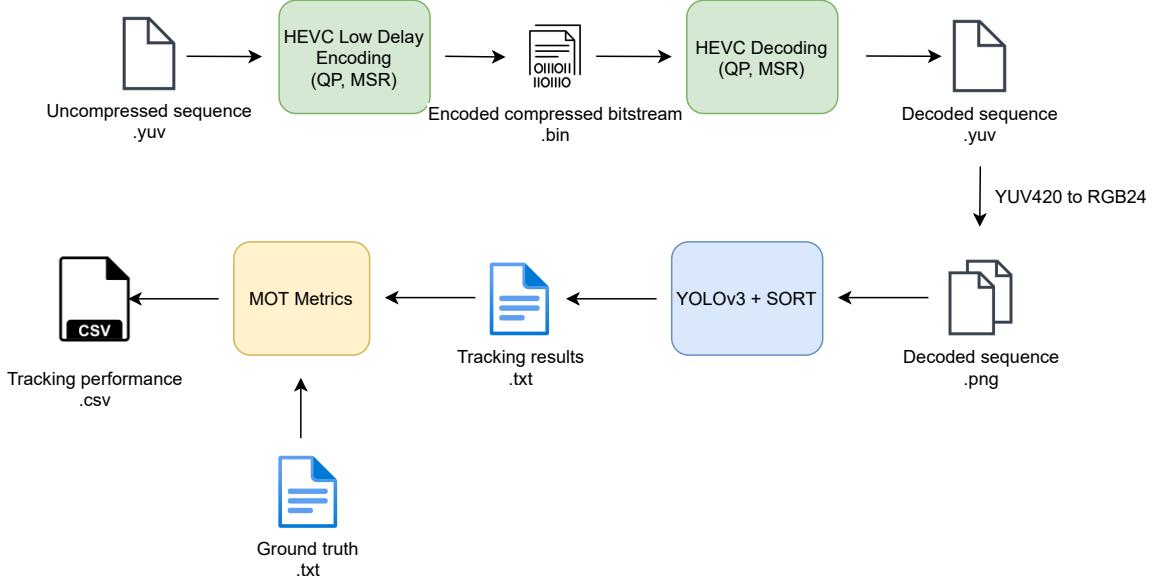
For the case of uncompressed sequences, we do not apply HEVC compression but run the object detector and tracker directly on the raw sequences. The uncompressed sequences from Table 3.1 were used, but we excluded the training sequence PartyScene since we have used this sequence to tune the parameters of the detector and tracker. Therefore, there are a total of 12 sequences in the experiments. Figure 3.5 shows the pipeline for the experiment, and each step is explained as follows.

1. **HEVC Encoding:** The experiment starts with applying HEVC encoding to the uncompressed sequences and generating the compressed bitstreams. Low Delay Encoding is used for encoding. Encoding is the process that took the most amount of time in the entire pipeline, so we encoded all the possible combinations of QP and MSR and saved all the compressed bitstreams before applying the decoder.
2. **HEVC Decoding:** After obtaining all the compressed bitstreams, we apply the HEVC decoder to the bitstreams and output the decoded sequences in the YUV420 format.
3. **YUV420 to RGB24 color conversion:** Since the YOLOv3 object detector accepts input images in the RGB24 format, we convert the decoded sequences from YUV420 to RGB24 format and save frames in PNG files.
4. **Running YOLOv3 and SORT:** Once the decoded sequence are stored in RGB24, we run the object detector and tracker to obtain the tracking results in MOT benchmark format.
5. **MOT metrics evaluation:** To assess the tracking performance, we input both the tracking results and ground truth, and the software [32] will generate the tracking performance results.

We ran YOLOv3 and SORT to detect and track each class ID shown in Table 3.1 as single-class tracking, and also detect and track the “all” object classes. In other words, we took measurements from single-class object tracking and multi-class multiple objects tracking.

## 3.4 Summary

In this chapter, we described the 13 uncompressed video sequences with the ground truth we generated to run the experiment. The object detection and tracking pipeline with YOLOv3 and SORT were described. The data conversion from the YOLO format to the MOT format was explained. Since there are 6 parameters to be set to run YOLOv3 and SORT, we tuned the parameters on the training sequence to optimize the detector and tracker and determined the suitable values for



**Figure 3.5:** Pipeline for the experiment.

the parameters before running the main experiment on other sequences. Finally, we explained the experiment pipeline. The pipeline starts with encoding to generate bitstreams, decoding, color conversion, running YOLOv3 and SORT, and evaluating the detection and tracking performance.

# Chapter 4

## Results and Discussion

This chapter presents the results from the experiment and insights found from the analysis. In the previous chapter, we explained that we ran the experiment with the various values of QP and MSR. After obtaining the tracking results, we analyze the numerical results using regression analysis and then examine tracking performance in detail on several sequences.

### 4.1 Results on Multiple Video Sequences

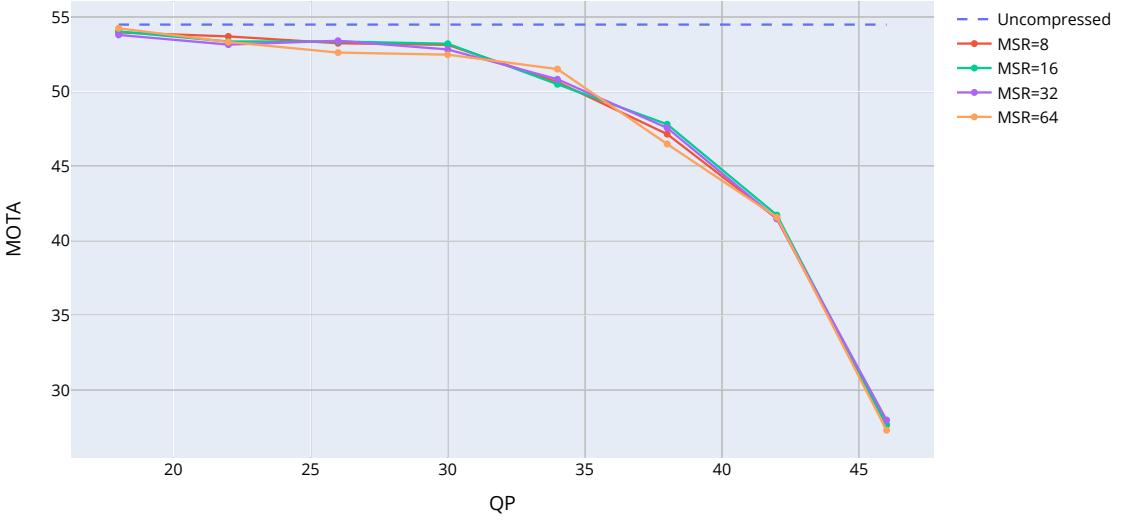
Running the experiment with all the possible combinations of QP and MSR as shown in Table 3.4, we obtained the result from all 12 video sequences. For each pair of QP and MSR, we averaged each score across all 12 video sequences. As there are 20 metrics for evaluating object tracking performance, and MOTA is a good indicator of the overall tracking performance as explained in Section 2.4, we focused our study on MOTA. In the following analysis, we visualized the data and conducted a regression analysis and t-test to quantify the results. The methodologies of statistical analysis we employed are according to the textbook [47].

#### 4.1.1 Visualization of Results across All Video Sequences

We visualized the MOTA score for “all” object classes across all video sequences at different QP and MSR as shown in Figure 4.1. As can be seen from the plot, the uncompressed video sequences achieved the highest MOTA score on average. For the compressed sequences, the average MOTA score is lower than the uncompressed result, and the higher the QP, the lower the MOTA score. Not only MOTA, we observed that the performance scores of most of the other metrics decrease as QP increases. Although the scores at different MSR are shown, we do not see any significant differences between MOTA scores at different MSR values from this plot.

Figure 4.2 shows the visualization of all the metrics<sup>1</sup> over different QP and MSR, and Table 4.1 tabulates their values. As discussed in Section 2.5, the higher the QP, the lower the bitrate,

<sup>1</sup>The metric GT is omitted from the visualization since it is a constant value, and the value can be confirmed from the table.

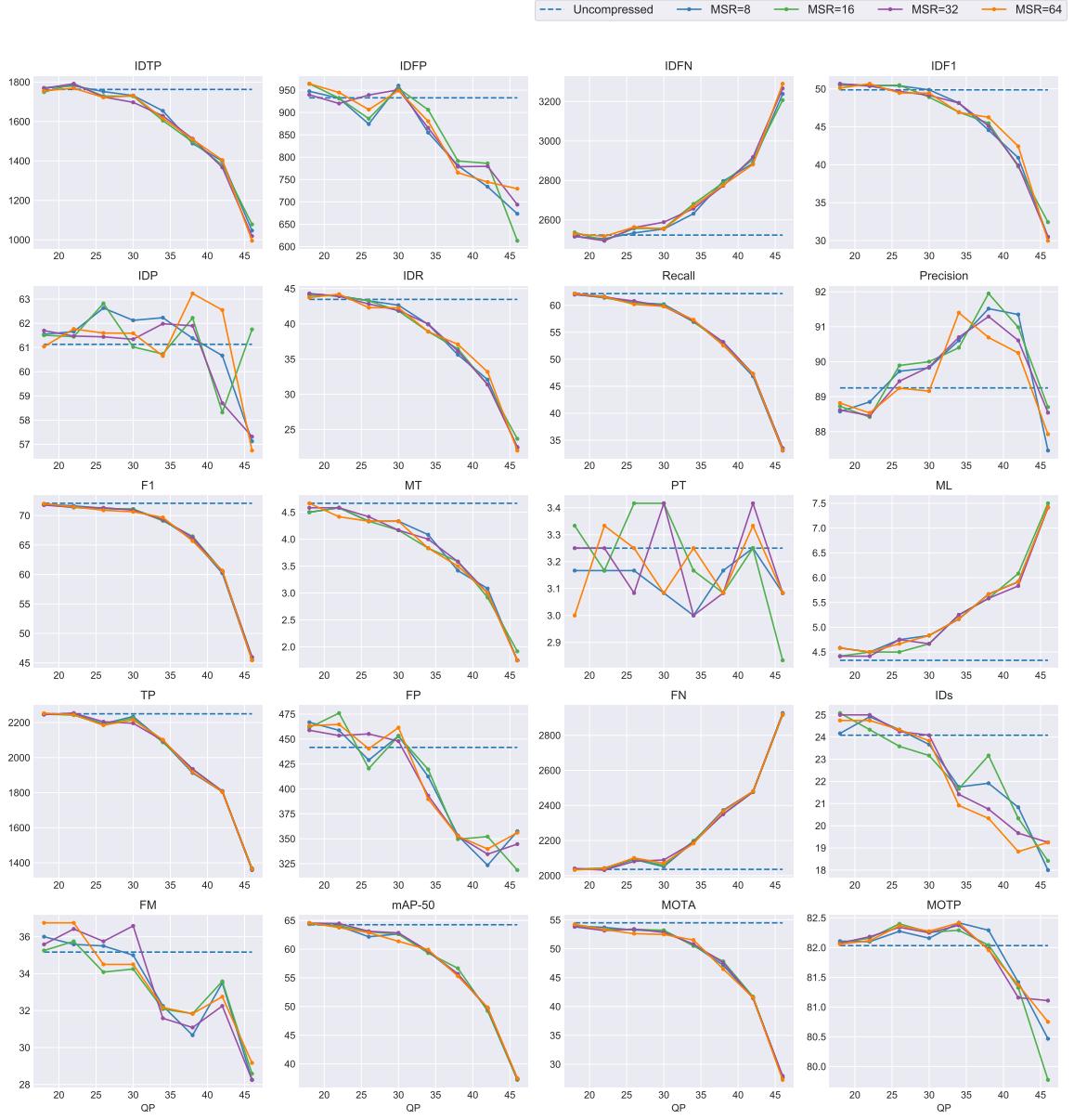


**Figure 4.1:** Average MOTA score across all video samples at different QP.

so we expected the tracking performance to be lower. Our performance results from most metrics are consistent with this expectation. To explain the decrease of MOTA defined in terms of FP, FN, and IDs based on Equation (2.24), we can see that as QP increases, FP and IDs decreases but FN increases, which is consistent with our expectation. Although the decrease of FP and IDs contribute to the increase of MOTA, the increase of FN is significantly larger than FP and IDs; thus, MOTA decreases. However, we observed that Precision and MOTP are not entirely consistent with our expectation and that the performance first increases and then decreases as QP is increased. Since video samples differ in resolution, frame rate, number of objects, and object classes, tracking performance is different. Due to the relatively small number of video samples, the standard deviation is relatively high around the average performance. We show standard deviations in Appendix A.2. To see if MSR impacts the performance scores, we visualized each performance score versus MSR as a horizontal axis, and we observed no significant dependence on MSR as shown in Figure 4.3.

#### 4.1.2 Regression Analysis

To examine the impact of QP and MSR on tracking accuracy quantitatively, we conducted a regression analysis on the entire data (all the video sequences). As we focused on the analysis on the MOTA score, we quantified the result on MOTA over QP and MSR values. As can be seen from the visualization, the relationship between MOTA and MSR is constant, but MOTA does change with QP. In order to conduct linear regression analysis, we transformed QP so that the relationship between MOTA and  $QP'$  is more linear (we call the transformed QP as  $QP'$ ). The determination of the exact QP transformation was found by plotting various forms of transformation on the scatter plots on the



**Figure 4.2:** Visualization of the average performance results at different QP (x-axis) across all video samples

individual video sequences by trial and error. The factor 52 was chosen in the denominator because since QP ranges from 0 to 51 in theory [27] and if choose 51 as a factor, QP at 51 will be undefined. To avoid this issue, we chose the factor 52. Figure 4.4 shows the scatter plots of MOTA before and after the chosen QP transformation on the example sequence BasketballPass. Using the form of Equation (4.1), the scatter plot of MOTA at different  $QP'$  shows a linear relationship, as shown in Figure 4.4b. Similar outcomes were observed in other video sequences as well.

$$QP' = \frac{1}{\frac{QP}{52} - 1} \quad (4.1)$$



**Figure 4.3:** Visualization of the average performance results at different MSR (x-axis) across all video samples.

**Table 4.1:** Average performance results across all video samples.

QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
Uncompressed	Uncompressed	1762.58	932.83	2522.67	49.88	61.13	43.48	62.17	89.25	72.08	12.25	4.67	3.25	4.33	2249.75	441.67	2035.50	24.08	35.17	64.26	54.49	82.03

(a) Mean values for the uncompressed sequence

QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	8	1770.08	947.50	2515.17	50.60	61.55	44.26	62.18	88.58	71.92	12.25	4.50	3.17	4.58	2246.75	466.83	2038.50	24.17	36.00	64.42	53.97	82.11
22	8	1780.83	931.92	2504.42	50.46	61.66	43.96	61.62	88.85	71.65	12.25	4.58	3.17	4.50	2250.00	458.75	2035.25	24.92	35.58	64.01	53.70	82.10
26	8	1751.67	874.17	2533.58	50.41	62.63	43.28	60.49	89.73	71.22	12.25	4.33	3.17	4.75	2192.92	428.92	2092.33	24.33	35.50	62.16	53.24	82.28
30	8	1731.50	960.08	2553.75	49.88	62.12	42.65	60.21	89.83	71.07	12.25	4.33	3.08	4.83	2234.42	453.17	2050.83	23.67	35.00	62.66	53.13	82.16
34	8	1653.92	855.08	2631.33	48.14	62.23	39.93	57.12	90.61	69.26	12.25	4.08	3.00	5.17	2092.58	412.42	2192.67	21.75	32.25	59.58	50.64	82.42
38	8	1488.50	781.08	2796.75	44.56	61.38	35.62	52.70	91.52	65.97	12.25	3.42	3.17	5.67	1912.42	353.17	2372.83	21.92	30.67	55.50	47.15	82.29
42	8	1399.25	733.83	2886.00	40.92	60.67	32.08	46.86	91.35	60.24	12.25	3.08	3.25	5.92	1805.67	323.42	2479.58	20.83	33.50	49.66	41.48	81.42
46	8	1046.50	673.33	3238.75	30.48	57.12	22.40	33.09	87.46	45.48	12.25	1.75	3.08	7.42	1358.33	357.50	2926.92	18.00	28.25	37.34	27.64	80.47

(b) Mean values for MSR = 8

QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	16	1749.25	964.25	2536.00	50.48	61.51	44.07	62.04	88.73	71.88	12.25	4.50	3.33	4.42	2248.00	461.50	2037.25	25.08	35.25	64.37	54.03	82.08
22	16	1790.08	932.17	2495.17	50.40	61.45	43.96	61.39	88.42	71.38	12.25	4.58	3.17	4.50	2242.17	476.08	2043.08	24.33	35.75	64.09	53.36	82.16
26	16	1728.75	886.25	2556.50	50.48	62.82	43.28	60.43	89.89	71.21	12.25	4.33	3.42	4.50	2190.50	420.50	2094.75	23.58	34.08	63.03	53.36	82.40
30	16	1728.33	954.08	2556.92	48.91	61.02	41.83	60.04	90.00	71.10	12.25	4.17	3.42	4.67	2225.42	453.50	2059.83	23.17	34.25	62.62	53.21	82.26
34	16	1604.67	906.00	2608.58	46.94	60.73	38.90	56.89	90.40	69.10	12.25	3.83	3.17	5.25	2087.08	419.58	2198.17	21.67	32.08	59.32	50.49	82.29
38	16	1496.58	791.58	2788.67	45.47	62.23	36.46	53.13	91.95	66.47	12.25	3.58	3.08	5.58	1934.58	349.58	2350.67	23.17	31.83	56.66	47.81	82.04
42	16	1379.33	786.17	2905.92	39.76	58.32	31.38	47.28	90.98	60.56	12.25	2.92	3.25	6.08	1809.33	352.17	2475.92	20.33	33.58	49.24	41.72	81.33
46	16	1078.17	613.00	3207.08	32.42	61.75	23.68	33.48	88.70	45.99	12.25	1.92	2.83	7.50	1368.58	318.58	2916.67	18.42	28.58	37.21	27.67	79.77

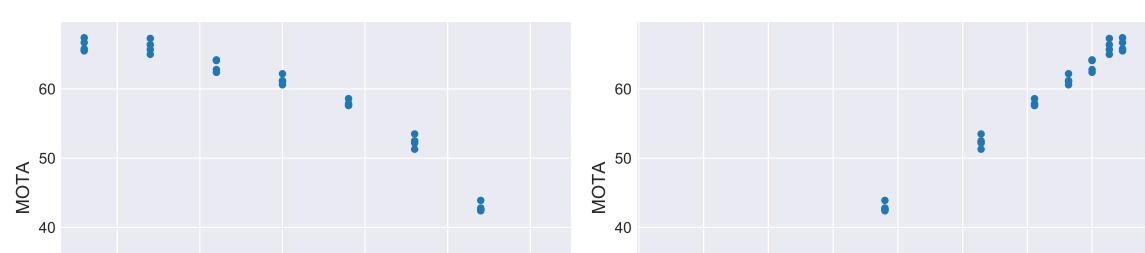
(c) Mean values for MSR = 16

QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	32	1768.58	939.58	2516.67	50.67	61.70	44.31	61.98	88.62	71.78	12.25	4.58	3.25	4.42	2245.33	458.83	2039.92	25.00	35.58	64.54	53.80	82.08
22	32	1791.42	920.08	2493.83	50.38	61.48	43.92	61.49	88.46	71.42	12.25	4.58	3.25	4.42	2254.00	453.50	2031.25	25.00	36.42	64.47	53.15	82.18
26	32	1724.42	939.08	2560.83	49.68	61.44	42.82	60.80	89.44	71.32	12.25	4.42	3.08	4.75	2204.33	455.17	2080.92	24.25	35.75	63.11	53.41	82.34
30	32	1697.08	950.92	2588.17	49.14	61.34	41.91	59.79	89.85	70.91	12.25	4.17	3.42	4.67	2196.08	447.92	2089.17	24.08	36.58	62.84	52.83	82.25
34	32	1628.42	865.50	2656.83	48.13	61.98	39.98	57.10	90.70	69.32	12.25	4.00	3.00	5.25	2096.75	393.17	2188.50	21.42	31.58	59.66	50.83	82.38
38	32	1512.50	778.75	2772.75	45.10	61.90	36.13	53.20	91.29	66.31	12.25	3.58	3.08	5.58	1934.67	352.58	2350.58	20.75	31.08	55.69	47.57	81.98
42	32	1367.25	779.58	2918.00	39.92	58.70	31.38	47.28	90.61	60.48	12.25	3.00	3.48	5.83	1808.25	334.58	2477.00	19.67	32.25	49.68	41.48	81.16
46	32	1018.67	693.75	3266.58	30.47	57.32	22.52	33.53	88.54	45.93	12.25	1.75	3.08	7.42	1363.75	344.67	2921.50	19.25	28.25	37.40	27.29	81.11

(d) Mean values for MSR = 32

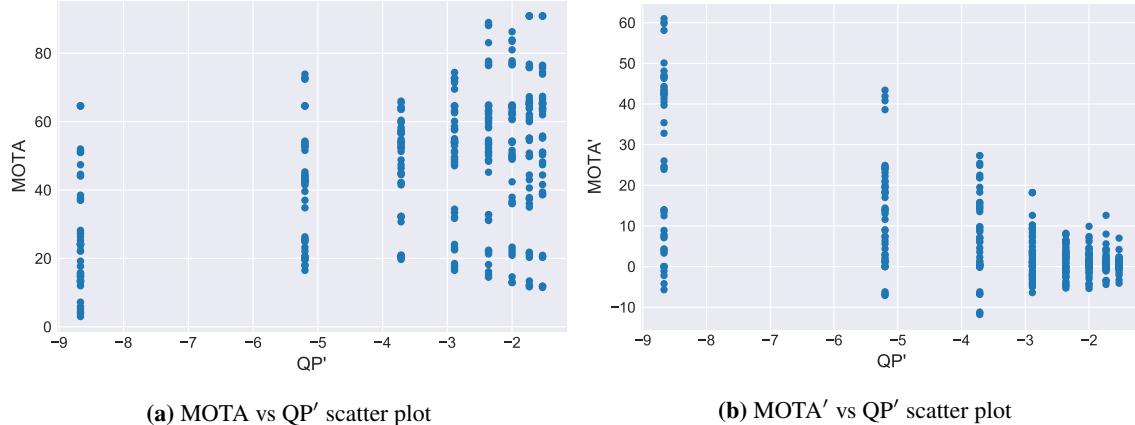
QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	64	1756.00	964.33	2529.25	50.10	61.04	43.73	62.17	88.82	72.02	12.25	4.67	3.00	4.58	2252.83	463.50	2032.42	24.75	36.75	64.59	54.24	82.06
22	64	1768.75	944.75	2516.50	50.68	61.77	44.22	61.56	88.53	71.48	12.25	4.42	3.33	4.50	2244.83	464.67	2040.42	24.75	36.75	63.76	53.33	82.12
26	64	1722.25	906.50	2563.00	49.46	61.60	42.30	60.18	89.24	70.90	12.25	4.33	3.25	4.67	2184.42	440.33	2100.83	24.33	34.50	62.87	52.62	82.37
30	64	1731.25	949.00	2554.00	49.44	61.59	42.27	59.80	89.16	70.64	12.25	4.33	3.08	4.83	2214.75	461.50	2070.50	23.83	34.50	61.37	52.47	82.28
34	64	1615.58	880.42	2669.67	46.90	60.65	38.91	57.32	91.40	69.69	12.25	3.83	3.25	5.17	2102.25	389.75	2183.00	20.92	32.17	59.90	51.51	82.42
38	64	1508.92	765.42	2776.33	46.25	63.23	37.09	52.56	90.69	65.69	12.25	3.50	3.08	5.67	1917.75	352.58	2367.50	20.33	31.83	55.29	46.48	81.96
42	64	1404.50	744.42	2880.75	42.43	62.55	33.18	47.31	90.25	60.65	12.25	3.00	3.33	5.92	1805.17	339.75	2480.08	18.83	32.75	49.85	41.57	81.38
46	64	995.00	729.42	3290.25	29.98	56.74	22.02	33.18	87.92	45.43	12.25	1.75	3.08	7.42	1364.33	356.08	2920.92	19.25	29.17	37.50	27.29	80.75

(e) Mean values for MSR = 64



**Figure 4.4:** Scatter plots before and after the QP transformation on the BasketballPass sequence.

However, using all the video sequences, which have different MOTA scores, the variance at each QP' is high, and little insight can be derived from the high variance of data as shown in Figure 4.5a.



**Figure 4.5:** Scatter plots before and after the MOTA transformation across all the video sequences.

Therefore, we transformed MOTA such that we subtract each MOTA score at different  $QP'$  from the uncompressed MOTA score, as represented below.

$$MOTA'_i = MOTA(\text{Uncompressed}) - MOTA(QP'_i) \quad (4.2)$$

Doing this transformation, we only consider the drop of the MOTA score from the uncompressed sequence, and the variance is lower at higher  $QP'$ , in other words, the variance is lower at lower QP. This is expected because  $MOTA(\text{Uncompressed})$  and  $MOTA(QP_i)$  are both random variables, and the difference of these two variables will be lower at lower QP. Figure 4.5b shows the scatter plot of all the video sequences after the MOTA transformation. We can see from this figure that the relationship between  $MOTA'$  and  $QP'$  is still linear but error variance at each  $QP'$  is not constant. Therefore, the weighted least squares method is adopted in the regression model. Also, since we have two continuous independent variables of  $QP'$  and MSR and each may impact the continuous response variable of the  $MOTA'$  score, we applied a multiple linear regression across all the video sequences. We represent the regression model as,

$$MOTA'_i = \beta_0 + \beta_1 \cdot QP'_i + \beta_2 \cdot \text{MSR}_i + \beta_3 \cdot QP'_i \cdot \text{MSR}_i + \epsilon_i, \quad (4.3)$$

where  $\beta_0$  is the intercept;  $\beta_1$  and  $\beta_2$  are the parameters of independent variables  $QP'$  and  $\text{MSR}$ ;  $\beta_3$  is the parameter of the interaction term of  $QP'$  and  $\text{MSR}$ . The interaction term is included to see if  $\text{MSR}$  and  $QP'$  depend on each other. We assume that  $\epsilon_i$  is the independent normally distributed error with mean 0 and variance  $\sigma_i^2$ , where  $i$  indicates the  $i$ -th trial from the independent variables. As we are applying the weighted least squares method, the weight at each  $QP'$  can be calculated as

$$w_i = \frac{1}{\sigma_i^2} \quad (4.4)$$

**Table 4.2:** Regression analysis result of the MOTA score for “all” object classes across all video sequences.

$\beta_0$	-4.32
$\beta_1$	-2.97
$\beta_2$	$<  10^{-2} $
$\beta_3$	$<  10^{-2} $
p-value( $\beta_0$ )	$<  10^{-6} $
p-value( $\beta_1$ )	$<  10^{-9} $
p-value( $\beta_2$ )	0.83
p-value( $\beta_3$ )	0.77

Equation (4.4) is according to [47, p. 422]. Since  $\sigma_i$  is the true value of standard deviation at each QP', we estimate  $w_i$  using the sample standard deviation  $s_i$  of data points at each QP':

$$w_i = \frac{1}{s_i^2} \quad (4.5)$$

We utilized the Python package [48] for computation, and Table 4.2 shows the results from the multiple linear regression analysis for the MOTA score. To see if the impact of the independent variable is statistically significant on the dependent variable, we conducted a hypothesis test at a significance level of 0.05. The null hypothesis is the case the parameter is 0 and the alternative hypothesis is the case the parameter is not zero.

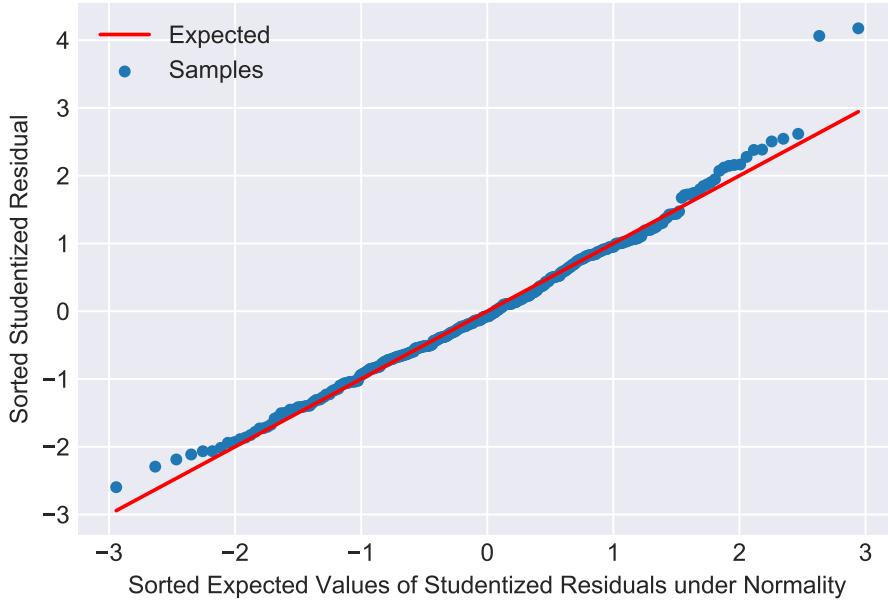
$$\begin{aligned} H_0 : \beta_c &= 0 \\ H_1 : \beta_c &\neq 0 \end{aligned} \quad (4.6)$$

where  $c = 0, 1, 2, 3$ . From the table, the p-value for the test on  $\beta_1$  is less than 0.05 while the p-values for the tests on  $\beta_2$  and  $\beta_3$  are greater than 0.05. This shows that we reject the null hypothesis of the test on  $\beta_1$ , so QP significantly impacts MOTA at 95% confidence. However, we fail to reject the null hypothesis for  $\beta_2$  and  $\beta_3$ . Hence, the available data is insufficient to prove that MSR and the interaction term have statistically significant impact on the MOTA score.

To further justify these results, we tested the adequacy of our regression model with the normal probability plot of the residuals. To plot this, we would have to obtain the residuals and the expected values of the residuals under normality. Since the normal probability plot requires variance to be constant and we have a model with unequal variance, we studentize the residuals in order to obtain the residuals with constant variance of 1. The studentized residuals  $r_i$  can be obtained as

$$r_i = \frac{\text{MOTA}'_{i,\text{measurement}} - \text{MOTA}'_{i,\text{predicted}}}{s_i} \quad (4.7)$$

where  $s_i$  is an estimator to the standard deviation  $\sigma_i$ . These studentized residuals will be sorted and be plotted in the normal probability plot. Adapted from [47, p. 111], the expected values of



**Figure 4.6:** Normal probability plot to test the adequacy of the regression model.

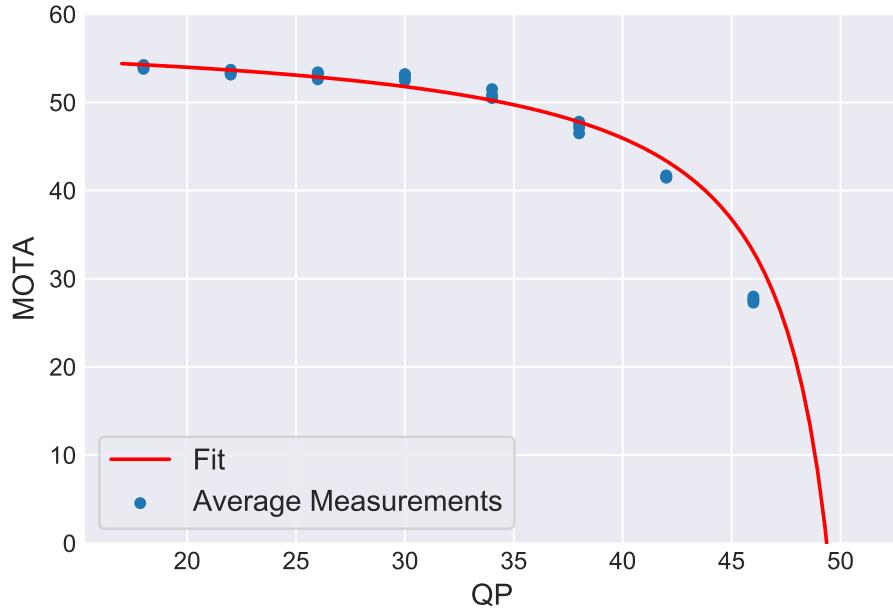
studentized residuals under normality can be represented as the following equation.

$$E(r_k) = \text{ppf}\left(\frac{k - 0.375}{n + 0.25}\right) \quad (4.8)$$

ppf is a percent point function (ppf) and  $k$  indicates the  $k$ -th smallest studentized residual. The range of  $k$  is  $1, 2, 3 \dots n$ , and the smallest residual will be  $k = 1$  and the largest residual will be  $k = n$ . This equation represents the sorted expected values of studentized residuals under the normal distribution with mean of 0 and variance of 1. Figure 4.6 shows the normal probability plot of sorted studentized residuals and the sorted expected values of studentized residuals under normality. Our samples are close to the normality which would be a straight line with a slope of 1.0. Fitting this plot, we obtain a slope of 1.001, intercept of 0.004, and the coefficient of determination  $R^2$  of 0.985. The relatively high value of  $R^2$  indicates that the error term  $\epsilon_i$  is close to the normal distribution. Therefore, we conclude that our regression model (4.3) is adequate. As we justified our regression model and from the hypothesis testing, we conclude that MSR does not impact the MOTA score, and QP and MSR do not depend on each other; therefore, our prediction for MSR, as explained in Section 2.5, is inconsistent with our results. Thus, we will limit our further study to QP only.

We showed that our regression model (4.3) is appropriate and therefore, we propose to formulate the relationship between MOTA and QP. From this regression, we have the following formulations.

$$\text{MOTA}' = \beta_0 + \beta_1 \cdot \text{QP}' \quad (4.9)$$



**Figure 4.7:** Scatter plot of average values of MOTA measurements and fitted model.

$$\text{MOTA} = \text{MOTA(Uncompressed)} - \beta_0 - \beta_1 \cdot \frac{1}{\frac{\text{QP}}{52} - 1} \quad (4.10)$$

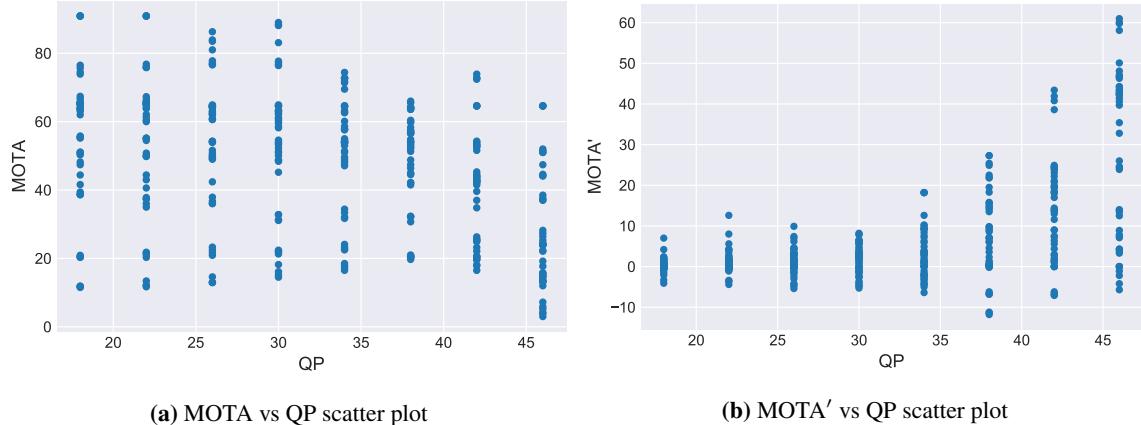
Substituting the estimated values of parameters from Table 4.2, we have the following equation.

$$\text{MOTA} = 58.81 + 2.97 \cdot \frac{1}{\frac{\text{QP}}{52} - 1} \quad (4.11)$$

We plotted the average values of MOTA measurements at each QP and overlaid the fitted model 4.11 on the scatter plot as shown in Figure 4.7. Note that our model is fitted according to the raw data points for “all” object classes while average MOTA values are plotted for visualizing purposes. Also, our model is fitted based on the QP range from 18 to 46, and since QP ranges from 0 to 51 in theory [27], our model is extrapolated. However, from the practical point of view, QP = 51 is virtually never used and the QP range from 18 to 46 is sufficient. Future study could include the whole range of QP to fit the model. In addition, not only MOTA, the same analysis procedure can be applied to all tracking metrics, and we tabulated the regression analysis results of all the metrics in Appendix A.3.

### 4.1.3 One-sided t-test

Finally, to quantify the results to answer the question - at which QP, the performance score is significantly lower than the score at the uncompressed case, we conducted the one-sided t-test. As explained in Section 4.1.2, without the MOTA transformation, the variance is high. To gain the better insight with the reduced variance, we consider the difference between the score at the uncompressed



**Figure 4.8:** Scatter plots before and after the MOTA transformation for t-test.

case and the score at each QP. For the one-sided t-test, we included all the metrics, so we will transform each performance score as

$$p_{\text{diff}} = \begin{cases} p_{\text{comp}} - p_{\text{uncomp}}, & \text{for IDFN, FN, PT, ML} \\ p_{\text{uncomp}} - p_{\text{comp}}, & \text{else} \end{cases} \quad (4.12)$$

where  $p_{\text{comp}}$  is the performance score at each QP,  $p_{\text{uncomp}}$  is the the performance score at the uncompressed case, and  $p_{\text{diff}}$  represents the difference of such scores. Since the scores IDFN, FN, PT, and ML increase as QP increases according to Figure 4.2, we subtract  $p_{\text{comp}}$  from  $p_{\text{uncomp}}$ . For other metrics, the scores decrease as QP increases, so we subtract  $p_{\text{uncomp}}$  from  $p_{\text{comp}}$ . Figure 4.8 shows that after the transformation of the score, MOTA as an example, the variance is reduced at lower QP. To perform t-test, the hypotheses are

$$\begin{aligned} H_0 : \mu_{\text{diff}} &\leq 0 \\ H_1 : \mu_{\text{diff}} &> 0 \end{aligned} \quad (4.13)$$

where  $\mu_{\text{diff}}$  is the mean of 12 transformed score values  $p_{\text{diff}}$ . The null hypothesis is the case when the mean  $\mu_{\text{diff}}$  is less than or equal to the population mean of 0. The alternative hypothesis is the case when the mean  $\mu_{\text{diff}}$  is greater than the population mean of 0. Applying the one-sided t-test for “all” object classes across all the video sequences at a significance level of 0.05, we obtain the results in Table 4.3 using the Python package [49]. The value less than the significance level of 0.05 is bolded in the table. From this result, for example, the MOTA score at QP = 18 is significantly lower than the score of the uncompressed sequence with 95% confidence as we reject the null hypothesis. In fact, we can conclude for most of the metrics except Precision and PT as following; the performance score is greater than the uncompressed score for the metrics IDFN, FN, and ML but the performance score is less than the uncompressed score for other metrics at specific QP with 95% confidence. From Figure 4.2, the Precision score increases first and then decreases as QP increases. Therefore,

**Table 4.3:** One-sided t-test for “all” object classes across all the video sequences.

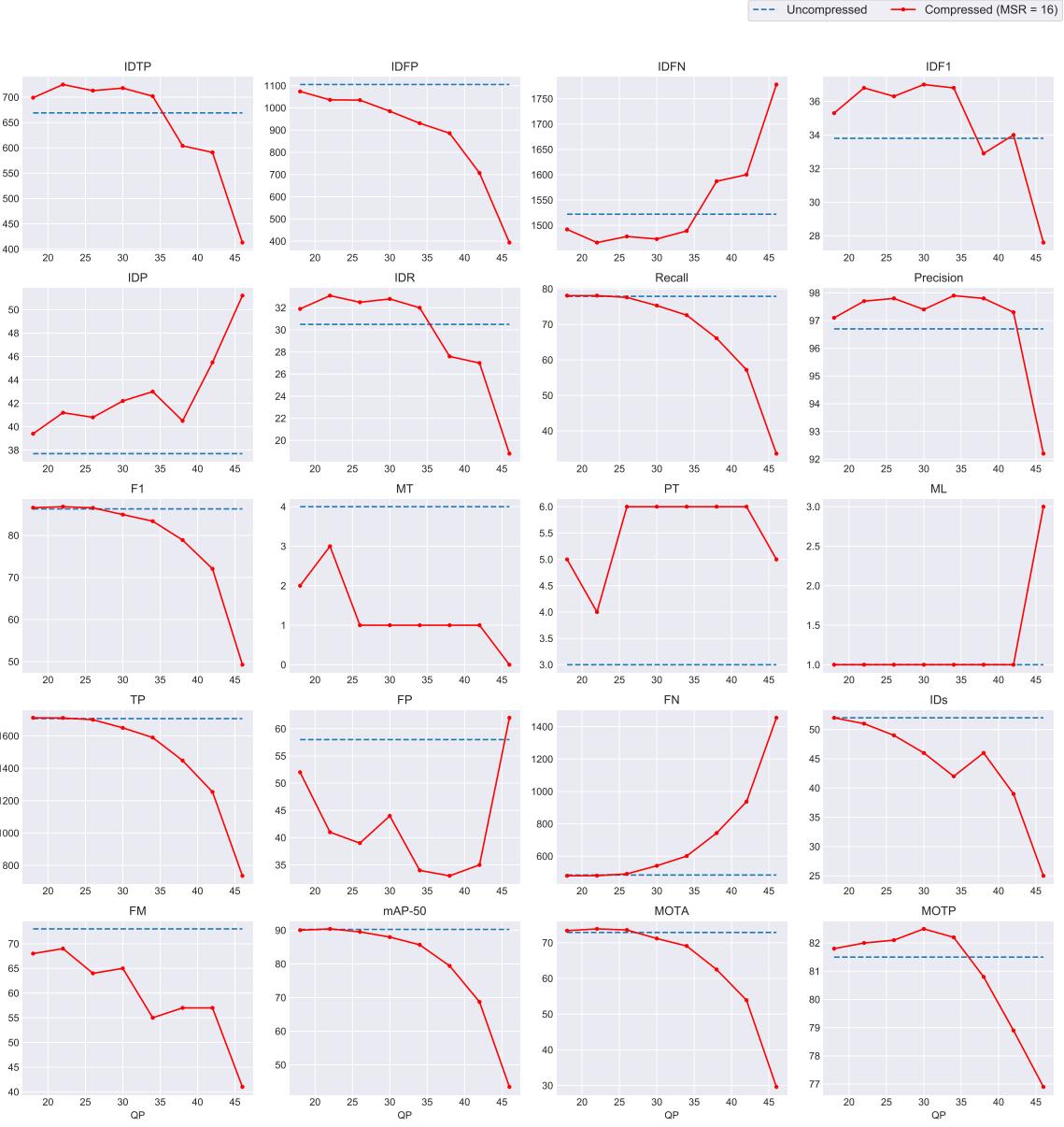
	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
p-value(QP=18)	0.44	0.90	0.44	0.87	0.70	0.90	0.31	<b>0.02</b>	0.06	<b>0.05</b>	0.79	<b>0.02</b>	0.29	0.99	0.29	0.98	0.94	0.68	<b>0.03</b>	0.97
p-value(QP=22)	0.91	0.49	0.91	0.79	0.71	0.79	< $ 10^{-2} $	<b>0.03</b>	< $ 10^{-2} $	<b>0.02</b>	0.63	<b>0.03</b>	0.39	0.99	0.39	0.92	0.93	0.38	< $ 10^{-2} $	0.97
p-value(QP=26)	0.16	0.07	0.16	0.58	0.91	0.20	< $ 10^{-3} $	0.89	< $ 10^{-3} $	< $ 10^{-3} $	0.59	< $ 10^{-3} $	< $ 10^{-2} $	0.27	< $ 10^{-2} $	0.53	0.41	<b>0.02</b>	< $ 10^{-3} $	1.00
p-value(QP=30)	0.08	0.76	0.08	0.24	0.65	<b>0.03</b>	< $ 10^{-3} $	0.83	< $ 10^{-2} $	< $ 10^{-3} $	0.50	< $ 10^{-2} $	0.10	0.84	0.10	0.29	0.47	0.08	< $ 10^{-3} $	0.99
p-value(QP=34)	< $ 10^{-2} $	< $ 10^{-2} $	< $ 10^{-2} $	< $ 10^{-2} $	0.60	< $ 10^{-4} $	< $ 10^{-3} $	1.00	< $ 10^{-4} $	< $ 10^{-3} $	0.95	< $ 10^{-3} $	< $ 10^{-3} $	<b>0.01</b>	< $ 10^{-3} $	< $ 10^{-2} $	<b>0.01</b>	< $ 10^{-4} $	0.99	
p-value(QP=38)	< $ 10^{-3} $	< $ 10^{-3} $	< $ 10^{-3} $	< $ 10^{-4} $	0.88	< $ 10^{-6} $	< $ 10^{-7} $	1.00	< $ 10^{-6} $	< $ 10^{-7} $	0.92	< $ 10^{-7} $	< $ 10^{-4} $	< $ 10^{-4} $	< $ 10^{-2} $	< $ 10^{-3} $	< $ 10^{-3} $	< $ 10^{-5} $	0.57	
p-value(QP=42)	< $ 10^{-5} $	< $ 10^{-4} $	< $ 10^{-5} $	< $ 10^{-6} $	0.17	< $ 10^{-7} $	< $ 10^{-10} $	1.00	< $ 10^{-8} $	< $ 10^{-9} $	0.40	< $ 10^{-6} $	< $ 10^{-8} $	< $ 10^{-5} $	< $ 10^{-8} $	< $ 10^{-3} $	0.12	< $ 10^{-3} $	< $ 10^{-8} $	<b>0.01</b>
p-value(QP=46)	< $ 10^{-8} $	< $ 10^{-4} $	< $ 10^{-9} $	0.04	< $ 10^{-9} $	< $ 10^{-13} $	0.23	< $ 10^{-12} $	< $ 10^{-13} $	0.69	< $ 10^{-8} $	< $ 10^{-11} $	0.02	< $ 10^{-11} $	0.01	< $ 10^{-9} $	< $ 10^{-11} $	< $ 10^{-3} $		

the result from the one-sided t-test is consistent with the the visualization. According to Equation 2.15, Precision depends on TP and FP, and the increase of p-value in FP can be confirmed at QP = 30. The detailed reasoning for this outcome is explained in the following Section 4.2. A slight increase of MOTP can also be confirmed from QP = 26. We also applied the one-sided t-test to the “person” object class and similar conclusions can be made. See Appendix A.4 for the “person” object case.

As we just showed the results from the visualization and the statistical analysis for the multiple video sequences, most metrics behave in a way as expected, but the outcomes of Precision and MOTP were not entirely expected. We will further analyze individual sequences in the next section.

## 4.2 Case Studies of Individual Sequences

In this section, we will look into tracking performance on the individual sequences to extend the analysis of all video sequences obtained in the previous section. Out of 12 video samples, we observed that most are consistent with the hypothesis that tracking performance scores decrease as QP increases, and the performance on the compressed sequences is lower than on uncompressed sequences. However, in some video samples, the performance scores increase midway through the QP range, but start decreasing at high QP, so it may happen that the performance on a compressed sequence with a moderate QP is better than the uncompressed one. These exceptional cases were observed in Cactus, BlowingBubbles, FourPeople, and KristenAndSara. To uncover the reason for this, we will examine the video samples: BasketballPass, Johnny, BlowingBubbles, and Cactus as case studies. Instead of tracking “all” object classes, we limit the tracking to the “person” object class in this analysis. The “person” object class is “cleaner” in terms of results than other classes, because the “person” class is more frequently present across the frames; in contrast, other object classes such as “sports ball” and “handbag” are rarely observed. We also included the “potted plant” class in Cactus. As different MSR did not impact the tracking performance according to the previous section, we selected MSR = 16 in the following analysis. Finally, we primarily focused on the MOTA score since this metric correlates the most with the human subjective impression of tracking according to [46]. Since MOTA is defined in terms of FN, FP, and IDs from Equation (2.24), we will examine each of these metrics closely.



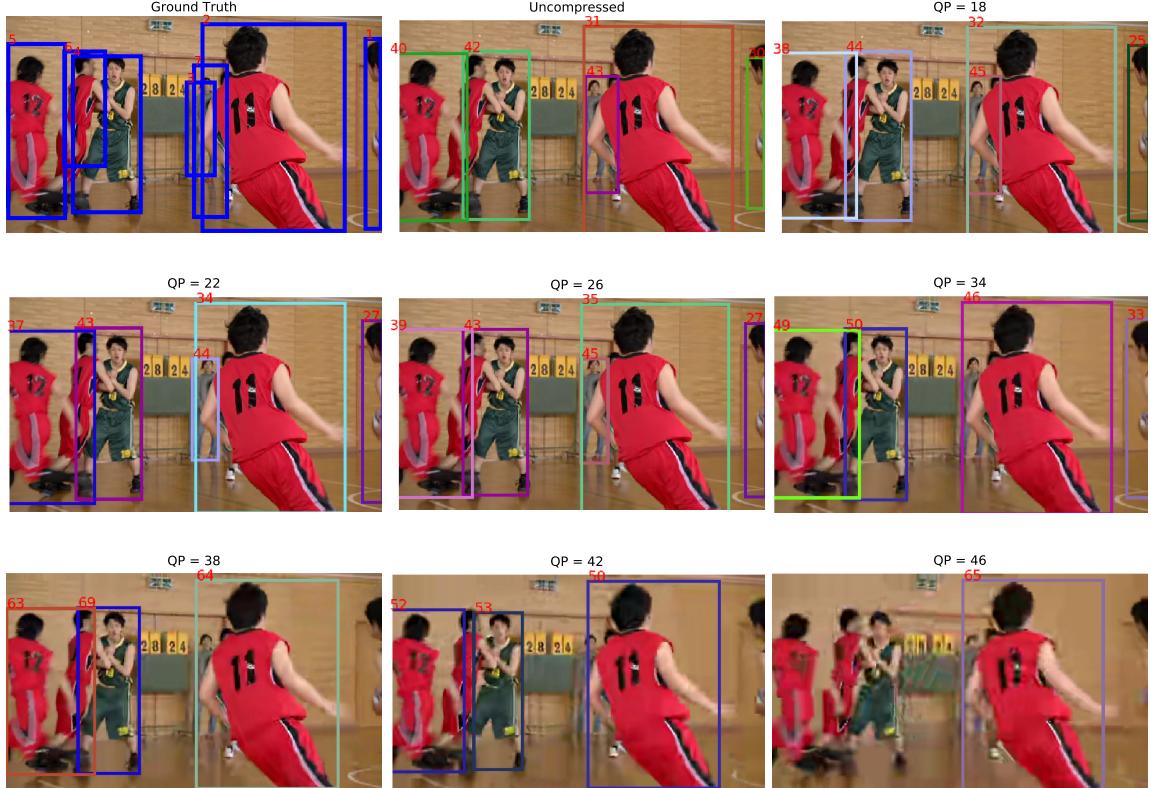
**Figure 4.9:** Visualization of the performance results on BasketballPass at different QP for the “person” object class.

#### 4.2.1 BasketballPass

We examined the results from the video sequence BasketballPass. This sequence consists of 7 “person” objects, and the occlusion frequently occurs in the video. Figure 4.9 shows all the performance scores at different QPs at MSR = 16 and Table 4.4 shows each numerical value. The results are consistent with the average results shown in Section 4.1, such that the performance decreases as QP increases on most of the metrics except IDP and MOTP. To examine the results more thoroughly, we inspected the video sequence frame by frame at different QP. Figure 4.10 shows

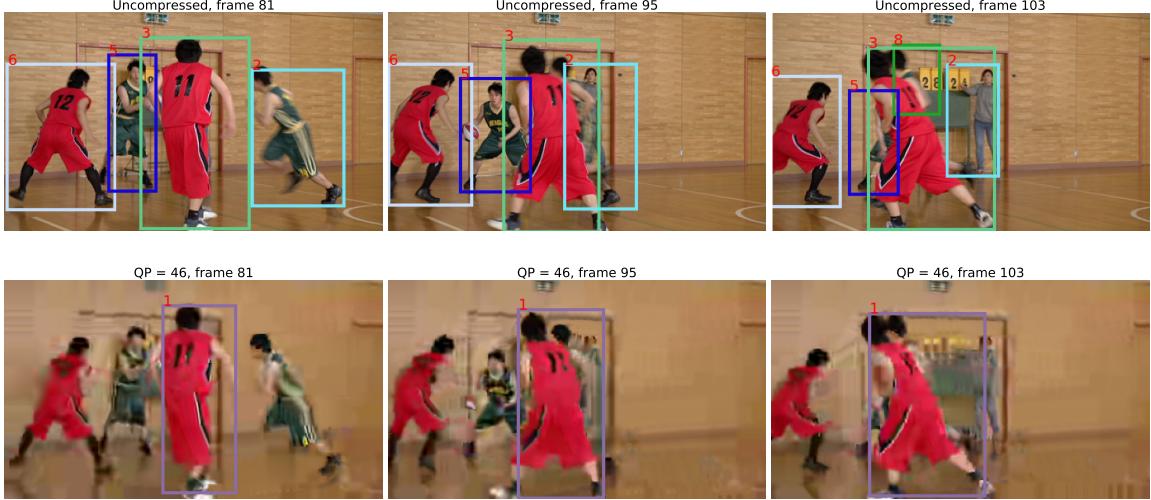
**Table 4.4:** Performance results on BasketballPass.

QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
Uncompressed	Uncompressed	669.00	1105.00	1522.00	33.80	37.70	30.50	77.90	96.70	86.29	8	4	3	1	1707	58	484	52	73	90.25	72.90	81.50
18	16	699.00	1074.00	1492.00	35.30	39.40	31.90	78.10	97.10	86.57	8	2	5	1	1712	52	479	52	68	90.03	73.40	81.80
22	16	725.00	1036.00	1466.00	36.80	41.20	33.10	78.10	97.70	86.81	8	3	4	1	1711	41	480	51	69	90.42	73.90	82.00
26	16	713.00	1035.00	1478.00	36.30	40.80	32.50	77.60	97.80	86.54	8	1	6	1	1700	39	491	49	64	89.54	73.60	82.10
30	16	718.00	985.00	1473.00	37.00	42.20	32.80	75.30	97.40	84.94	8	1	6	1	1650	44	541	46	65	87.98	71.20	82.50
34	16	702.00	931.00	1489.00	36.80	43.00	32.00	72.60	97.90	83.37	8	1	6	1	1590	34	601	42	55	85.69	69.10	82.20
38	16	604.00	886.00	1587.00	32.90	40.50	27.60	66.10	97.80	78.88	8	1	6	1	1448	33	743	46	57	79.42	62.50	80.80
42	16	591.00	707.00	1600.00	34.00	45.50	27.00	57.20	97.30	72.05	8	1	6	1	1254	35	937	39	57	68.72	53.90	78.90
46	16	413.00	394.00	1778.00	27.60	51.20	18.80	33.60	92.20	49.25	8	0	5	3	736	62	1455	25	41	43.45	29.60	76.90



**Figure 4.10:** Comparison of ground truth and tracking results on the BasketballPass sequence in frames at 320 at different QP.

the comparison of ground truth, tracking results without compression, and results with compression at different QP in the BasketballPass sequence at frame 320, as an example. This comparison illustrates that the higher the QP, the lower the image quality. As the quality decreases, the YOLOv3 detector starts failing to detect the “person” objects, and hence the FN increases. We also confirmed that IDs decreases because the number of detected objects decreases; therefore, the number of times the objects that are considered occluded also decreases. When the number of detected occlusion decreases, ID switch on the trajectories should decrease, since SORT does not have a feature to perform re-identification (re-ID) of the object identities. An example of this situation is shown in Figure 4.11. The detected occlusion is shown in the uncompressed frames from 81 to 103, where the identities of two “person” objects are swapped due to occlusion. That same occlusion is not detected in the compressed frames at QP = 46. From FN, FP, and IDs, the increase of FN is significantly



**Figure 4.11:** Comparison of BasketballPass frames 81, 95, 103 to explain IDs results.

**Table 4.5:** Performance results on Johnny.

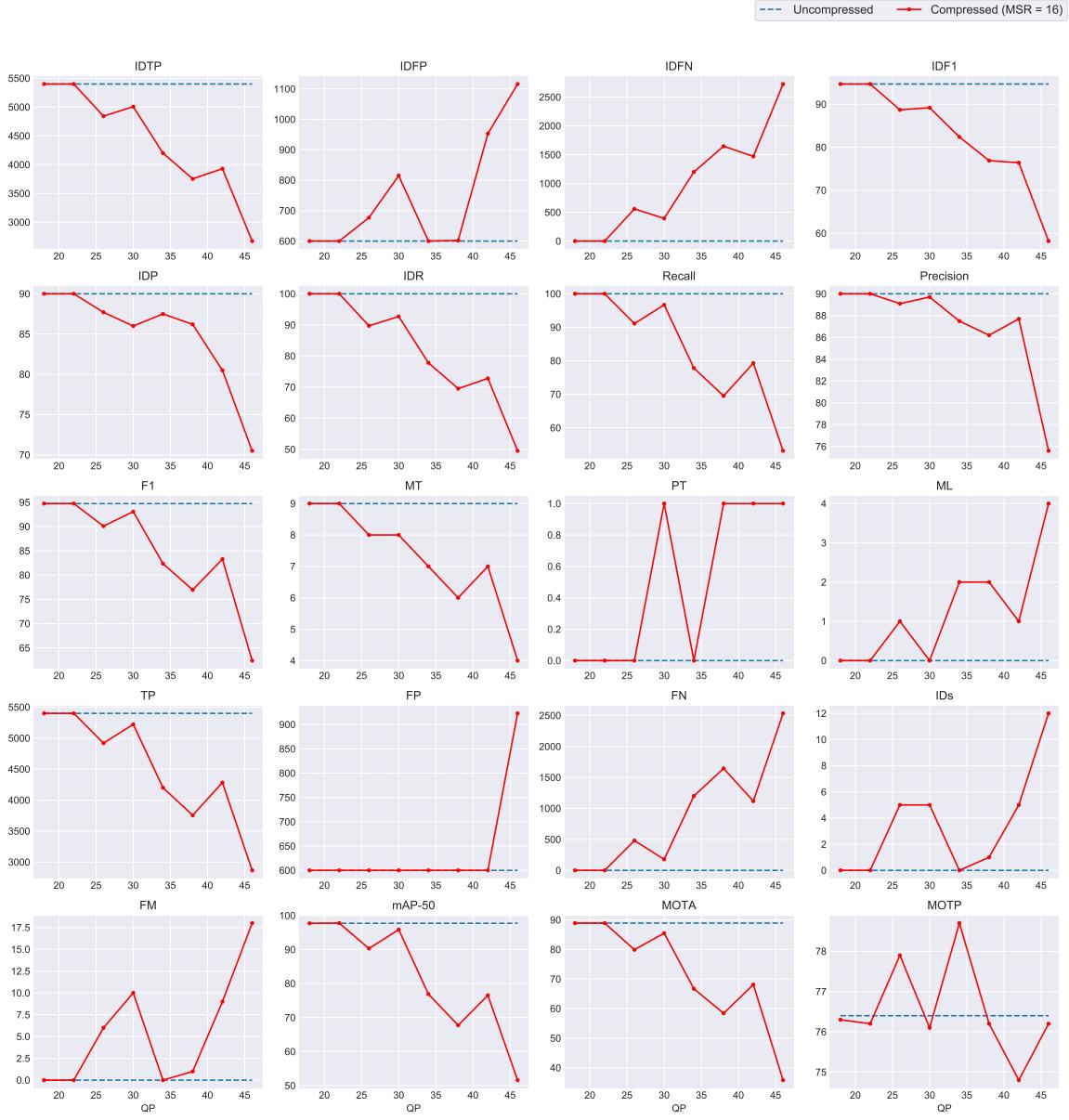
QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
Uncompressed	Uncompressed	5400.00	600.00	0.00	94.70	90.00	100.00	100.00	90.00	94.74	9	9	0	0	5400	600	0	0	0	97.72	88.90	76.40
18	16	5400.00	600.00	0.00	94.70	90.00	100.00	100.00	90.00	94.74	9	9	0	0	5400	600	0	0	0	97.70	88.90	76.30
22	16	5400.00	600.00	0.00	94.70	90.00	100.00	100.00	90.00	94.74	9	9	0	0	5400	600	0	0	0	97.76	88.90	76.20
26	16	4842.00	677.00	558.00	88.70	87.70	89.70	91.10	89.10	90.09	9	8	0	1	4919	600	481	5	6	90.29	79.90	77.90
30	16	5007.00	815.00	393.00	89.20	86.00	92.70	96.70	89.70	93.07	9	8	1	0	5222	600	178	5	10	95.85	85.50	76.10
34	16	4200.00	600.00	1200.00	82.40	87.50	77.80	77.80	87.50	82.37	9	7	0	2	4200	600	1200	0	0	76.89	66.70	78.70
38	16	3753.00	602.00	1647.00	76.90	86.20	69.50	86.50	86.20	76.95	9	6	1	2	3755	600	1645	1	1	67.73	58.40	76.20
42	16	3930.00	953.00	1470.00	76.40	80.50	72.80	79.30	87.70	83.29	9	7	1	1	4283	600	1117	5	9	76.54	68.10	74.80
46	16	2673.00	1116.00	2727.00	58.20	70.50	49.50	53.10	75.60	62.38	9	4	1	4	2866	923	2534	12	18	51.60	35.80	76.20

larger than the decrease of IDs and FP, so we conclude that MOTA decreases as QP increases based on Equation (2.24).

From Figure 4.9, we observed that IDP increases. As we confirmed that the number of occlusions decreases as QP increases, we can verify that IDFP will also decrease. Hence, based on the IDP equation (2.21), due to the drop of IDFP, IDP increases. This result can also be explained qualitatively as follows; as QP increases, the detected occlusion occurs less, so the incorrect ID assignments occur less, making ID precision higher at a higher QP.

#### 4.2.2 Johnny

The Johnny sequence consists of 9 “person” objects. Occlusion was not observed, and the objects scarcely move. Figure 4.12 shows a visualization of all the performance metrics at different QP, and Table 4.5 shows the corresponding numerical values. This result reveals that the most performance metrics decrease as QP increases similar to the case of BasketballPass. The decrease of performance can be verified from Figure 4.13, as the objects are starting to not be detected at higher QP values. The decrease of detections explains the increase of FN. ID switch did not occur in the uncompressed sequence but was observed in the higher QP in the compressed sequence. Therefore, unlike in BasketballPass there is an increase in IDs at higher QP. The increase of IDs can be explained qualitatively as follows; since there is no occlusion in this sequence, ID switch between

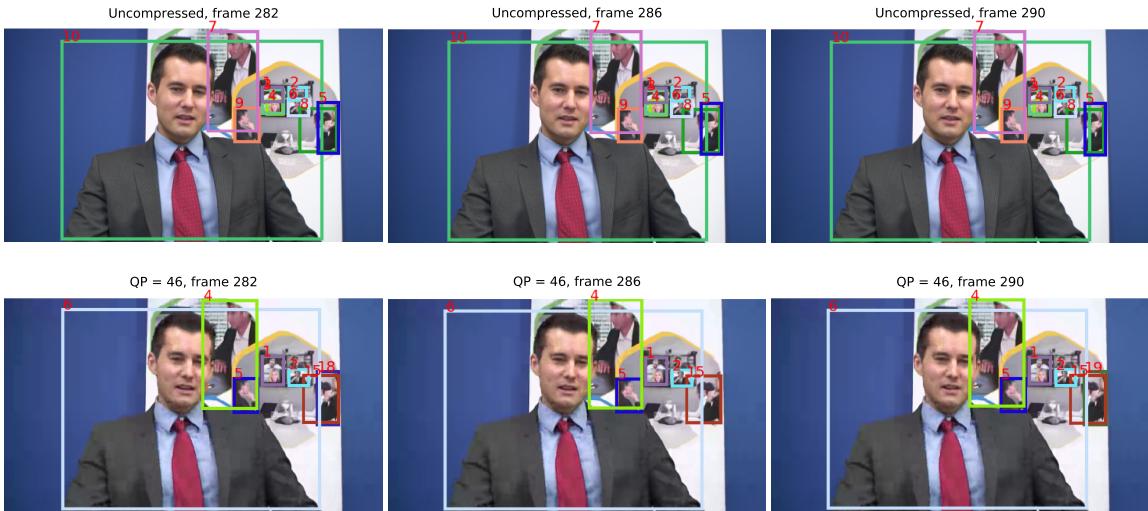


**Figure 4.12:** Visualization of the performance results on Johnny at different QP for the “person” object class.

the trajectories does not occur at lower QP, but starts to occur at high QP as the image quality drops and different identities are assigned to the same trajectory. As the image quality drops, the objects starts to not be detected in some sequence of frames. As a result of “gap” in the trajectory, in other words, after a certain number of frames in which objects are not detected, different object identities are assigned to the targets after objects are detected again. Therefore, IDs increases as QP increases. As an example from Figure 4.14, the identity of one “person” object did not change in the uncompressed frames 282 to 290, but changed in the compressed frames at QP = 46. Unlike the case of BasketballPass where the ID switch is caused by occlusion, ID switch in this sequence is caused by the “gap” in the trajectory, i.e., the detector fails to detect a particular object in a certain



**Figure 4.13:** Comparison of ground truth and tracking results on the Johnny sequence in frame 400 at different QP.



**Figure 4.14:** Comparison of Johnny frames 282, 286, 290 to explain IDs results.

number of frames due to the drop of image quality in the trajectory but detect again. Based on all the increases of FP, FN, and IDs, the general tracking performance metric of MOTA decreases as QP increases.

**Table 4.6:** Performance results on BlowingBubbles.

QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
Uncompressed	Uncompressed	753.00	373.00	279.00	69.80	66.90	73.00	94.10	86.30	90.03	3	2	0	1	971	154	61	2	5	89.71	79.00	90.00
18	16	874.00	240.00	158.00	81.50	78.50	84.70	95.70	88.80	92.12	3	2	0	1	988	125	44	1	2	91.53	83.50	89.90
22	16	986.00	105.00	46.00	92.90	90.40	95.50	95.40	90.40	92.83	3	2	0	1	985	105	47	0	3	91.72	85.30	90.00
26	16	986.00	46.00	86.00	93.80	92.00	95.50	95.40	92.00	93.67	3	2	0	1	985	86	47	0	3	92.46	87.10	90.10
30	16	991.00	49.00	41.00	95.70	95.30	96.00	95.90	95.30	95.60	3	2	0	1	990	49	42	0	2	93.64	91.20	90.10
34	16	949.00	59.00	83.00	93.10	94.10	92.00	94.50	96.80	95.64	3	2	0	1	975	32	57	1	4	93.62	91.30	89.90
38	16	929.00	89.00	103.00	90.70	91.30	90.00	95.10	96.50	95.79	3	2	0	1	981	36	51	2	2	94.24	91.40	89.70
42	16	675.00	350.00	357.00	65.70	65.90	65.40	91.70	92.40	92.05	3	2	0	1	946	78	86	5	6	91.87	83.60	88.60
46	16	340.00	98.00	692.00	46.30	77.60	32.90	42.00	99.10	59.00	3	0	2	1	433	4	599	11	13	53.32	40.50	86.40

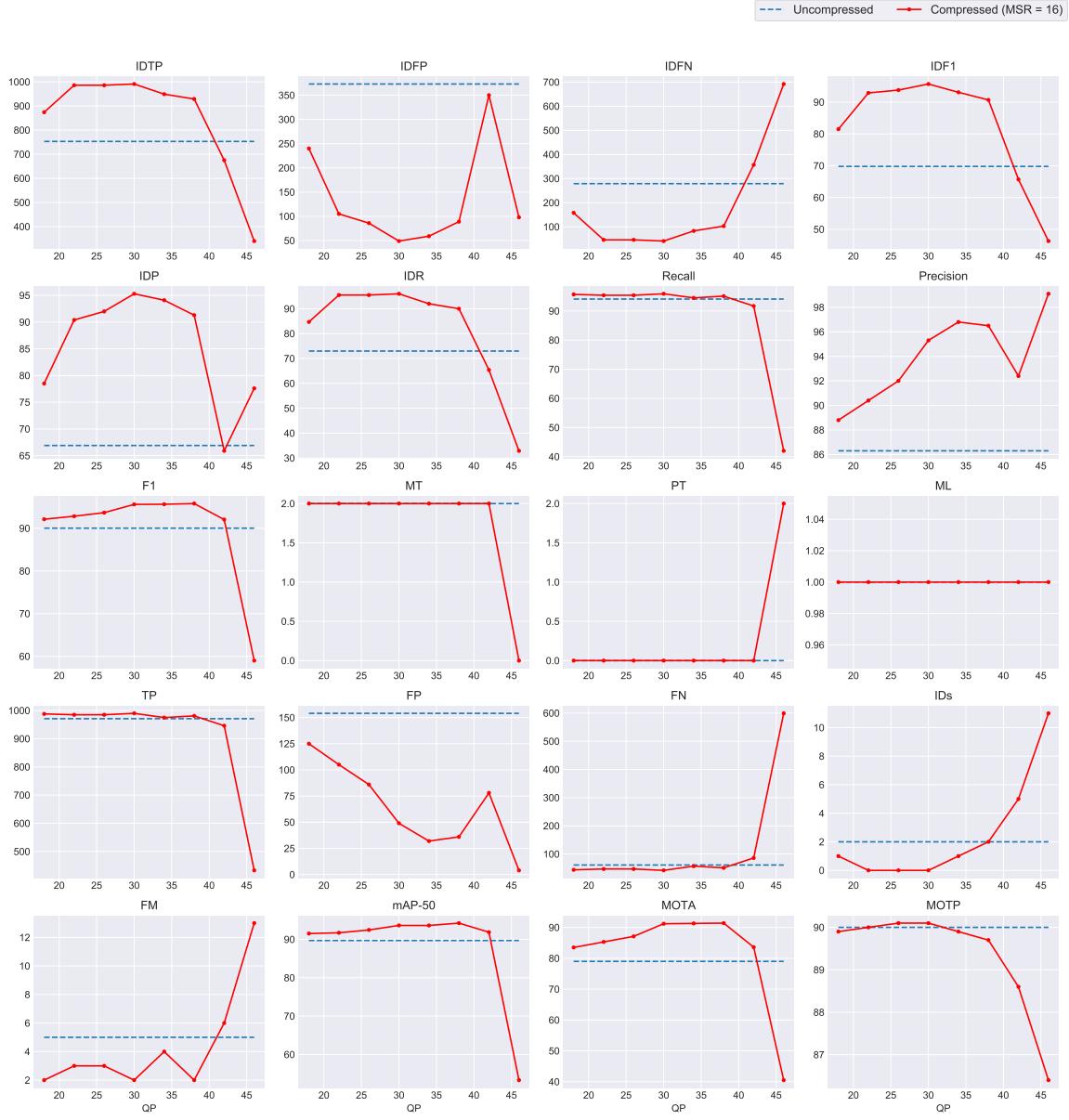
Since IDs increases, we expect IDP to decrease. This is because the more ID switch occurs, the lower ID precision should be, and indeed IDR also decreases. The correlation between IDs and IDP is high, which can also be seen in Appendix A.5. Therefore, we confirmed that ID performance drops as QP increases. We also observed that other metrics such as detection performance and track quality drop as QP increases. The results in this sequence are different from the results on BasketballPass because the occlusion does not occur, so ID precision decreases as image quality drops.

### 4.2.3 BlowingBubbles

We will now examine the sequence where we observed an increase of MOTA with the increase of QP, which can be confirmed in the visualization in Figure 4.15 and the numerical values in Table 4.6. The MOTA visualization plot shows that the score increases until QP = 38 but decreases thereafter. To analyze this outcome, we inspected the video carefully frame by frame. From Figure 4.16, which shows frame 465, we observed a toy, a humanlike object, was detected as “person” in the uncompressed case and at QP = 18, 22, 26, but not thereafter. This reveals that the detector incorrectly detects and recognizes this object as “person” at the lower QP, but as the image quality drops at the higher QP, the detector starts to not detect this particular object. Note that this toy object is not a part of the ground truth. This outcome was also observed in the other frames as well. This observation illustrates that as QP increases, FP may decrease, leading to increased MOTA. In this sequence, after QP = 38, the image quality further drops, and as shown in Figure 4.16, one “person” object is not detected at QP = 46, thereby FN increases. We observed that not only MOTA, the detection performance and ID measure also behave in the same way such that the performance scores increase up to the middle of the QP range and decrease thereafter. Though there is no occlusion in this sequence, we observed IDs increasing due to the drop of the image quality that causes “gap” in a trajectory similar to the outcome in Johnny.

### 4.2.4 Cactus

Here, we will present the analysis of tracking results in this sequence, where “potted plant” is the only object. Similar to BlowingBubbles, we observed an increase of MOTA and F1 in the middle of the QP range, as shown in Figure 4.17. Table 4.7 shows the corresponding numerical results. To explain this result, we inspected the video sequence frame by frame and compared various QP



**Figure 4.15:** Visualization of the performance results on BlowingBubbles at different QP for the “person” object class.

**Table 4.7:** Performance results on Cactus.

QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
Uncompressed	Uncompressed	120.00	232.00	380.00	28.20	34.10	24.00	60.20	85.50	70.65	1	0	1	0	301	51	199	7	11	67.96	48.60	88.20
18	16	119.00	252.00	381.00	27.30	32.10	23.80	60.00	80.90	68.90	1	0	1	0	300	71	200	7	12	68.21	44.40	88.10
22	16	105.00	236.00	395.00	25.00	30.80	21.00	55.00	80.60	65.38	1	0	1	0	275	66	225	6	14	67.81	40.60	87.90
26	16	131.00	243.00	369.00	30.00	35.00	26.20	63.20	84.50	72.31	1	0	1	0	316	58	184	7	14	70.53	50.20	87.60
30	16	62.00	274.00	438.00	14.80	18.50	12.40	60.60	90.20	72.49	1	0	1	0	303	33	197	9	16	70.83	52.20	87.30
34	16	110.00	249.00	390.00	25.60	30.60	22.00	64.20	89.40	74.73	1	0	1	0	321	38	179	8	13	73.29	55.00	87.20
38	16	114.00	232.00	386.00	27.00	32.90	22.80	65.40	94.50	77.30	1	0	1	0	327	19	173	8	17	78.77	60.00	86.80
42	16	89.00	201.00	411.00	22.50	30.70	17.80	49.80	85.90	63.05	1	0	1	0	249	41	251	10	19	49.37	39.60	86.10
46	16	70.00	190.00	430.00	18.40	26.90	14.00	29.20	56.20	38.43	1	0	1	0	146	114	354	6	12	20.74	5.20	83.60

cases at frame 32, as shown in Figure 4.18.

From this comparison, we observed that YOLOv3

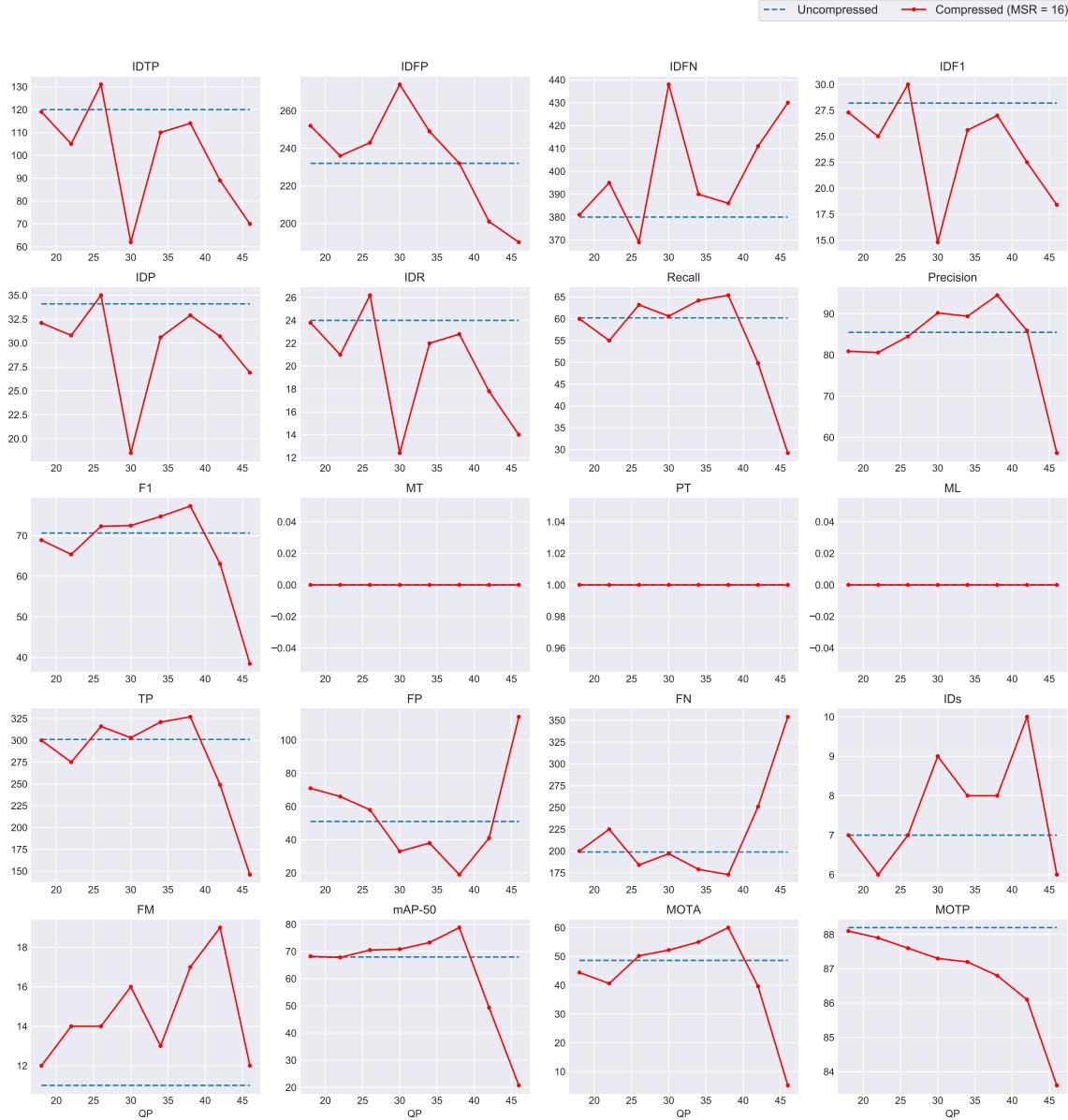


**Figure 4.16:** Comparison of ground truth and tracking results on the BlowingBubbles sequence in frame at 465 at different QP.

incorrectly detect two objects at a single target until QP = 26. Ground truth only contains one “potted plant” object. At QP = 34, we can see the correctly detected object, but as QP increases further, the object is no longer detected and at QP = 46, YOLOv3 detected a wrong target. This illustrates the decrease of FP midway through the QP range and its increase again at high QP. It is also observed that FN decreases slightly at the middle of the QP range but increases at high QP, since the correct target is no longer detected. In fact, not only at frame 32, but we also observed similar outcomes in other frames. This reveals that the cause of an increase of MOTA midway through the QP range is due to the decrease of FP and FN. One reason for the outcome we observed at frame 32 could be that the type of object is problematic because there are three cactuses in a vase, so YOLOv3 incorrectly detect one of the cactuses as a “potted plant” instead of detecting all three cactuses in a vase as a “potted plant”.

#### 4.2.5 Discussion

We inspected the four individual video sequences to gain the better insight into the general tracking performance measured by MOTA. From the inspected video samples, we found that MOTA scores in BasketballPass and Johnny are consistent with the expectation that tracking performance decreases as QP increases. However, the MOTA scores in BlowingBubbles and Cactus increase midway



**Figure 4.17:** Visualization of the performance results on Cactus at different QP for the “potted plant” object class.

through the QP range and decrease thereafter. We found that these results are due to the type of object that makes the YOLOv3 detector incorrectly detect wrong targets, and FP was observed to be high even in the uncompressed video. Recall that we are using SORT based on YOLOv3, so the tracking performance is dependent on the detection performance. Note that since FP was high in the uncompressed sequence, as QP increases and the image quality drops, FP was observed to decrease. This insight also explains why we observed an increase in Precision in the average results. From Equation (2.15), Precision depends on TP and FP, and since FP decreases in some video samples, Precision was observed to increase.



**Figure 4.18:** Comparison of ground truth and tracking results on the Cactus sequence in frame 32 at different QP.

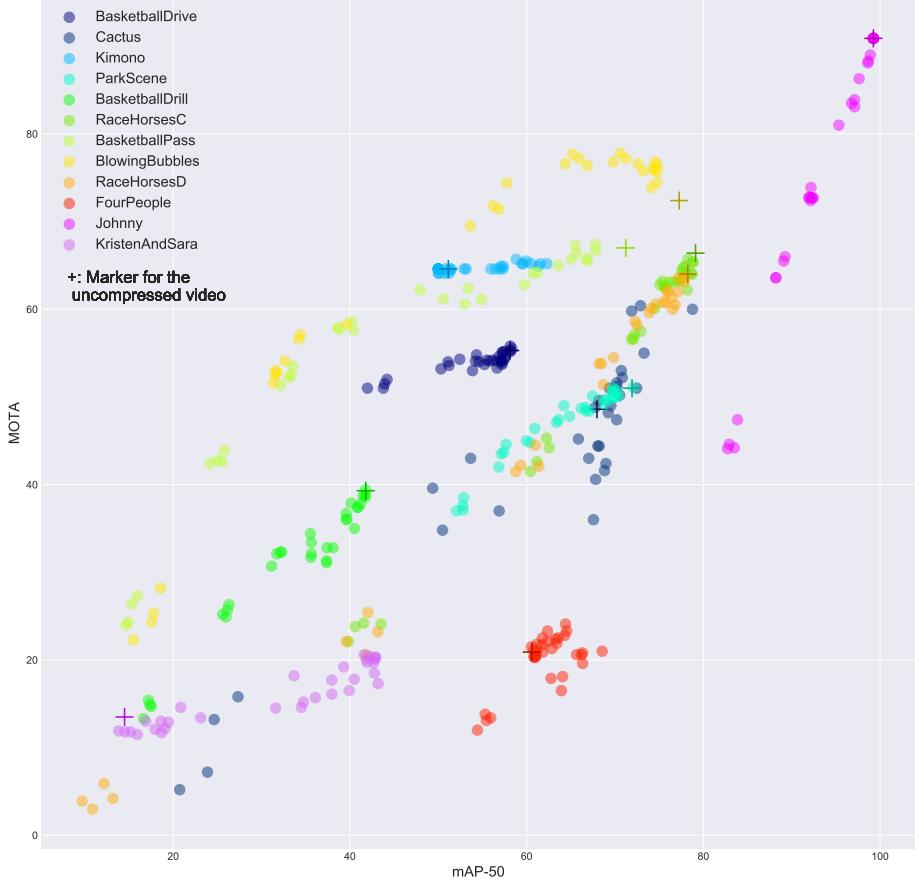
Not only does the tracking performance depend on the quality of the detector, but it also depends on the quality of the tracker. Re-ID for the long-term undetection and occlusion is not implemented in our tracker, SORT, so it affects the ID measure and IDs.

Finally, for the MOTP metric, we observed an increase of MOTP with increasing QP in some sequences: BasketballPass, Johnny, and BlowingBubbles. Since MOTP measures how well the objects are localized relative to the ground truth, the results from individual sequences tell us that object localization improves midway through the QP range, but decreases thereafter as the image quality further drops. Not all the video sequences behave in this way, but these outcomes contributed to the average results.

### 4.3 Impact of Detection on Tracking Performance

To analyze how much the detection performance impacts the tracking performance, MOTA is plotted against mAP-50 across all video sequences for “all” object classes as shown in Figure 4.19.

This figure illustrates each scatter plot of video sequence colored differently, and “+” indicates uncompressed sequences. It is trivial from this scatter plot that the detection performance impacts the tracking performance because the higher the mAP-50, the higher the MOTA. As we are analyzing the relation between the two variables MOTA and mAP-50, we computed the Pearson correlation

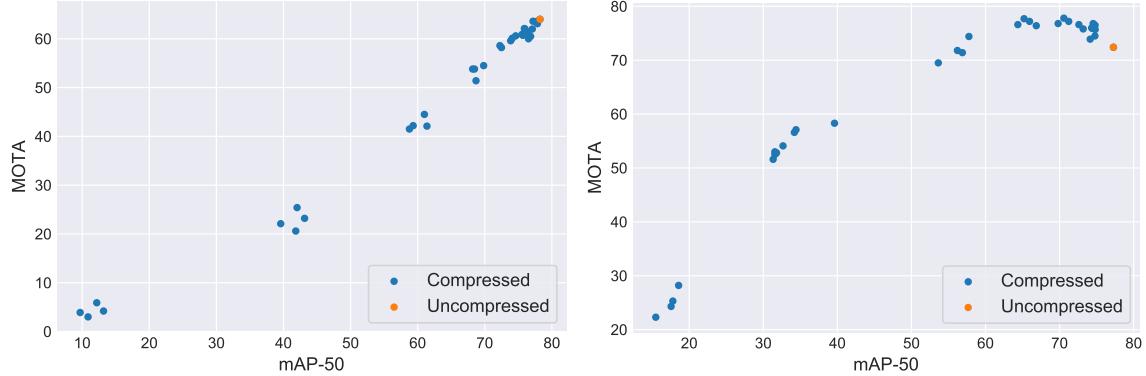


**Figure 4.19:** Scatter plot of MOTA vs. mAP-50 across all video sequence.

coefficient and obtained as 0.703. This value explains that the two variables mAP-50 and MOTA are positively related. To examine more carefully for each video sequence, from Figure 4.19, we can see that not all the sequences look linear. For example, RaceHorsesD shows a linear relationship but BlowingBubbles shows a non-linear relationship between MOTA and mAP-50, as shown in Figure 4.20. Similar cases for the non-linear outcome can be observed in other sequences as well. The non-linear cases of result indicate that some factors contribute to the concave down of the curve, i.e., the decrease of MOTA growth rate as mAP-50 increases, as one of the non-linear outcomes shown in Figure 4.20b. Although we tested the parameters QP and MSR, there are other possible factors that we did not experiment with, for example, size of the bounding boxes and parameters in the Kalman filter framework. A more thorough analysis will need to be conducted to identify such factors as a future study.

## 4.4 Summary

We presented the average tracking results of 12 video sequences for “all” object classes. Regression analysis was conducted to quantify the impact of QP and MSR on MOTA and propose the



(a) MOTA vs. mAP-50 scatter plot for RaceHorsesD      (b) MOTA vs. mAP-50 scatter plot for BlowingBubbles

**Figure 4.20:** Linear and non-linear cases of scatter plots of MOTA and mAP-50.

relationship between MOTA and QP. One-sided t-test was conducted to identify the specific QP at which performance on compressed sequences is lower than on uncompressed sequences with 95% confidence. The results indicate that most metrics except Precision and MOTP are consistent with our expectation. To achieve a better understanding of the average results, we inspected four individual sequences. From each video, we saw that as the image quality drops, the detector starts to not detect the target, showing the increase of FN. We also found that there are cases that make certain performance scores increase as the image quality decreases. This is because the tracking performance depends on the detector's performance. Our detector YOLOv3 sometimes incorrectly detects wrong targets, but as the image quality drops, the frequency of wrong detections decreases. We also learned that for videos with frequent occlusions, IDs decreases due to fewer detections and therefore less detected occlusions as the image quality drops. However, for videos with fewer occlusions, IDs increases due to “gap” in trajectories, and different identities would be assigned to the trajectories. Note that because SORT does not have a re-ID feature for long-term occlusions or undetection, we observed such IDs result. Finally, the impact of detection performance on tracking performance was examined. From the scatter plots of MOTA and mAP-50, we observed a linear relationship in some videos. However, some videos show a non-linear relationship such that as mAP-50 increases, MOTA increases but the rate of increase of MOTA decreases. A more thorough analysis will need to be conducted to understand this outcome.

# Chapter 5

## Conclusion

In this work, we studied the impact of video compression on multiple object tracking performance. We started with annotating the ground truth that is suitable for object tracking, conducted the experiment by running the HEVC codec at different QP and MSR on the uncompressed video sequences, and ran the multiple object tracker on the decoded sequences. Analyzing the results from the compressed and uncompressed video sequences, the visualization of average results across 12 video sequences shows that as QP increases, detection and tracking performance decrease, which is consistent with the expectation, except for the Precision and MOTP scores. Applying the regression analysis to all the video sequences, we learned that MSR does not impact the MOTA score but QP does with 95% confidence based on the available data. We also proposed a relationship between MOTA and QP. One-sided t-test was also conducted to determine the specific QP at which tracking performance on compressed sequences is lower than on uncompressed sequences.

By analyzing further on four individual sequences as case studies, we illustrated that FN increases as QP increases due to less detection as the image quality drops. Since FN is significantly larger than the change in FP and IDs, we observed a decrease of MOTA. However, we found that MOTA sometimes increases midway through the QP range in some sequences. This is because in certain cases, YOLOv3 detects wrong objects, leading to a higher FP than the case on uncompressed video sequences. As the image quality drops, the frequency of incorrect detection decreases; hence, FP decreases and MOTA increases. However, as the image quality drops further, MOTA decreases due to the significant increase of FN. The reason the Precision score increased in the average results is also due to the decrease of FP.

Occlusion is inevitable in some video sequences. We found that IDs decreases in the video sequences with frequent occlusions due to the less detected occlusions, as QP increases. For sequences with fewer occlusions, IDs increases as QP increases due to the “gap” in trajectories caused by the drop of image quality. For MOTP, the increase of score midway through the QP range was observed. From the analysis of four individual sequences, we observed such increases in some sequences, which indicate that as the image quality drops, the object localization improves in some sequences, but further analysis is required to identify its cause.

These case studies explained that the results are due to the design of YOLOv3 and SORT. Tracking performance is dependent on the design of the tracker. In our case, we are using a detection-based tracker; therefore, the tracking performance is dependent on the detection performance. Analyzing how detection impacts tracking performance, MOTA and mAP-50 are positively correlated, and linear and non-linear relationships of MOTA and mAP were observed. A more thorough analysis will be necessary to identify the reason for the non-linear outcome, which is a decrease of MOTA growth rate as mAP-50 increases. The possible factors to consider could be size of bounding boxes and parameters in the Kalman filter framework in SORT.

We employed a simple and effective approach to detecting and tracking multiple objects using YOLOv3 and SORT. However, since there are better detectors and trackers, we could run the experiment with a better multiple object tracker. For example, for a detector, Ultralytics developed YOLOv5 [50], achieving better performance than YOLOv3. Since SORT does not have a feature to perform re-ID for the long-term occlusions and undetection, we could improve the tracker by implementing a re-ID feature. Finally, since QP and MSR are not the only configuration parameters in the video compression standard of HEVC, we could explore and examine further with different configurations.

# References

- [1] D. M. Gavrila and V. Philomin, "Real-time object detection for "smart" vehicles", in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 1, Sep. 1999, 87–93 vol.1. doi: [10.1109/ICCV.1999.791202](https://doi.org/10.1109/ICCV.1999.791202).
- [2] W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu, and T.-K. Kim, "Multiple object tracking: A literature review", *Artificial Intelligence*, vol. 293, p. 103448, 2021, issn: 0004-3702. doi: <https://doi.org/10.1016/j.artint.2020.103448>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0004370220301958>.
- [3] L. Lin, Y. Lu, C. Li, H. Cheng, and W. Zuo, "Detection-free multiobject tracking by reconfigurable inference with bundle representations", *IEEE Transactions on Cybernetics*, vol. 46, no. 11, pp. 2447–2458, 2016, issn: 2168-2275. doi: [10.1109/TCYB.2015.2478515](https://doi.org/10.1109/TCYB.2015.2478515).
- [4] Z. Zou, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey", *arXiv:1905.05055 [cs]*, May 15, 2019. arXiv: 1905.05055. [Online]. Available: <http://arxiv.org/abs/1905.05055> (visited on 02/05/2021).
- [5] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 580–587. [Online]. Available: [https://openaccess.thecvf.com/content\\_cvpr\\_2014/html/Girshick\\_Rich\\_Feature\\_Hierarchies\\_2014\\_CVPR\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2014/html/Girshick_Rich_Feature_Hierarchies_2014_CVPR_paper.html) (visited on 05/18/2021).
- [6] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015. doi: [10.1109/TPAMI.2015.2389824](https://doi.org/10.1109/TPAMI.2015.2389824).
- [7] R. Girshick, "Fast r-CNN", presented at the Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1440–1448. [Online]. Available: [https://openaccess.thecvf.com/content\\_iccv\\_2015/html/Girshick\\_Fast\\_R-CNN\\_ICCV\\_2015\\_paper.html](https://openaccess.thecvf.com/content_iccv_2015/html/Girshick_Fast_R-CNN_ICCV_2015_paper.html) (visited on 05/18/2021).
- [8] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-CNN: Towards real-time object detection with region proposal networks", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017. doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031).
- [9] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection", presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2117–2125. [Online]. Available: [https://openaccess.thecvf.com/content\\_cvpr\\_2017/html/Lin\\_Feature\\_Pyramid\\_Networks\\_CVPR\\_2017\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2017/html/Lin_Feature_Pyramid_Networks_CVPR_2017_paper.html) (visited on 05/18/2021).

- [10] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask r-CNN”, presented at the Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2961–2969. [Online]. Available: [https://openaccess.thecvf.com/content\\_iccv\\_2017/html/He\\_Mask\\_R-CNN\\_ICCV\\_2017\\_paper.html](https://openaccess.thecvf.com/content_iccv_2017/html/He_Mask_R-CNN_ICCV_2017_paper.html) (visited on 05/18/2021).
- [11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection”, presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 779–788. [Online]. Available: [https://www.cv-foundation.org/openaccess/content\\_cvpr\\_2016/html/Redmon\\_You\\_Only\\_Look\\_CVPR\\_2016\\_paper.html](https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Redmon_You_Only_Look_CVPR_2016_paper.html) (visited on 03/01/2021).
- [12] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “SSD: Single shot MultiBox detector”, in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., ser. Lecture Notes in Computer Science, Cham: Springer International Publishing, 2016, pp. 21–37, ISBN: 978-3-319-46448-0. doi: [10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2).
- [13] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, “Focal loss for dense object detection”, presented at the Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2980–2988. [Online]. Available: [https://openaccess.thecvf.com/content\\_iccv\\_2017/html/Lin\\_Focal\\_Loss\\_for\\_ICCV\\_2017\\_paper.html](https://openaccess.thecvf.com/content_iccv_2017/html/Lin_Focal_Loss_for_ICCV_2017_paper.html) (visited on 05/18/2021).
- [14] J. Redmon and A. Farhadi, “YOLOv3: An incremental improvement”, 2018.
- [15] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, “Simple online and realtime tracking”, in *2016 IEEE International Conference on Image Processing (ICIP)*, ISSN: 2381-8549, Sep. 2016, pp. 3464–3468. doi: [10.1109/ICIP.2016.7533003](https://doi.org/10.1109/ICIP.2016.7533003).
- [16] M. Jacobs and J. Probell, “A brief history of video coding”, 2009.
- [17] T. Zhang and S. Mao, “An overview of emerging video coding standards”, *GetMobile: Mobile Comp. and Comm.*, vol. 22, no. 4, pp. 13–20, 2019, Place: New York, NY, USA Publisher: Association for Computing Machinery, ISSN: 2375-0529. doi: [10.1145/3325867.3325873](https://doi.org/10.1145/3325867.3325873). [Online]. Available: <https://doi.org/10.1145/3325867.3325873>.
- [18] J. Redmon and A. Farhadi, “YOLO9000: Better, faster, stronger”, presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7263–7271. [Online]. Available: [https://openaccess.thecvf.com/content\\_cvpr\\_2017/html/Redmon\\_YOLO9000\\_Better\\_Faster\\_CVPR\\_2017\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2017/html/Redmon_YOLO9000_Better_Faster_CVPR_2017_paper.html) (visited on 03/04/2021).
- [19] G. Jocher, Y. Kwon, guigarfr, perry0418, J. Veitch-Michaelis, Ttayu, D. Suess, F. Baltacı, G. Bianconi, IlyaOvodov, Marc, e96031413, C. Lee, D. Kendall, Falak, F. Reveriano, FuLin, GoogleWiki, J. Nataprawira, J. Hu, LinCoce, LukeAI, NanoCode012, NirZarrabi, O. Reda, P. Skalski, SergioSanchezMontesUAM, S. Song, T. Havlik, and T. M. Shead, “Ultralytics/yolov3: V9.5.0 - YOLOv5 v5.0 release compatibility update for YOLOv3”, version v9.5.0, Apr. 2021. doi: [10.5281/zenodo.4681234](https://doi.org/10.5281/zenodo.4681234). [Online]. Available: <https://doi.org/10.5281/zenodo.4681234>.
- [20] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft COCO: Common objects in context”, in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., Cham: Springer International Publishing, 2014, pp. 740–755, ISBN: 978-3-319-10602-1.
- [21] (). “COCO - common objects in context”, [Online]. Available: <https://cocodataset.org/#detection-eval> (visited on 05/09/2021).

- [22] R. E. Kalman *et al.*, “A new approach to linear filtering and prediction problems”, *Journal of basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.
- [23] X. Li, K. Wang, W. Wang, and Y. Li, “A multiple object tracking method using kalman filter”, in *The 2010 IEEE International Conference on Information and Automation*, Jun. 2010, pp. 1862–1866. doi: [10.1109/ICINFA.2010.5512258](https://doi.org/10.1109/ICINFA.2010.5512258).
- [24] abewley, “Abewley/sort”, original-date: 2016-02-03T03:16:23Z, Apr. 9, 2021. [Online]. Available: <https://github.com/abewley/sort> (visited on 04/09/2021).
- [25] R. Labbe, “Rlabbe/filterpy”, original-date: 2014-07-15T02:15:19Z, Jun. 24, 2021. [Online]. Available: <https://github.com/rlabbe/filterpy> (visited on 06/24/2021).
- [26] H. W. Kuhn and H. W. Kuhn, “The hungarian method for the assignment problem”, *Naval research logistics quarterly.*, vol. 2, no. 1, pp. 83–97, 1955, Place: Arlington, Va. : Washington, D.C. : Publisher: Office of Naval Research, ISSN: 00281441.
- [27] G. J. Sullivan, J. Ohm, W. Han, and T. Wiegand, “Overview of the high efficiency video coding (HEVC) standard”, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012, ISSN: 1558-2205. doi: [10.1109/TCSVT.2012.2221191](https://doi.org/10.1109/TCSVT.2012.2221191).
- [28] M. Budagavi, A. Fuldseth, and G. Bjøntegaard, “HEVC transform and quantization”, in, 2014, pp. 141–169, ISBN: 978-3-319-06894-7. doi: [10.1007/978-3-319-06895-4\\_6](https://doi.org/10.1007/978-3-319-06895-4_6).
- [29] Y. Sharab and N. Sarhan, “Modeling and analysis of power consumption in live video streaming systems”, *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 13, pp. 11–12, 2017. doi: [10.1145/3115505](https://doi.org/10.1145/3115505).
- [30] C. Lou, S. Lee, and C. J. Kuo, “Adaptive motion search range prediction for video encoding”, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 12, pp. 1903–1908, Dec. 2010, ISSN: 1558-2205. doi: [10.1109/TCSVT.2010.2087551](https://doi.org/10.1109/TCSVT.2010.2087551).
- [31] S. Bachu and K. M. Chari, “A review on motion estimation in video compression”, in *2015 International Conference on Signal Processing and Communication Engineering Systems*, Jan. 2015, pp. 250–256. doi: [10.1109/SPACES.2015.7058259](https://doi.org/10.1109/SPACES.2015.7058259).
- [32] C. Heindl, “Cheind/py-motmetrics”, original-date: 2017-04-07T15:16:59Z, Mar. 24, 2021. [Online]. Available: <https://github.com/cheind/py-motmetrics> (visited on 03/25/2021).
- [33] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, “Performance measures and a data set for multi-target, multi-camera tracking”, in *Computer Vision – ECCV 2016 Workshops*, G. Hua and H. Jégou, Eds., Cham: Springer International Publishing, 2016, pp. 17–35, ISBN: 978-3-319-48881-3.
- [34] A. Milan, L. Leal-Taixé, I. D. Reid, S. Roth, and K. Schindler, “MOT16: A benchmark for multi-object tracking”, *CoRR*, vol. abs/1603.00831, 2016, \_eprint: 1603.00831. [Online]. Available: <http://arxiv.org/abs/1603.00831>.
- [35] R. Padilla, S. L. Netto, and E. A. B. da Silva, “A survey on performance metrics for object-detection algorithms”, in *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, ISSN: 2157-8702, Jul. 2020, pp. 237–242. doi: [10.1109/IWSSIP48289.2020.9145130](https://doi.org/10.1109/IWSSIP48289.2020.9145130).

- [36] R. Padilla, W. L. Passos, T. L. B. Dias, S. L. Netto, and E. A. B. da Silva, “A comparative analysis of object detection metrics with a companion open-source toolkit”, *Electronics*, vol. 10, no. 3, 2021, ISSN: 2079-9292. doi: [10.3390/electronics10030279](https://doi.org/10.3390/electronics10030279). [Online]. Available: <https://www.mdpi.com/2079-9292/10/3/279>.
- [37] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, “The pascal visual object classes challenge: A retrospective”, *International Journal of Computer Vision*, vol. 111, no. 1, pp. 98–136, Jan. 1, 2015, ISSN: 1573-1405. doi: [10.1007/s11263-014-0733-5](https://doi.org/10.1007/s11263-014-0733-5). [Online]. Available: <https://doi.org/10.1007/s11263-014-0733-5>.
- [38] Cartucho, “Cartucho/mAP”, original-date: 2018-03-08T11:24:23Z, Jun. 27, 2021. [Online]. Available: <https://github.com/Cartucho/mAP> (visited on 06/28/2021).
- [39] J. Cartucho, R. Ventura, and M. Veloso, “Robust object recognition through symbiotic deep learning in mobile robots”, in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, ISSN: 2153-0866, Oct. 2018, pp. 2336–2341. doi: [10.1109/IROS.2018.8594067](https://doi.org/10.1109/IROS.2018.8594067).
- [40] K. Bernardin and R. Stiefelhagen, “Evaluating multiple object tracking performance: The CLEAR MOT metrics”, *EURASIP Journal on Image and Video Processing*, vol. 2008, 2008. doi: [10.1155/2008/246309](https://doi.org/10.1155/2008/246309).
- [41] L. Leal-Taixé, A. Milan, K. Schindler, D. Cremers, I. D. Reid, and S. Roth, “Tracking the trackers: An analysis of the state of the art in multiple object tracking”, *CoRR*, vol. abs/1704.02781, 2017, \_eprint: 1704.02781. [Online]. Available: <http://arxiv.org/abs/1704.02781>.
- [42] F. Bossen *et al.*, “Common test conditions and software reference configurations”, *JCTVC-L1100*, vol. 12, no. 7, 2013.
- [43] Alexey, “AlexeyAB/yolo\_mark”, original-date: 2016-12-17T21:25:07Z, Jun. 28, 2021. [Online]. Available: [https://github.com/AlexeyAB/Yolo\\_mark](https://github.com/AlexeyAB/Yolo_mark) (visited on 06/28/2021).
- [44] H. Choi, E. Hosseini, S. R. Alvar, R. A. Cohen, and I. V. Bajić, “A dataset of labelled objects on raw video sequences”, *Data in Brief*, vol. 34, p. 106701, 2021, ISSN: 2352-3409. doi: <https://doi.org/10.1016/j.dib.2020.106701>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2352340920315808>.
- [45] F. Zhao, Q. Huang, and W. Gao, “Image matching by normalized cross-correlation”, in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, ISSN: 2379-190X, vol. 2, May 2006, pp. II–II. doi: [10.1109/ICASSP.2006.1660446](https://doi.org/10.1109/ICASSP.2006.1660446).
- [46] L. Leal-Taixé, A. Milan, I. D. Reid, S. Roth, and K. Schindler, “MOTChallenge 2015: Towards a benchmark for multi-target tracking”, *CoRR*, vol. abs/1504.01942, 2015, \_eprint: 1504.01942. [Online]. Available: <http://arxiv.org/abs/1504.01942>.
- [47] M. Kutner, “Applied linear statistical models”, 5th ed. Boston: McGraw-Hill Irwin, 2005, ISBN: 0-07-238688-6.
- [48] S. Seabold and J. Perktold, “Statsmodels: Econometric and statistical modeling with python”, in *9th Python in Science Conference*, 2010.

- [49] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, “SciPy 1.0: Fundamental algorithms for scientific computing in python”, *Nature Methods*, vol. 17, pp. 261–272, 2020. doi: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2).
- [50] G. Jocher, A. Stoken, J. Borovec, NanoCode012, A. Chaurasia, TaoXie, L. Changyu, A. V, Laughing, tkianai, yxNONG, A. Hogan, lorenzomammana, AlexWang1900, J. Hajek, L. Diaconu, Marc, Y. Kwon, oleg, wanghaoyang0106, Y. Defretin, A. Lohia, ml5ah, B. Milanko, B. Fineran, D. Khromov, D. Yiwei, Doug, Durgesh, and F. Ingham, “Ultralytics/yolov5: V5.0 - YOLOv5-p6 1280 models, AWS, supervise.ly and YouTube integrations”, version v5.0, Apr. 2021. doi: [10.5281/zenodo.4679653](https://doi.org/10.5281/zenodo.4679653). [Online]. Available: <https://doi.org/10.5281/zenodo.4679653>.

# Appendix A

## Additional Information

### A.1 Justification of SORT Optimization

We optimized SORT by maximizing MOTA on the training video sequence of PartyScene as explained in Section 3.2. Since MOTA is not the only metric in MOT metrics, we will justify that other scores will achieve good performance at max age of 1, min hits of 5, and IOU threshold of 0.4. Table A.1 shows the results of all the performance metrics at different min hits. Note that we fixed max age of 1 as suggested by [24] and IOU threshold of 0.4 based on MOTA result.

**Table A.1:** Justification of SORT Optimization on “person” and “all” object classes.

max_age	min_hits	iou_thres	IDF1	IDP	IDR	Rcell	Prcn	F1	GT	MT	PT	ML	FP	FN	IDs	FM	MOTA	MOTP
1	1	0.40	80.70	81.40	80.10	95.50	97.20	96.34	3	3	0	0	21	34	6	9	92.00	86.60
1	2	0.40	81.40	83.20	79.80	94.60	98.60	96.56	3	3	0	0	10	41	6	8	92.50	86.70
1	3	0.40	81.80	84.30	79.50	93.60	99.20	96.32	3	2	1	0	6	49	6	6	92.00	86.80
1	4	0.40	82.10	85.10	79.30	92.80	99.60	96.08	3	2	1	0	3	55	4	4	91.90	86.80
<b>1</b>	<b>5</b>	<b>0.40</b>	<b>82.20</b>	<b>85.50</b>	<b>79.10</b>	<b>92.30</b>	<b>99.70</b>	<b>95.86</b>	<b>3</b>	<b>2</b>	<b>1</b>	<b>0</b>	<b>2</b>	<b>59</b>	<b>4</b>	<b>4</b>	<b>91.50</b>	<b>86.80</b>
1	6	0.40	82.40	86.00	79.00	91.70	99.90	95.62	3	2	1	0	1	63	4	4	91.10	86.90
1	7	0.40	82.50	86.50	78.90	91.20	100.00	95.40	3	2	1	0	0	67	3	3	90.80	86.90
1	8	0.40	82.50	86.70	78.70	90.80	100.00	95.18	3	2	1	0	0	70	3	3	90.40	86.90
1	9	0.40	82.60	86.90	78.60	90.40	100.00	94.96	3	1	2	0	0	73	3	3	90.00	86.90

(a) Tuning parameters on “person” object class.

max_age	min_hits	iou_thres	IDF1	IDP	IDR	Rcell	Prcn	F1	GT	MT	PT	ML	FP	FN	IDs	FM	MOTA	MOTP
1	1	0.40	70.00	67.80	72.40	86.70	81.10	83.81	7	5	2	0	430	285	20	30	65.60	85.80
1	2	0.40	71.10	69.90	72.30	85.50	82.70	84.08	7	5	2	0	382	309	18	24	66.80	85.90
1	3	0.40	71.80	71.50	72.10	84.60	83.80	84.20	7	5	2	0	350	330	13	17	67.60	85.90
1	4	0.40	72.40	72.80	72.00	83.90	84.80	84.35	7	4	3	0	321	344	10	14	68.40	86.00
<b>1</b>	<b>5</b>	<b>0.40</b>	<b>72.90</b>	<b>73.90</b>	<b>71.90</b>	<b>83.40</b>	<b>85.80</b>	<b>84.58</b>	<b>7</b>	<b>4</b>	<b>3</b>	<b>0</b>	<b>295</b>	<b>355</b>	<b>10</b>	<b>13</b>	<b>69.10</b>	<b>86.00</b>
1	6	0.40	73.40	75.10	71.80	83.00	86.80	84.86	7	4	3	0	270	364	11	12	69.80	86.10
1	7	0.40	73.90	76.20	71.70	82.50	87.70	85.02	7	4	3	0	247	373	10	12	70.50	86.10
1	8	0.40	74.30	77.10	71.60	82.10	88.50	85.18	7	4	3	0	228	382	10	12	71.00	86.10
1	9	0.40	74.60	78.00	71.50	81.70	89.10	85.24	7	4	3	0	213	391	10	12	71.30	86.10

(b) Tuning parameters on “all” object classes.

## A.2 Standard Deviation of Average Result

We showed the averaged result of 12 video sequences in Section 4.1. Due to the relatively small number of video samples and each video has a different resolution, frame rate, number of objects, and object classes, each performance result is different. Therefore, the standard deviations are large. We show the standard deviation table in this section.

**Table A.2:** Standard deviation of performance results across all video sequences.

QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Rcll	Prcn	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
Uncompressed	Uncompressed	2064.78	920.95	3363.60	21.39	19.91	23.64	20.12	10.80	16.29	7.70	3.52	3.72	5.42	2050.74	731.69	3253.93	28.13	36.08	21.60	21.79	5.32

**(a) Standard Deviation for the Uncompressed Sequence**

QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Rcll	Prcn	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	8	2060.04	979.45	3377.44	22.75	21.09	24.90	19.98	11.66	16.32	7.70	3.61	3.43	5.30	2058.07	778.56	3241.03	27.89	36.76	20.95	22.21	5.25
22	8	2090.58	933.34	3335.78	22.71	20.92	24.90	19.95	11.46	16.29	7.70	3.55	3.61	5.32	2072.84	756.57	3219.89	27.76	35.89	19.98	22.36	5.18
26	8	2030.44	843.94	3236.65	21.82	20.84	23.07	18.51	11.00	15.54	7.70	3.47	3.46	5.51	1994.64	714.86	3185.43	27.07	35.28	20.14	20.68	5.11
30	8	1980.42	973.34	3268.86	20.86	20.46	22.09	18.64	10.38	15.17	7.70	3.70	3.34	5.18	2085.29	770.80	3132.23	26.29	33.79	17.35	20.91	5.40
34	8	1879.30	868.16	3230.65	19.17	20.30	18.72	15.42	10.75	13.68	7.70	3.50	3.49	5.27	1884.81	702.41	3113.31	24.99	33.42	16.27	17.89	5.03
38	8	1702.59	757.26	3232.29	17.56	20.63	16.08	14.58	9.68	13.15	7.70	3.12	3.38	5.28	1707.19	594.40	3227.73	25.31	33.33	17.86	15.75	4.73
42	8	1706.42	726.48	3431.96	17.76	18.58	17.50	17.53	9.54	15.68	7.70	3.23	3.19	5.19	1798.04	547.93	3299.26	22.50	29.19	18.93	17.55	5.02
46	8	1392.69	859.76	3567.74	16.97	17.51	15.36	18.60	12.29	19.96	7.70	2.09	2.47	5.53	1530.73	732.53	3438.87	16.26	24.77	21.35	18.38	4.33

**(b) Standard Deviation at MSR = 8**

QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Rcll	Prcn	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	16	2011.62	1027.53	3447.42	22.45	20.88	24.53	20.15	11.59	16.48	7.70	3.61	3.65	5.28	2055.22	768.69	3246.40	28.47	36.29	21.07	22.45	5.28
22	16	2113.02	934.99	3310.19	23.44	21.89	24.55	20.20	11.64	16.63	7.70	3.55	3.61	5.42	2063.34	786.66	3233.91	28.68	36.85	20.19	22.93	5.11
26	16	1977.59	886.72	3304.55	21.98	21.12	23.09	18.54	10.99	15.57	7.70	3.52	3.78	5.21	1993.54	694.27	3178.99	27.47	34.43	19.55	20.66	4.98
30	16	1991.49	997.00	3336.38	23.54	23.96	24.26	18.50	10.96	15.35	7.70	3.61	3.92	5.28	2067.08	745.75	3142.65	26.21	33.11	16.94	21.30	5.46
34	16	1822.60	933.65	3313.62	19.61	20.81	19.14	15.28	10.60	13.68	7.70	3.35	3.43	5.24	1883.27	693.22	3115.29	26.03	34.02	16.65	17.93	5.05
38	16	1674.77	814.20	3399.62	18.00	20.17	16.87	14.47	9.65	13.09	7.70	3.09	3.29	5.32	1726.97	600.13	3232.38	25.45	32.75	18.01	15.83	4.79
42	16	1699.11	782.77	3465.90	17.88	18.58	17.77	17.71	10.17	15.78	7.70	2.91	2.86	5.38	1781.06	629.23	3354.80	22.51	29.91	18.72	17.94	4.75
46	16	1385.56	796.19	3581.80	16.58	19.46	15.00	18.15	14.03	19.15	7.70	2.23	2.33	5.76	1514.79	636.69	3450.12	17.41	24.10	21.62	18.67	4.22

**(c) Standard Deviation at MSR = 16**

QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Rcll	Prcn	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	32	2043.60	978.25	3401.33	23.06	21.33	25.21	20.14	11.93	16.49	7.70	3.55	3.72	5.37	2052.97	772.83	3251.43	28.59	35.81	20.85	22.60	5.20
22	32	2120.42	901.69	3298.39	23.13	21.77	25.10	19.87	11.67	16.22	7.70	3.55	3.70	5.21	2077.58	742.03	3204.20	28.72	35.76	19.12	22.54	5.11
26	32	1996.29	974.00	3345.79	21.15	19.81	22.76	19.07	11.20	15.89	7.70	3.45	3.48	5.41	2017.05	773.88	3200.13	27.29	35.99	19.87	21.35	5.06
30	32	1928.22	979.63	3326.39	22.31	22.40	22.85	17.73	10.88	15.02	7.70	3.41	3.42	5.26	2004.73	736.84	3133.16	26.69	34.16	17.04	20.50	5.26
34	32	1825.76	900.55	3318.31	21.58	22.86	20.77	15.35	10.10	13.65	7.70	3.52	3.49	5.24	1890.05	641.81	3105.59	25.81	33.10	16.39	17.81	5.05
38	32	1685.45	781.39	3360.86	18.07	20.98	16.68	14.66	9.34	13.01	7.70	3.09	3.29	5.32	1727.57	581.72	3225.25	23.83	32.00	18.34	15.73	5.08
42	32	1648.36	734.24	3389.41	15.87	16.67	15.80	17.39	10.31	15.50	7.70	3.05	3.15	5.41	1790.08	590.90	3320.16	21.98	30.56	18.58	17.76	4.74
46	32	1349.58	853.75	3656.11	16.71	16.56	15.46	18.52	12.12	19.61	7.70	2.30	2.91	5.79	1524.22	654.05	3456.88	16.90	23.08	21.24	17.82	4.00

**(d) Standard Deviation at MSR = 32**

QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Rcll	Prcn	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	64	2046.04	982.64	3395.07	21.92	20.27	24.08	20.05	11.46	16.41	7.70	3.52	3.46	5.48	2059.74	767.82	3239.35	28.20	36.91	20.65	22.29	5.29
22	64	2065.76	985.02	3377.98	23.29	21.89	25.24	19.86	11.62	16.18	7.70	3.63	3.65	5.35	2072.95	771.13	3214.79	28.75	37.09	19.94	22.30	5.36
26	64	1973.12	860.48	3209.07	20.43	20.00	21.36	18.20	11.00	15.46	7.70	3.50	3.62	5.14	1962.76	739.29	3167.65	27.64	35.15	18.93	20.74	4.98
30	64	2017.15	914.00	3239.18	21.74	21.74	22.71	18.74	11.22	15.66	7.70	3.68	3.42	5.32	2060.76	748.33	3164.25	27.02	33.51	18.05	21.70	5.27
34	64	1819.18	913.01	3317.98	21.57	22.88	20.80	15.38	10.22	13.66	7.70	3.35	3.55	5.27	1884.95	663.61	3108.10	23.93	33.79	16.43	17.90	5.18
38	64	1669.71	764.32	3358.91	17.47	19.86	16.16	13.86	9.62	12.50	7.70	2.91	3.34	5.28	1698.78	583.67	3238.10	23.42	30.96	17.65	14.96	4.74
42	64	1621.58	730.11	3428.76	16.23	18.02	16.08	16.95	9.65	15.01	7.70	2.95	3.06	5.37	1764.65	572.65	3350.99	19.69	29.51	18.59	17.48	5.03
46	64	1288.47	945.93	3654.56	16.93	20.13	15.05	18.78	13.81	19.93	7.70	2.49	2.75	5.74	1544.48	696.43	3445.22	18.40	25.99	21.35	18.86	4.08

**(e) Standard Deviation at MSR = 64**

### A.3 Regression Analysis of All Metrics

As we focused on the analysis of the MOTA score, we showed its result in Section 4.1.2. In this appendix section, the regression analysis result for all the metrics are included in Table A.3. The regression analysis is applied on all the raw data points for “all” object classes across all the video sequences. The equivalent transformations of QP and each performance score, and the hypotheses can be established. The value less than the significance level of 0.05 is bolded in the table.

**Table A.3:** Regression analysis result of all metrics for “all” object classes across all video sequences.

	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
$\beta_0$	-173.56	-79.55	-173.56	-5.58	-1.67	-6.14	-5.71	1.09	-3.76	-0.54	<b>0.04</b>	-0.59	-162.88	-61.11	-162.88	-2.09	-1.90	-6.23	-4.32	-0.22
$\beta_1$	-100.23	-48.83	-100.23	-2.79	-0.49	-3.31	-3.7	0.66	-2.50	-0.40	<b>0.05</b>	-0.45	-107.25	-28.81	-107.25	-1.05	-1.11	-3.82	-2.97	-0.07
$\beta_2$	0.11	-0.17	0.11	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>	< $ 10^{-3} $	< $ 10^{-2} $	< $ 10^{-2} $	< $ 10^{-3} $	< $ 10^{-3} $	< $ 10^{-3} $	-0.25	<b>0.04</b>	-0.25	<b>-0.01</b>	<b>-0.02</b>	< $ 10^{-3} $	< $ 10^{-2} $	< $ 10^{-3} $
$\beta_3$	<b>-0.60</b>	<b>&lt; <math> 10^{-3} </math></b>	<b>-0.04</b>	< $ 10^{-2} $	< $ 10^{-3} $	< $ 10^{-3} $	< $ 10^{-3} $	< $ 10^{-2} $	< $ 10^{-3} $	< $ 10^{-3} $	< $ 10^{-3} $	< $ 10^{-3} $	-0.11	< $ 10^{-2} $	< $ 10^{-2} $	< $ 10^{-3} $	< $ 10^{-2} $	< $ 10^{-3} $	< $ 10^{-2} $	< $ 10^{-3} $
p-value( $\beta_0$ )	< $ 10^{-5} $	<b>0.01</b>	< $ 10^{-5} $	< $ 10^{-6} $	0.10	< $ 10^{-7} $	< $ 10^{-12} $	<b>0.04</b>	< $ 10^{-8} $	< $ 10^{-4} $	0.79	< $ 10^{-3} $	< $ 10^{-9} $	< $ 10^{-3} $	< $ 10^{-9} $	<b>0.01</b>	0.10	< $ 10^{-5} $	< $ 10^{-6} $	0.08
p-value( $\beta_1$ )	< $ 10^{-6} $	< $ 10^{-4} $	< $ 10^{-6} $	< $ 10^{-4} $	0.10	< $ 10^{-10} $	< $ 10^{-16} $	< $ 10^{-3} $	< $ 10^{-10} $	< $ 10^{-12} $	0.45	< $ 10^{-7} $	< $ 10^{-20} $	< $ 10^{-4} $	< $ 10^{-10} $	<b>0.02</b>	< $ 10^{-7} $	< $ 10^{-9} $	0.32	
p-value( $\beta_2$ )	0.92	0.84	0.92	0.64	0.72	0.71	0.90	0.82	0.77	0.89	0.82	0.85	0.72	0.93	0.72	0.56	0.47	0.96	0.83	0.59
p-value( $\beta_3$ )	0.95	1.00	0.95	0.72	0.73	0.81	0.89	0.51	0.77	0.83	0.76	0.90	0.80	0.98	0.80	0.53	0.71	0.97	0.77	0.64

## A.4 One-sided t-test for “person” object class

In Section 4.1.3, we showed the result of one-sided t-test for “all” object classes across all the video sequences. In this appendix section, we included the result for “person” object class across all video sequences. Although the QP at which the score is significantly greater or less than that of the uncompressed case is different, the similar conclusions to “all” object classes can be made in “person” object class. The result shows, for example, the MOTA score at QP = 34 is lower than the score at the uncompressed sequences with 95% confidence. Similar to “all” object classes, the metrics Precision and PT shows the result that is not expected according to Section 2.5. The value less than the significance level of 0.05 is bolded in the table.

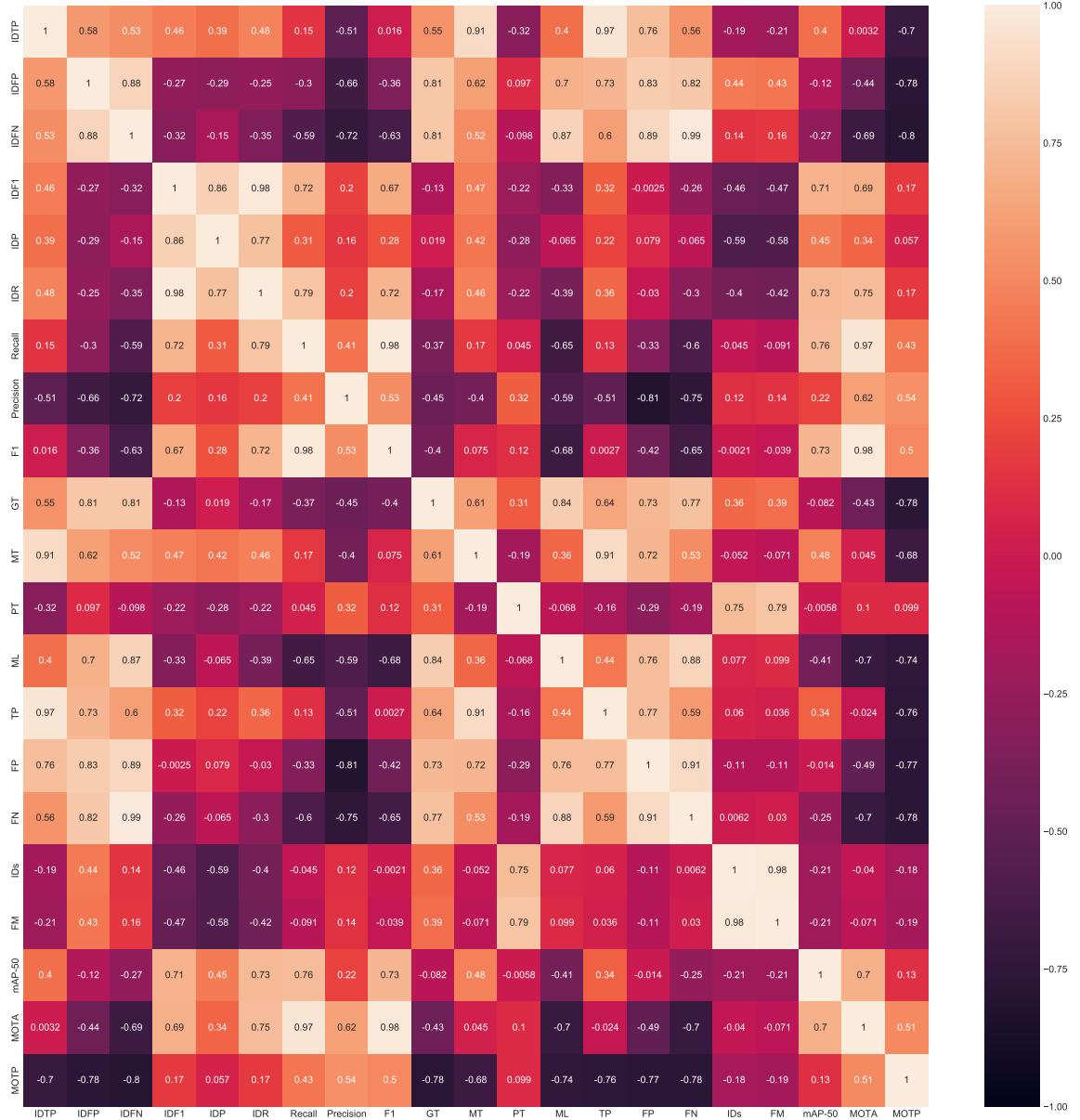
**Table A.4:** One-sided t-test for “person” object class across all the video sequences.

	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
p-value(QP=18)	0.91	0.77	0.91	0.97	0.88	0.98	0.97	0.09	0.81	<b>0.05</b>	0.31	0.13	0.97	0.99	0.97	0.77	0.31	0.11	0.69	0.98
p-value(QP=22)	0.98	0.31	0.98	0.96	0.92	0.96	0.52	0.30	0.77	<b>0.03</b>	0.50	<b>0.01</b>	0.91	0.96	0.91	0.64	0.59	<b>0.04</b>	0.63	0.98
p-value(QP=26)	0.32	0.05	0.32	0.84	0.96	0.59	<  10 <sup>-3</sup>	0.94	0.08	<  10 <sup>-3</sup>	0.19	<  10 <sup>-3</sup>	<b>0.04</b>	0.21	<b>0.04</b>	0.55	0.38	<  10 <sup>-3</sup>	0.11	1.00
p-value(QP=30)	0.22	0.46	0.22	0.78	0.90	0.55	<  10 <sup>-4</sup>	0.94	0.07	<  10 <sup>-2</sup>	0.09	<  10 <sup>-2</sup>	0.07	0.44	0.07	0.82	0.82	<  10 <sup>-3</sup>	0.11	1.00
p-value(QP=34)	<  10 <sup>-2</sup>	<  10 <sup>-2</sup>	<  10 <sup>-2</sup>	0.25	0.78	<b>0.04</b>	<  10 <sup>-5</sup>	0.98	<  10 <sup>-3</sup>	<  10 <sup>-4</sup>	0.15	<  10 <sup>-4</sup>	<  10 <sup>-2</sup>	<b>0.03</b>	<  10 <sup>-2</sup>	0.08	0.10	<  10 <sup>-4</sup>	<  10 <sup>-2</sup>	1.00
p-value(QP=38)	<  10 <sup>-3</sup>	<  10 <sup>-4</sup>	<  10 <sup>-3</sup>	<  10 <sup>-2</sup>	0.57	<  10 <sup>-3</sup>	<  10 <sup>-6</sup>	0.99	<  10 <sup>-6</sup>	<  10 <sup>-6</sup>	<b>0.01</b>	<  10 <sup>-4</sup>	<  10 <sup>-4</sup>	<  10 <sup>-4</sup>	<  10 <sup>-3</sup>	0.86	0.24	<  10 <sup>-5</sup>	<  10 <sup>-4</sup>	0.82
p-value(QP=42)	<  10 <sup>-5</sup>	0.09	<  10 <sup>-5</sup>	<  10 <sup>-5</sup>	<b>0.02</b>	<  10 <sup>-6</sup>	<  10 <sup>-8</sup>	0.90	<  10 <sup>-6</sup>	<  10 <sup>-9</sup>	<  10 <sup>-3</sup>	<  10 <sup>-4</sup>	<  10 <sup>-8</sup>	<  10 <sup>-8</sup>	0.78	0.93	<  10 <sup>-7</sup>	<  10 <sup>-7</sup>	0.16	
p-value(QP=46)	<  10 <sup>-6</sup>	<b>0.02</b>	<  10 <sup>-6</sup>	<  10 <sup>-2</sup>	<  10 <sup>-9</sup>	<  10 <sup>-12</sup>	<b>0.03</b>	<  10 <sup>-11</sup>	<  10 <sup>-16</sup>	<  10 <sup>-4</sup>	<  10 <sup>-7</sup>	<  10 <sup>-9</sup>	0.27	<  10 <sup>-9</sup>	0.78	0.75	<  10 <sup>-11</sup>	<  10 <sup>-11</sup>	0.10	

## A.5 Correlation of Performance Metrics

We examined the video sequences with each performance metric, and some metrics are defined in terms of other metrics according to the definitions listed in Section 2.4. We did not explicitly describe which metrics correlate or relate to other metrics except mAP-50 and MOTA in Section 4.3 but included the following figure, the potentially useful information to explain the correlation between different performance metrics. Figure A.1 shows the heatmap of Pearson correlation coefficients between different metrics. The result is based on 12 video sequences for “all” object class.

The potential useful insights are, for example, MOTA is defined in terms of FN, FP, IDs, and GT. From the correlation heatmap, excluding the constant GT, FN correlates the most compared to FP and IDs. This explains why the increase of FN is significantly larger than the change of FP and IDs so that MOTA decreases. FP still correlates significantly with MOTA but IDs correlates little. The p-value of correlation between MOTA and IDs is 0.43, therefore IDs does not significantly correlate with MOTA. For IDs, FM is correlated the most. This is logical since the ID switch occurs when the object gets undetected and re-detected, while FM occurs when the object state changes from being detected to undetected, losing on the track. IDP is also negatively correlated with IDs.



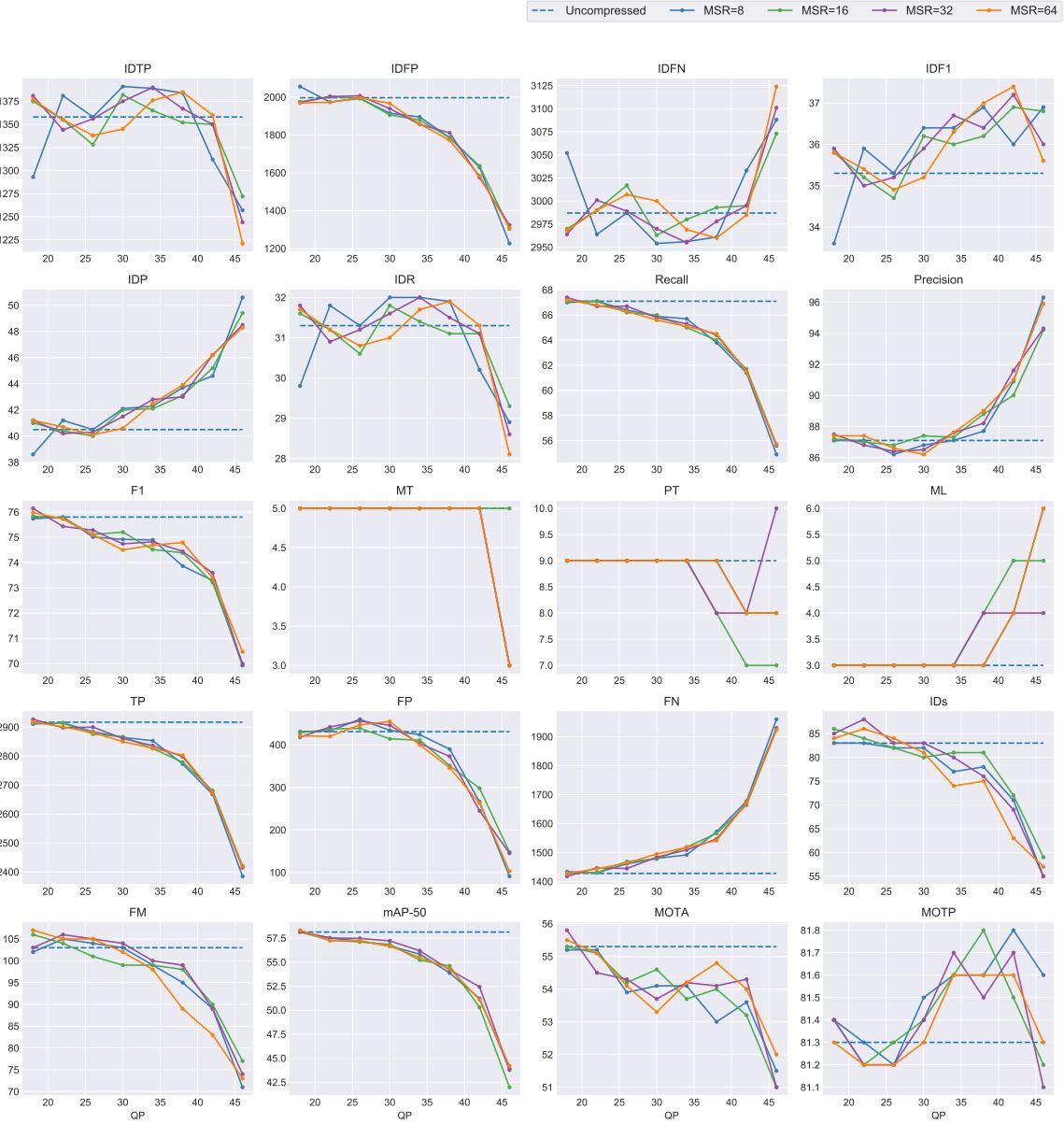
**Figure A.1:** Pearson correlation heatmap of performance metrics for “all” object classes.

## **Appendix B**

# **All Results of Individual Sequences**

We showed the averaged result in Section 3.1 but will show each individual result in this appendix. There are 12 video sequences, showing performance metrics with different QP and MSR as well as tabular results. The results are based on “all” object classes, showing all MSR values.

## B.1 BasketballDrive



**Figure B.1:** Visualization of performance results on BasketballDrive at different QP.

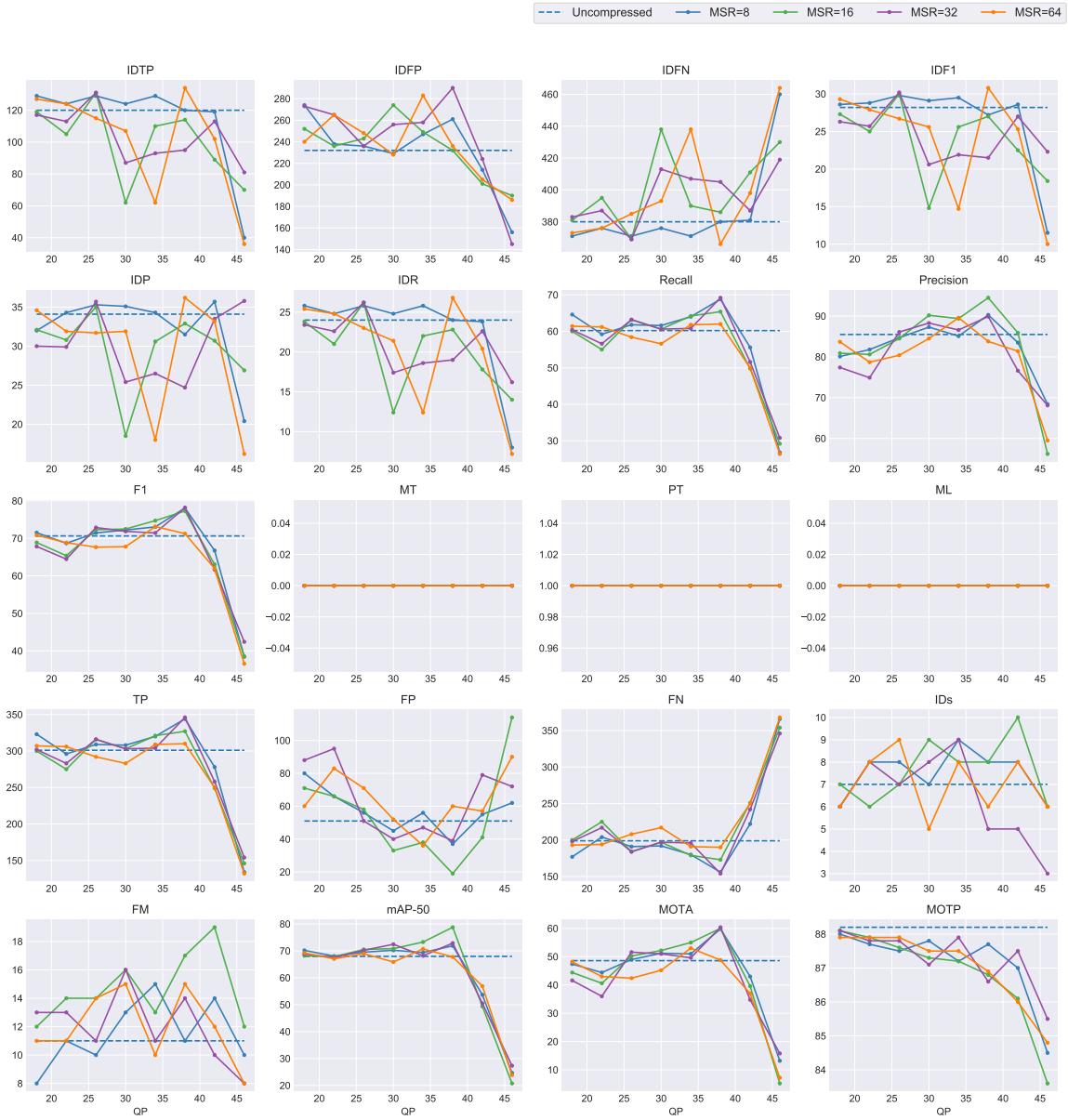


**Figure B.2:** Visualization of performance results on BasketballDrive at different MSR.

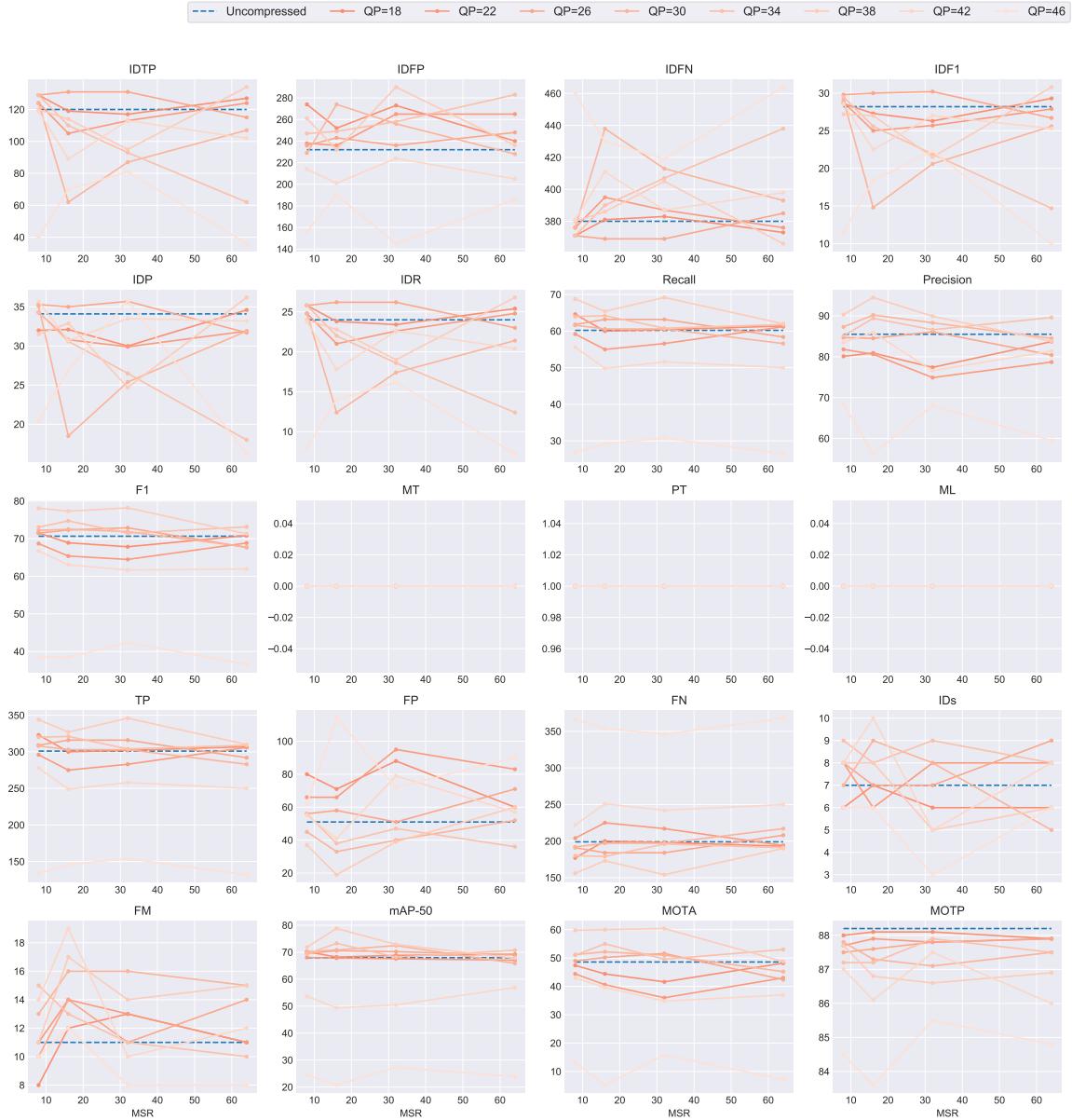
**Table B.1:** Performance results on BasketballDrive.

QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
Uncompressed	Uncompressed	1358.00	1997.00	2987.00	35.30	40.50	31.30	67.10	87.10	75.80	17	5	9	3	2917	431	1428	83	103	58.13	55.30	81.30
<b>(a) Uncompressed Sequence</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	8	1293.00	2056.00	3052.00	33.60	38.60	29.80	67.00	87.10	75.74	17	5	9	3	2911	431	1434	83	102	58.20	55.20	81.40
22	8	1381.00	1974.00	2964.00	35.90	41.20	31.80	67.10	87.10	75.80	17	5	9	3	2915	433	1430	83	105	57.51	55.20	81.30
26	8	1358.00	1993.00	2987.00	35.30	40.50	31.30	66.40	86.20	75.02	17	5	9	3	2884	460	1461	82	104	57.20	53.90	81.20
30	8	1391.00	1915.00	2954.00	36.40	42.10	32.00	65.90	86.80	74.92	17	5	9	3	2865	434	1480	82	103	56.67	54.10	81.50
34	8	1389.00	1895.00	2956.00	36.40	42.30	32.00	65.70	87.10	74.90	17	5	9	3	2853	424	1492	77	99	55.84	54.10	81.60
38	8	1384.00	1786.00	2961.00	36.90	43.70	31.90	63.80	87.70	73.86	17	5	9	3	2773	390	1572	78	95	53.87	53.00	81.60
42	8	1312.00	1629.00	3033.00	36.00	44.60	30.20	61.40	90.90	73.29	17	5	8	4	2668	266	1677	71	89	51.22	53.60	81.80
46	8	1257.00	1226.00	3088.00	36.90	50.60	28.90	54.90	96.30	69.93	17	3	8	6	2385	91	1960	55	71	43.93	51.50	81.60
<b>(b) MSR = 8</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	16	1375.00	1975.00	2970.00	35.80	41.00	31.60	67.10	87.20	75.84	17	5	9	3	2915	428	1430	86	106	58.14	55.30	81.40
22	16	1355.00	2003.00	2990.00	35.20	40.40	31.20	67.10	87.00	75.77	17	5	9	3	2914	437	1431	84	104	57.29	55.10	81.20
26	16	1328.00	1994.00	3017.00	34.70	40.00	30.60	66.20	86.80	75.11	17	5	9	3	2876	439	1469	82	101	57.10	54.20	81.30
30	16	1382.00	1906.00	2963.00	36.20	42.00	31.80	66.00	87.40	75.21	17	5	9	3	2867	414	1478	80	99	56.85	54.60	81.40
34	16	1365.00	1879.00	2980.00	36.00	42.10	31.40	65.00	87.30	74.52	17	5	9	3	2826	411	1519	81	99	55.22	53.70	81.60
38	16	1352.00	1785.00	2993.00	36.20	43.10	31.10	64.00	88.80	74.39	17	5	8	4	2779	351	1566	81	98	54.60	54.00	81.80
42	16	1350.00	1636.00	2995.00	36.90	45.20	31.10	61.70	90.00	73.21	17	5	7	5	2681	298	1664	72	90	50.30	53.20	81.50
46	16	1272.00	1304.00	3073.00	36.80	49.40	29.30	55.70	94.20	70.01	17	5	7	5	2421	148	1924	59	77	41.98	51.00	81.20
<b>(c) MSR = 16</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	32	1381.00	1971.00	2964.00	35.90	41.20	31.80	67.40	87.50	76.15	17	5	9	3	2927	418	1418	85	103	58.15	55.80	81.40
22	32	1344.00	2003.00	3001.00	35.00	40.20	30.90	66.70	86.80	75.43	17	5	9	3	2898	442	1447	88	106	57.51	54.50	81.20
26	32	1356.00	2007.00	2989.00	35.20	40.30	31.20	66.70	86.40	75.28	17	5	9	3	2900	456	1445	83	105	57.47	54.30	81.20
30	32	1375.00	1939.00	2970.00	35.90	41.50	31.60	65.80	86.50	74.74	17	5	9	3	2861	446	1484	83	104	57.21	53.70	81.40
34	32	1390.00	1856.00	2955.00	36.70	42.80	32.00	65.30	87.60	74.82	17	5	9	3	2836	403	1509	80	100	56.19	54.20	81.70
38	32	1367.00	1811.00	2978.00	36.40	43.00	31.50	64.40	88.20	74.44	17	5	8	4	2798	373	1547	76	99	54.20	54.10	81.50
42	32	1350.00	1575.00	2995.00	37.20	46.20	31.10	61.50	91.60	73.59	17	5	8	4	2673	245	1672	69	89	52.43	54.30	81.70
46	32	1244.00	1323.00	3101.00	36.00	48.50	28.60	55.60	94.30	69.95	17	3	10	4	2415	145	1930	55	74	43.78	51.00	81.10
<b>(d) MSR = 32</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	64	1377.00	1969.00	2968.00	35.80	41.20	31.70	67.20	87.40	75.98	17	5	9	3	2918	421	1427	84	107	58.32	55.50	81.30
22	64	1355.00	1973.00	2990.00	35.40	40.70	31.20	66.80	87.40	75.72	17	5	9	3	2901	420	1444	86	105	57.24	55.10	81.20
26	64	1338.00	1997.00	3007.00	34.90	40.10	30.80	66.30	86.60	75.10	17	5	9	3	2881	447	1464	84	105	57.24	54.10	81.20
30	64	1345.00	1967.00	3000.00	35.20	40.60	31.00	65.60	86.20	74.50	17	5	9	3	2850	455	1495	81	102	56.66	53.30	81.30
34	64	1376.00	1858.00	2969.00	36.30	42.50	31.70	65.10	87.60	74.69	17	5	9	3	2827	400	1518	74	98	55.49	54.20	81.60
38	64	1385.00	1771.00	2960.00	37.00	43.90	31.90	64.50	89.00	74.79	17	5	9	3	2803	346	1542	75	89	54.33	54.80	81.60
42	64	1360.00	1584.00	2985.00	37.40	46.20	31.30	61.50	91.00	73.40	17	5	8	4	2674	263	1671	63	83	51.10	54.00	81.60
46	64	1221.00	1309.00	3124.00	35.60	48.30	28.10	55.70	95.90	70.47	17	3	8	6	2420	103	1925	57	73	44.20	52.00	81.30
<b>(e) MSR = 64</b>																						

## B.2 Cactus



**Figure B.3:** Visualization of performance results on Cactus at different QP.



**Figure B.4:** Visualization of performance results on Cactus at different MSR.

**Table B.2:** Performance results on Cactus.

QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
Uncompressed	Uncompressed	120.00	232.00	380.00	28.20	34.10	24.00	60.20	85.50	70.65	1	0	1	0	301	51	199	7	11	67.96	48.60	88.20

(a) Uncompressed Sequence

QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	8	129.00	274.00	371.00	28.60	32.00	25.80	64.60	80.10	71.52	1	0	1	0	323	80	177	6	8	70.19	47.40	88.00
22	8	124.00	238.00	376.00	28.80	34.30	24.80	59.20	81.80	68.69	1	0	1	0	296	66	204	8	11	68.06	44.40	87.70
26	8	129.00	236.00	371.00	29.80	35.30	25.80	61.80	84.70	71.46	1	0	1	0	309	56	191	8	10	69.53	49.00	87.50
30	8	124.00	229.00	376.00	29.10	35.10	24.80	61.60	87.30	72.23	1	0	1	0	308	45	192	7	13	70.21	51.20	87.80
34	8	129.00	247.00	371.00	29.50	34.30	25.80	64.00	85.10	73.06	1	0	1	0	320	56	180	9	15	69.38	51.00	87.20
38	8	120.00	261.00	380.00	27.20	31.50	24.00	68.80	90.30	78.10	1	0	1	0	344	37	156	8	11	71.89	59.80	87.70
42	8	119.00	214.00	381.00	28.60	35.70	23.80	55.60	83.50	66.75	1	0	1	0	278	55	222	8	14	53.66	43.00	87.00
46	8	40.00	156.00	460.00	11.50	20.40	8.00	26.80	68.40	38.51	1	0	1	0	134	62	366	6	10	24.64	13.20	84.50

(b) MSR = 8

QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	16	119.00	252.00	381.00	27.30	32.10	23.80	60.00	80.90	68.90	1	0	1	0	300	71	200	7	12	68.21	44.40	88.10
22	16	105.00	236.00	395.00	25.00	30.80	21.00	55.00	80.60	65.38	1	0	1	0	275	66	225	6	14	67.81	40.60	87.90
26	16	131.00	243.00	369.00	30.00	35.00	26.20	63.20	84.50	72.31	1	0	1	0	316	58	184	7	14	70.53	50.20	87.60
30	16	62.00	274.00	438.00	14.80	18.50	12.40	60.60	90.20	72.49	1	0	1	0	303	33	197	9	16	70.83	52.20	87.30
34	16	110.00	249.00	390.00	25.60	30.60	22.00	64.20	89.40	74.73	1	0	1	0	321	38	179	8	13	73.29	55.00	87.20
38	16	114.00	232.00	386.00	27.00	32.90	22.80	65.40	94.50	77.30	1	0	1	0	327	19	173	8	17	78.77	60.00	86.80
42	16	89.00	201.00	411.00	22.50	30.70	17.80	49.80	85.90	63.05	1	0	1	0	249	41	251	10	19	49.37	39.60	86.10
46	16	70.00	190.00	430.00	18.40	26.90	14.00	29.20	56.20	38.43	1	0	1	0	146	114	354	6	12	20.74	5.20	83.60

(c) MSR = 16

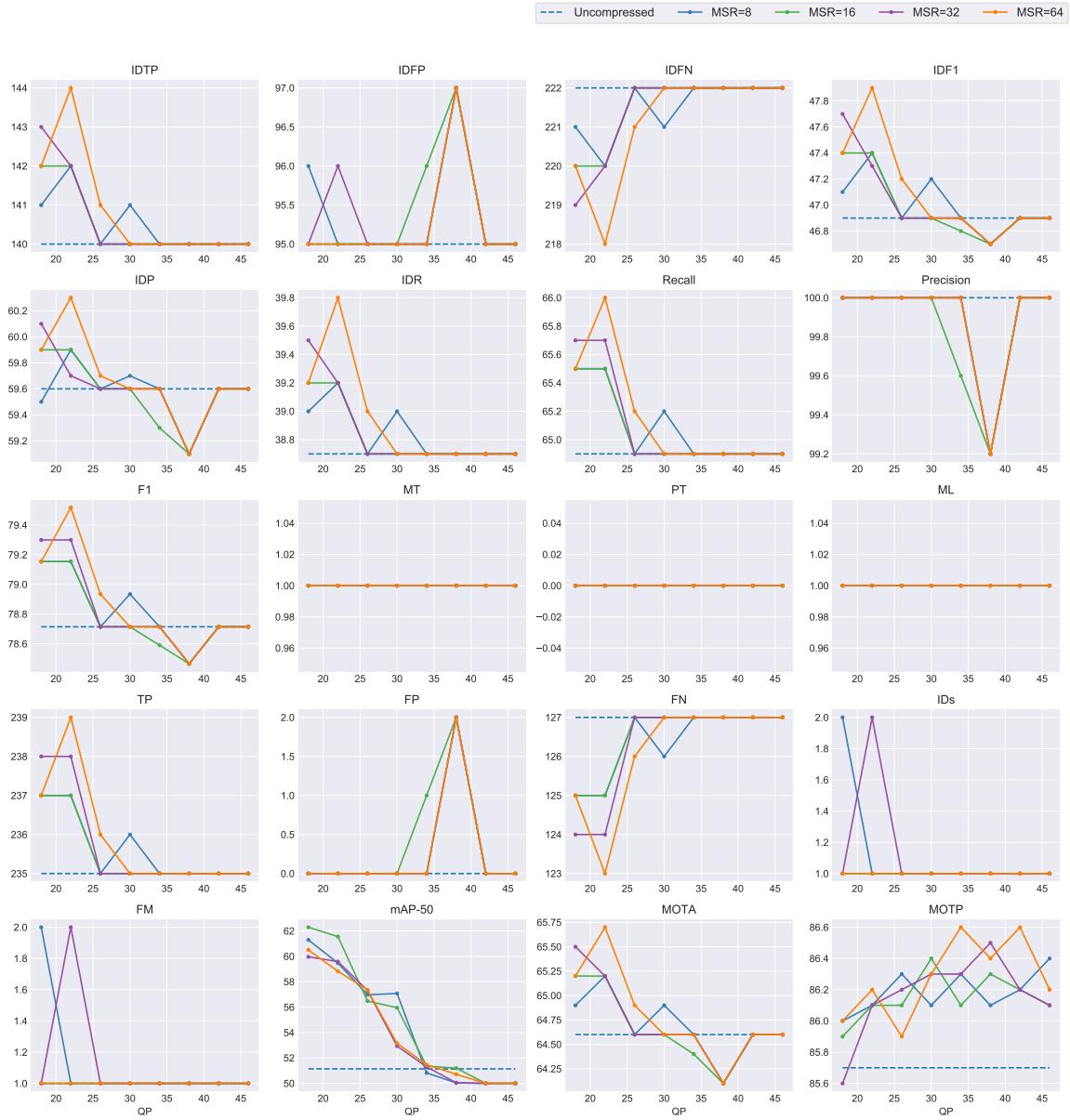
QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	32	117.00	273.00	383.00	26.30	30.00	23.40	60.40	77.40	67.85	1	0	1	0	302	88	198	6	13	68.81	41.60	88.10
22	32	113.00	265.00	387.00	25.70	29.90	22.60	56.60	74.90	64.48	1	0	1	0	283	95	217	8	13	67.56	36.00	87.80
26	32	131.00	236.00	369.00	30.20	35.70	26.20	63.20	86.10	72.89	1	0	1	0	316	51	184	7	11	70.20	51.60	87.80
30	32	87.00	256.00	413.00	20.60	25.40	17.40	60.60	88.30	71.87	1	0	1	0	303	40	197	8	16	72.45	51.00	87.10
34	32	93.00	258.00	407.00	21.90	26.50	18.60	60.80	86.60	71.44	1	0	1	0	304	47	196	9	11	68.17	49.60	87.90
38	32	95.00	290.00	405.00	21.50	24.70	19.00	69.20	89.90	78.20	1	0	1	0	346	39	154	5	14	72.88	60.40	86.60
42	32	113.00	224.00	387.00	27.00	33.50	22.60	51.60	76.60	61.66	1	0	1	0	258	79	242	5	10	50.48	34.80	87.50
46	32	81.00	145.00	419.00	22.30	35.80	16.20	30.80	68.10	42.42	1	0	1	0	154	72	346	3	8	27.33	15.80	85.50

(d) MSR = 32

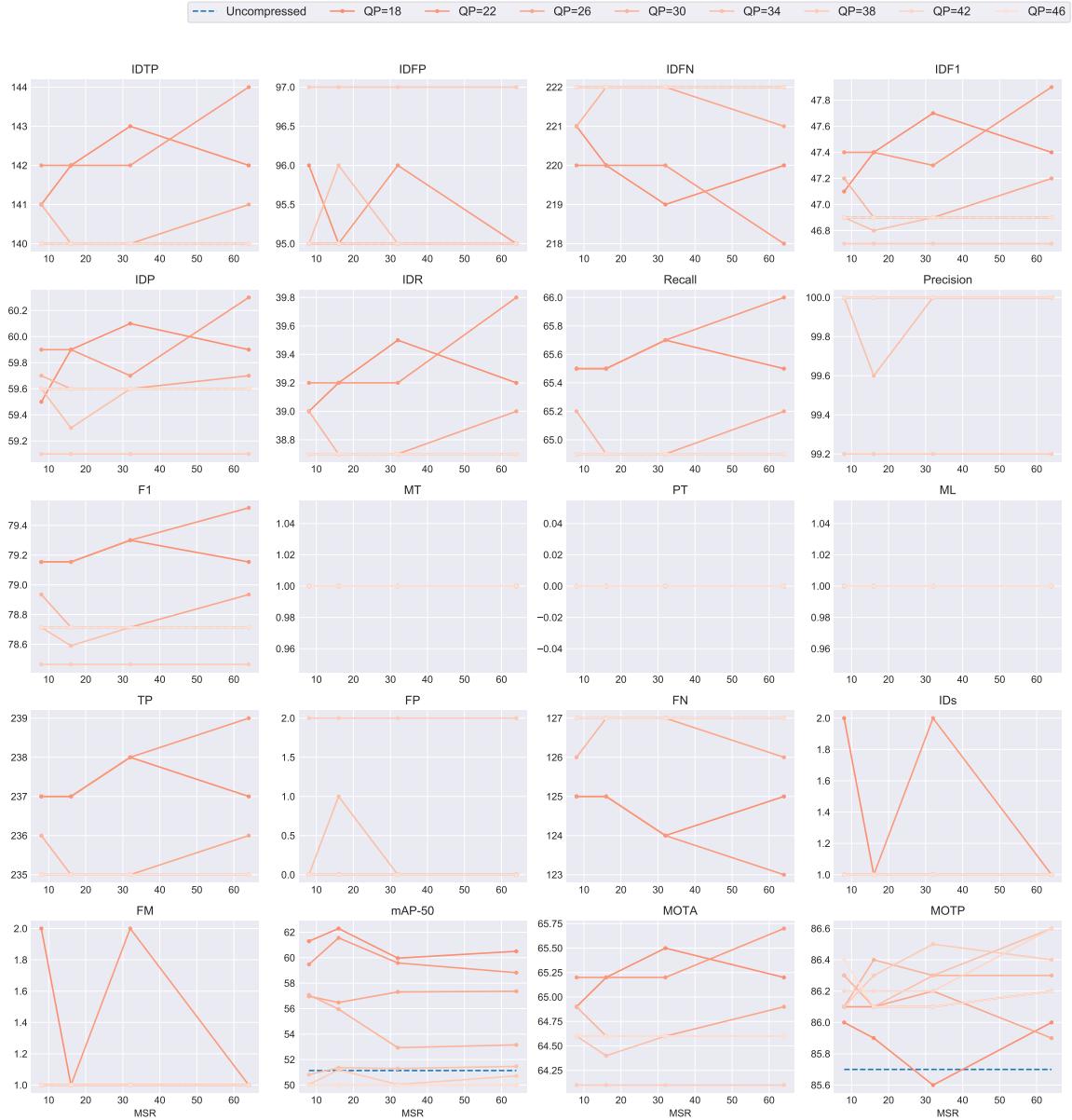
QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	64	127.00	240.00	373.00	29.30	34.60	25.40	61.40	83.70	70.84	1	0	1	0	307	60	193	6	11	69.23	48.20	87.90
22	64	124.00	265.00	376.00	27.90	31.90	24.80	61.20	78.70	68.86	1	0	1	0	306	83	194	8	11	67.03	43.00	87.90
26	64	115.00	248.00	385.00	26.70	31.70	23.00	58.40	80.40	67.66	1	0	1	0	292	71	208	9	14	68.98	42.40	87.90
30	64	107.00	228.00	393.00	25.60	31.90	21.40	56.60	84.50	67.79	1	0	1	0	283	52	217	5	15	65.85	45.20	87.50
34	64	62.00	283.00	438.00	14.70	18.00	12.40	61.80	89.60	73.15	1	0	1	0	309	36	191	8	10	70.71	53.00	87.50
38	64	134.00	236.00	366.00	30.80	36.20	26.80	62.00	83.80	71.27	1	0	1	0	310	60	190	6	15	67.75	48.80	86.90
42	64	102.00	205.00	398.00	25.30	33.20	20.40	50.00	81.40	61.95	1	0	1	0	250	57	250	8	12	56.89	37.00	86.00
46	64	36.00	186.00	464.00	10.00	16.20	7.20	26.40	59.50	36.57	1	0	1	0	132	90	368	6	8	23.88	7.20	84.80

(e) MSR = 64

### B.3 Kimono



**Figure B.5:** Visualization of performance results on Kimono at different QP.



**Figure B.6:** Visualization of performance results on Kimono at different MSR.

**Table B.3:** Performance results on Kimono.

QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
Uncompressed	Uncompressed	140.00	95.00	222.00	46.90	59.60	38.70	64.90	100.00	78.71	2	1	0	1	235	0	127	1	1	51.14	64.60	85.70

(a) Uncompressed Sequence

QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	8	141.00	96.00	221.00	47.10	59.50	39.00	65.50	100.00	79.15	2	1	0	1	237	0	125	2	2	61.29	64.90	86.00
22	8	142.00	95.00	220.00	47.40	59.90	39.20	65.50	100.00	79.15	2	1	0	1	237	0	125	1	1	59.47	65.20	86.10
26	8	140.00	95.00	222.00	46.90	59.60	38.70	64.90	100.00	78.71	2	1	0	1	235	0	127	1	1	56.97	64.60	86.30
30	8	141.00	95.00	221.00	47.20	59.70	39.00	65.20	100.00	78.93	2	1	0	1	236	0	126	1	1	57.08	64.90	86.10
34	8	140.00	95.00	222.00	46.90	59.60	38.70	64.90	100.00	78.71	2	1	0	1	235	0	127	1	1	50.82	64.60	86.30
38	8	140.00	97.00	222.00	46.70	59.10	38.70	64.90	99.20	78.47	2	1	0	1	235	2	127	1	1	50.04	64.10	86.10
42	8	140.00	95.00	222.00	46.90	59.60	38.70	64.90	100.00	78.71	2	1	0	1	235	0	127	1	1	50.00	64.60	86.20
46	8	140.00	95.00	222.00	46.90	59.60	38.70	64.90	100.00	78.71	2	1	0	1	235	0	127	1	1	50.00	64.60	86.40

(b) MSR = 8

QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	16	142.00	95.00	220.00	47.40	59.90	39.20	65.50	100.00	79.15	2	1	0	1	237	0	125	1	1	62.29	65.20	85.90
22	16	142.00	95.00	220.00	47.40	59.90	39.20	65.50	100.00	79.15	2	1	0	1	237	0	125	1	1	61.56	65.20	86.10
26	16	140.00	95.00	222.00	46.90	59.60	38.70	64.90	100.00	78.71	2	1	0	1	235	0	127	1	1	56.48	64.60	86.10
30	16	140.00	95.00	222.00	46.90	59.60	38.70	64.90	100.00	78.71	2	1	0	1	235	0	127	1	1	55.96	64.60	86.40
34	16	140.00	96.00	222.00	46.80	59.30	38.70	64.90	99.60	78.59	2	1	0	1	235	1	127	1	1	51.36	64.40	86.10
38	16	140.00	97.00	222.00	46.70	59.10	38.70	64.90	99.20	78.47	2	1	0	1	235	2	127	1	1	51.20	64.10	86.30
42	16	140.00	95.00	222.00	46.90	59.60	38.70	64.90	100.00	78.71	2	1	0	1	235	0	127	1	1	50.00	64.60	86.20
46	16	140.00	95.00	222.00	46.90	59.60	38.70	64.90	100.00	78.71	2	1	0	1	235	0	127	1	1	50.00	64.60	86.10

(c) MSR = 16

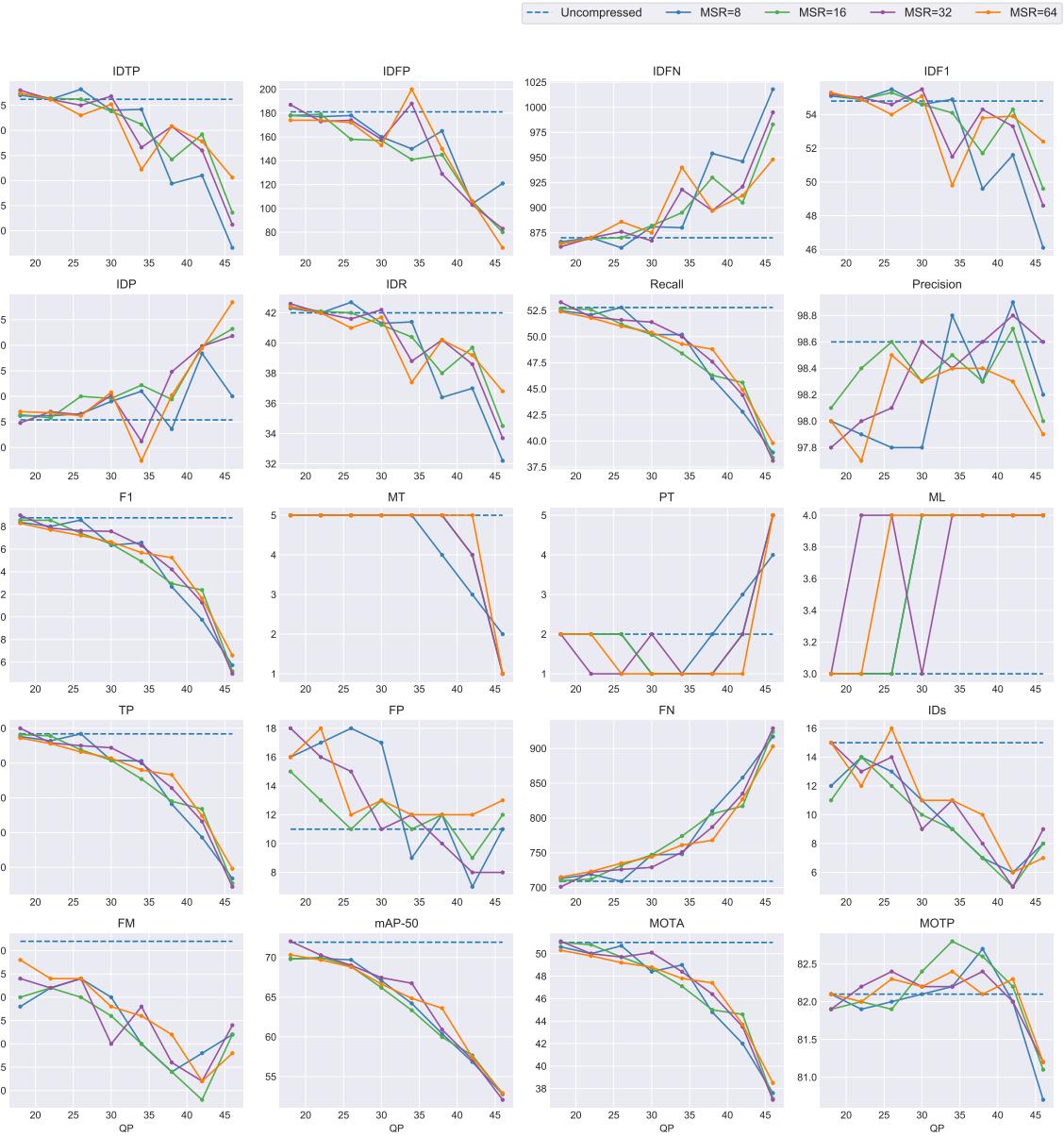
QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	32	143.00	95.00	219.00	47.70	60.10	39.50	65.70	100.00	79.30	2	1	0	1	238	0	124	1	1	59.96	65.50	85.60
22	32	142.00	96.00	220.00	47.30	59.70	39.20	65.70	100.00	79.30	2	1	0	1	238	0	124	2	2	59.59	65.20	86.10
26	32	140.00	95.00	222.00	46.90	59.60	38.70	64.90	100.00	78.71	2	1	0	1	235	0	127	1	1	57.31	64.60	86.20
30	32	140.00	95.00	222.00	46.90	59.60	38.70	64.90	100.00	78.71	2	1	0	1	235	0	127	1	1	52.93	64.60	86.30
34	32	140.00	95.00	222.00	46.90	59.60	38.70	64.90	100.00	78.71	2	1	0	1	235	0	127	1	1	51.30	64.60	86.30
38	32	140.00	97.00	222.00	46.70	59.10	38.70	64.90	99.20	78.47	2	1	0	1	235	2	127	1	1	50.05	64.10	86.50
42	32	140.00	95.00	222.00	46.90	59.60	38.70	64.90	100.00	78.71	2	1	0	1	235	0	127	1	1	50.00	64.60	86.20
46	32	140.00	95.00	222.00	46.90	59.60	38.70	64.90	100.00	78.71	2	1	0	1	235	0	127	1	1	50.00	64.60	86.10

(d) MSR = 32

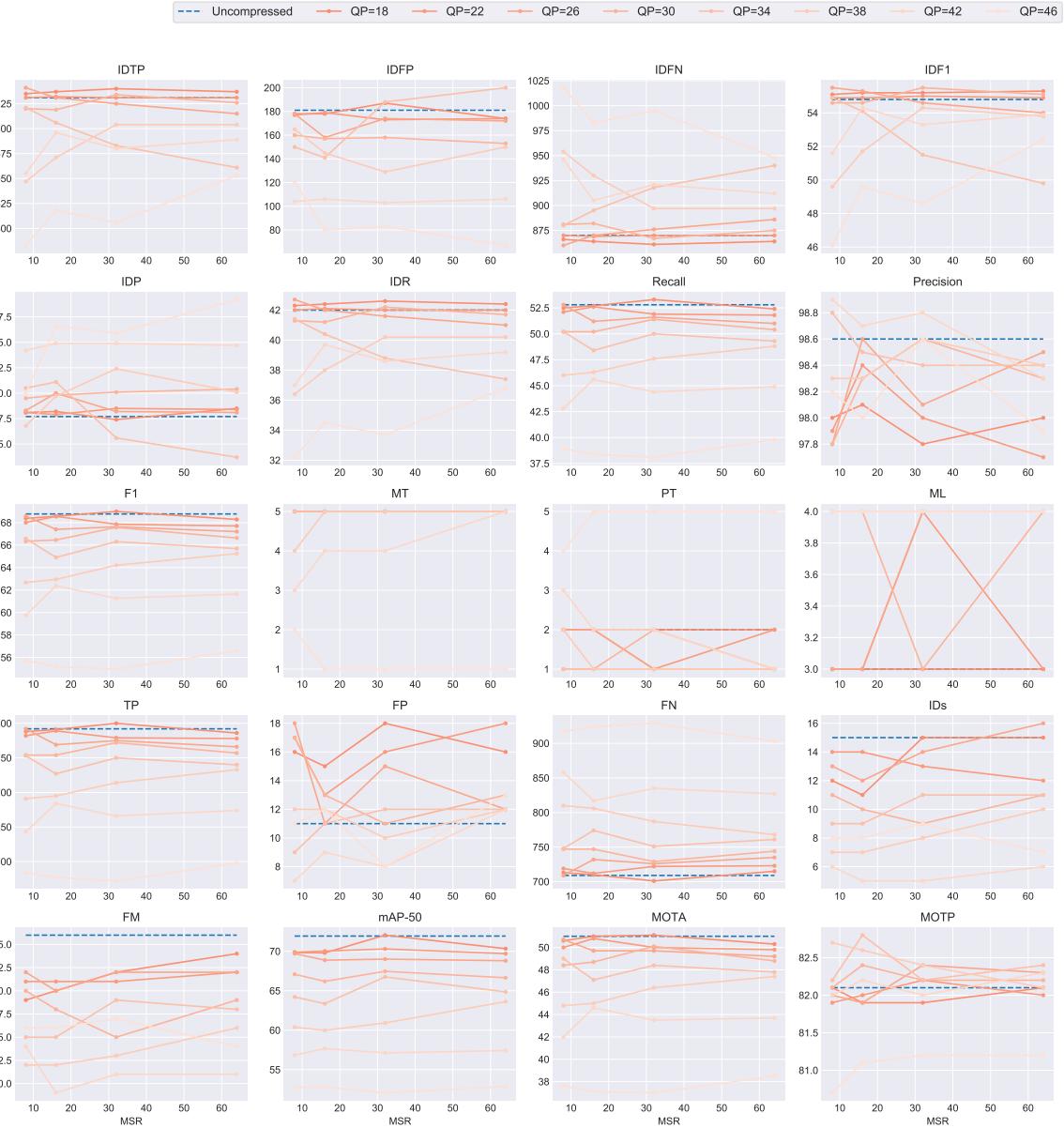
QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	64	142.00	95.00	220.00	47.40	59.90	39.20	65.50	100.00	79.15	2	1	0	1	237	0	125	1	1	60.50	65.20	86.00
22	64	144.00	95.00	218.00	47.90	60.30	39.80	66.00	100.00	79.52	2	1	0	1	239	0	123	1	1	58.83	65.70	86.20
26	64	141.00	95.00	221.00	47.20	59.70	39.00	65.20	100.00	78.93	2	1	0	1	236	0	126	1	1	57.36	64.90	85.90
30	64	140.00	95.00	222.00	46.90	59.60	38.70	64.90	100.00	78.71	2	1	0	1	235	0	127	1	1	53.15	64.60	86.30
34	64	140.00	95.00	222.00	46.90	59.60	38.70	64.90	100.00	78.71	2	1	0	1	235	0	127	1	1	51.46	64.60	86.60
38	64	140.00	97.00	222.00	46.70	59.10	38.70	64.90	99.20	78.47	2	1	0	1	235	2	127	1	1	50.72	64.10	86.40
42	64	140.00	95.00	222.00	46.90	59.60	38.70	64.90	100.00	78.71	2	1	0	1	235	0	127	1	1	50.00	64.60	86.60
46	64	140.00	95.00	222.00	46.90	59.60	38.70	64.90	100.00	78.71	2	1	0	1	235	0	127	1	1	50.00	64.60	86.20

(e) MSR = 64

## B.4 ParkScene



**Figure B.7:** Visualization of performance results on ParkScene at different QP.



**Figure B.8:** Visualization of performance results on ParkScene at different MSR.

**Table B.4:** Performance results on ParkScene.

QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
Uncompressed	Uncompressed	631.00	181.00	870.00	54.80	77.70	42.00	52.80	98.60	68.77	10	5	2	3	792	11	709	15	26	71.93	51.00	82.10

(a) Uncompressed Sequence

QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	8	635.00	178.00	866.00	55.10	78.10	42.30	52.50	98.00	68.37	10	5	2	3	788	16	713	12	19	69.85	50.60	82.10
22	8	631.00	177.00	870.00	54.90	78.10	42.00	52.10	97.90	68.01	10	5	2	3	782	17	719	14	21	69.89	50.00	81.90
26	8	641.00	178.00	860.00	55.50	78.30	42.70	52.80	97.80	68.58	10	5	2	3	792	18	709	13	22	69.70	50.70	82.00
30	8	620.00	160.00	881.00	54.60	79.50	41.30	50.20	97.80	66.35	10	5	1	4	754	17	747	11	20	67.09	48.40	82.10
34	8	621.00	150.00	880.00	54.90	80.50	41.40	50.20	98.80	66.57	10	5	1	4	753	9	748	9	15	64.21	49.00	82.20
38	8	547.00	165.00	954.00	49.60	76.80	36.40	46.00	98.30	62.67	10	4	2	4	691	12	810	7	12	60.39	44.80	82.70
42	8	555.00	104.00	946.00	51.60	84.20	37.00	42.80	98.90	59.74	10	3	3	4	643	7	858	6	14	56.84	42.00	82.00
46	8	483.00	121.00	1018.00	46.10	80.00	32.20	38.90	98.20	55.73	10	2	4	4	584	11	917	8	16	52.77	37.60	80.70

(b) MSR = 8

QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	16	637.00	178.00	864.00	55.20	78.20	42.40	52.70	98.10	68.57	10	5	2	3	791	15	710	11	20	69.79	51.00	81.90
22	16	632.00	179.00	869.00	54.90	77.90	42.10	52.60	98.40	68.55	10	5	2	3	789	13	712	14	21	70.03	50.80	82.00
26	16	631.00	158.00	870.00	55.30	80.00	42.00	51.20	98.60	67.40	10	5	2	3	769	11	732	12	20	68.88	49.70	81.90
30	16	619.00	157.00	882.00	54.60	79.80	41.20	50.20	98.30	66.46	10	5	1	4	754	13	747	10	18	66.19	48.70	82.40
34	16	606.00	141.00	895.00	54.10	81.10	40.40	48.40	98.50	64.91	10	5	1	4	727	11	774	9	15	63.36	47.10	82.80
38	16	571.00	145.00	930.00	51.70	79.70	38.00	46.30	98.30	62.95	10	5	1	4	695	12	806	7	12	59.97	45.00	82.60
42	16	596.00	106.00	905.00	54.30	84.90	39.70	45.60	98.70	62.38	10	4	2	4	684	9	817	5	9	57.67	44.60	82.20
46	16	518.00	80.00	983.00	49.60	86.60	34.50	38.40	98.00	55.18	10	1	5	4	577	12	924	8	16	52.82	37.10	81.10

(c) MSR = 16

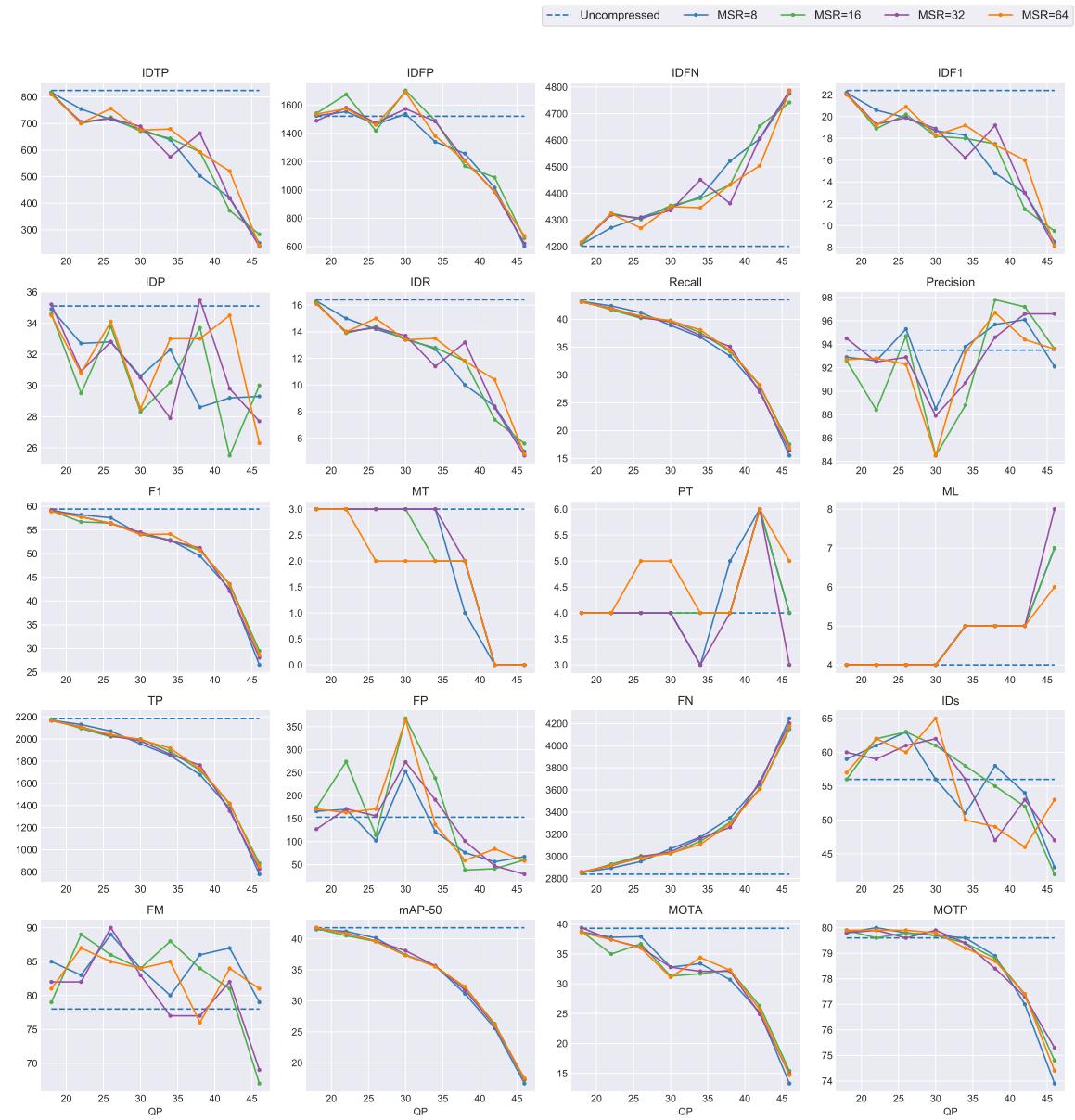
QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	32	640.00	187.00	861.00	55.20	77.40	42.60	53.30	97.80	69.00	10	5	2	3	800	18	701	15	22	72.04	51.10	81.90
22	32	631.00	173.00	870.00	55.00	78.50	42.00	51.90	98.00	67.86	10	5	1	4	779	16	722	13	21	70.29	50.00	82.20
26	32	625.00	174.00	876.00	54.60	78.20	41.60	51.60	98.10	67.63	10	5	1	4	775	15	726	14	22	69.00	49.70	82.40
30	32	634.00	158.00	867.00	55.50	80.10	42.20	51.40	98.60	67.57	10	5	2	3	772	11	729	9	15	67.46	50.10	82.20
34	32	583.00	188.00	918.00	51.50	75.60	38.80	50.00	98.40	66.31	10	5	1	4	750	12	751	11	19	66.75	48.40	82.20
38	32	604.00	129.00	897.00	54.30	82.40	40.20	47.60	98.60	64.20	10	5	1	4	714	10	787	8	13	60.92	46.40	82.40
42	32	580.00	103.00	921.00	53.30	84.90	38.60	44.40	98.80	61.27	10	4	2	4	666	8	835	5	11	57.13	43.50	82.00
46	32	506.00	83.00	995.00	48.60	85.90	33.70	38.10	98.60	54.96	10	1	5	4	572	8	929	9	17	52.07	37.00	81.20

(d) MSR = 32

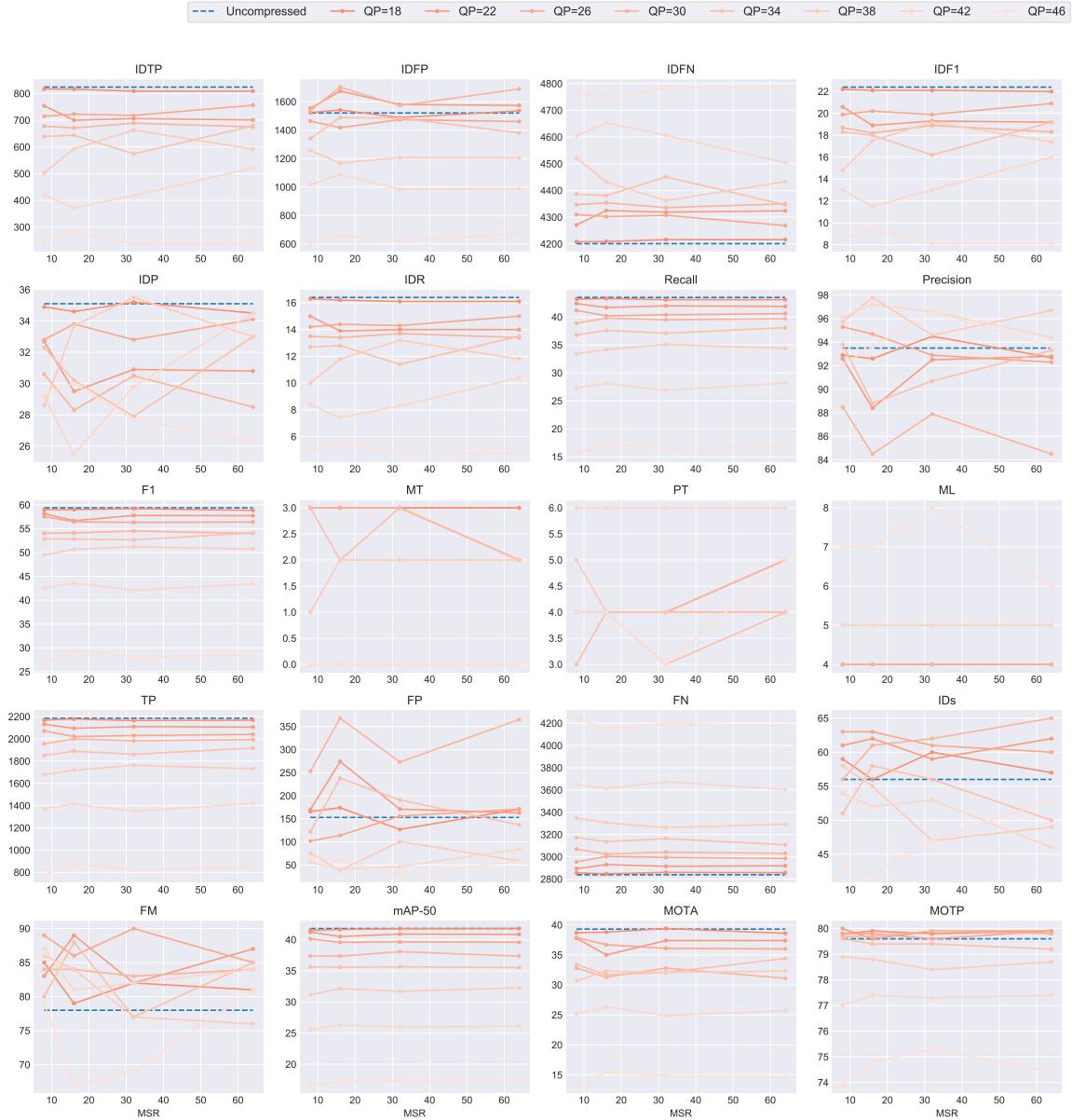
QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	64	637.00	174.00	864.00	55.30	78.50	42.40	52.40	98.00	68.29	10	5	2	3	786	16	715	15	24	70.33	50.30	82.10
22	64	631.00	174.00	870.00	54.90	78.40	42.00	51.80	97.70	67.70	10	5	2	3	778	18	723	12	22	69.70	49.80	82.00
26	64	615.00	172.00	886.00	54.00	78.10	41.00	51.00	98.50	67.20	10	5	1	4	766	12	735	16	22	68.80	49.20	82.30
30	64	626.00	153.00	875.00	55.10	80.40	41.70	50.40	98.30	66.64	10	5	1	4	757	13	744	11	19	66.65	48.80	82.20
34	64	561.00	200.00	940.00	49.80	73.70	37.40	49.30	98.40	65.69	10	5	1	4	740	12	761	11	18	64.86	47.80	82.40
38	64	604.00	150.00	897.00	53.80	80.10	40.20	48.80	98.40	65.24	10	5	1	4	733	12	768	10	16	63.62	47.40	82.10
42	64	589.00	106.00	912.00	53.90	84.70	39.20	44.90	98.30	61.64	10	5	1	4	674	12	827	6	11	57.42	43.70	82.30
46	64	553.00	67.00	948.00	52.40	89.20	36.80	39.80	97.90	56.59	10	1	5	4	598	13	903	7	14	52.88	38.50	81.20

(e) MSR = 64

## B.5 BasketballDrill



**Figure B.9:** Visualization of performance results on BasketballDrill at different QP.

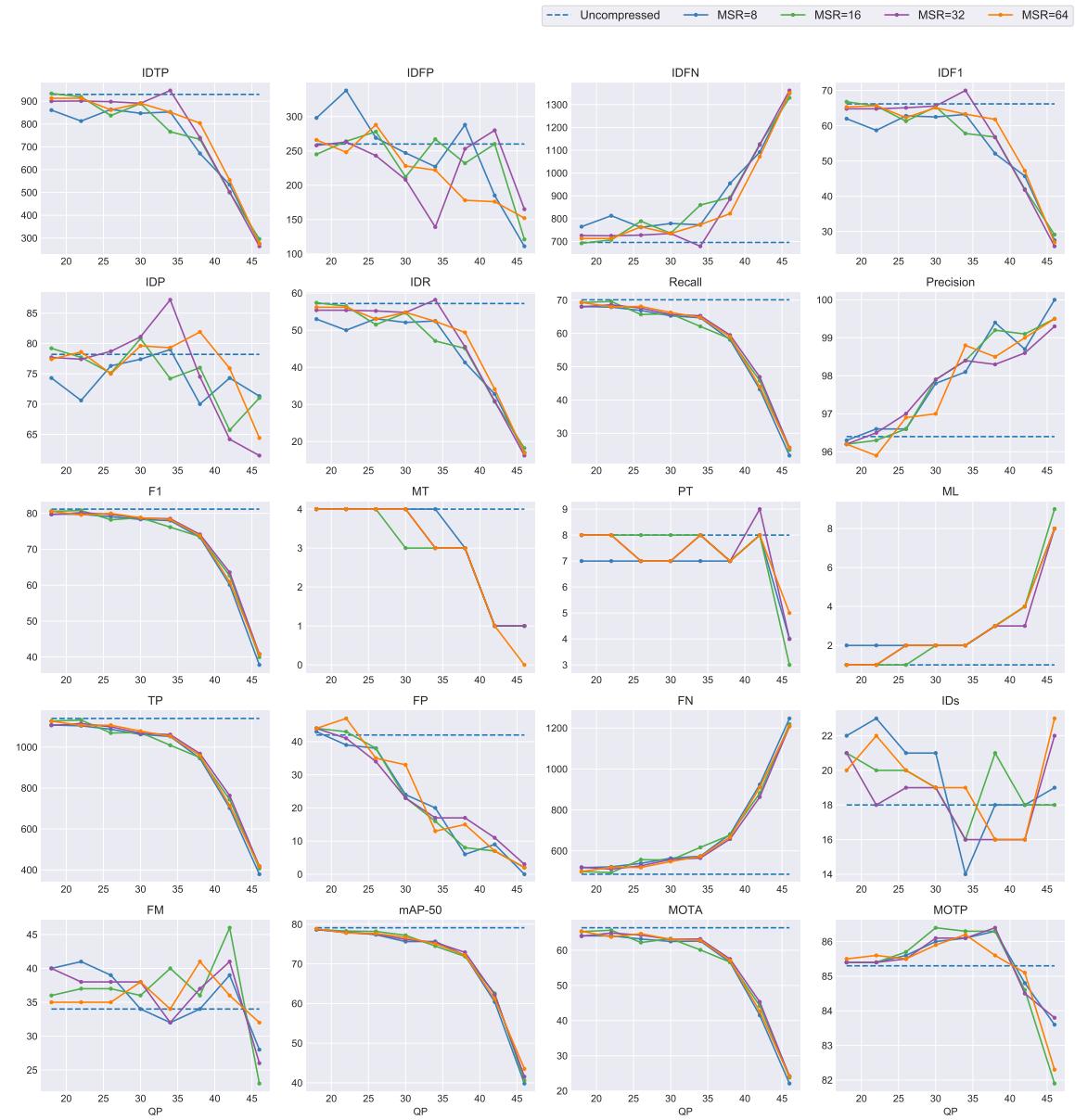


**Figure B.10:** Visualization of performance results on BasketballDrill at different MSR.

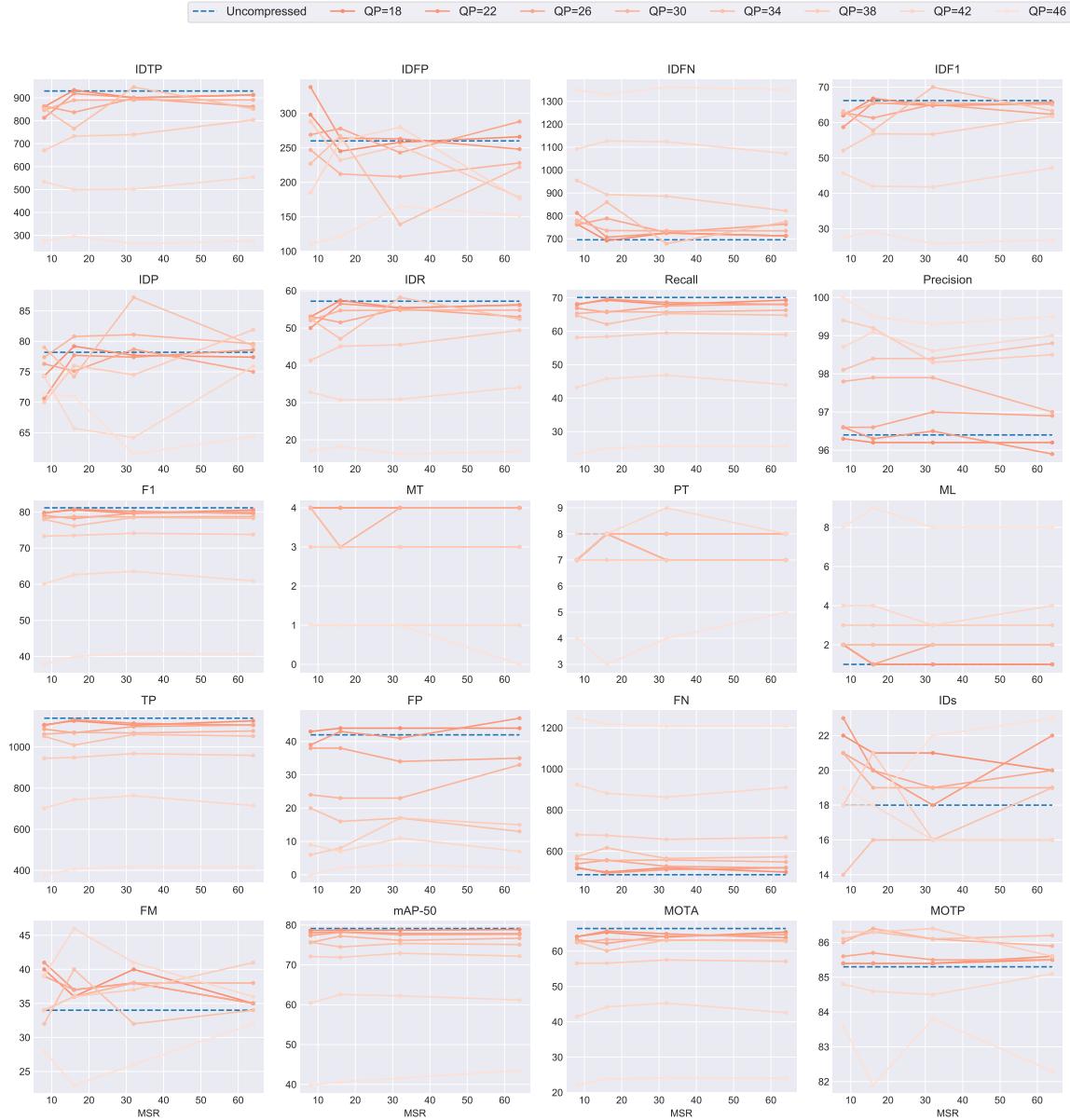
**Table B.5:** Performance results on BasketballDrill.

QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
Uncompressed	Uncompressed	824.00	1521.00	4201.00	22.40	35.10	16.40	43.50	93.50	59.38	11	3	4	4	2186	153	2839	56	78	41.79	39.30	79.60
<b>(a) Uncompressed Sequence</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	8	817.00	1524.00	4208.00	22.20	34.90	16.30	43.20	92.90	58.98	11	3	4	4	2169	166	2856	59	85	41.51	38.70	79.80
22	8	754.00	1553.00	4271.00	20.60	32.70	15.00	42.40	92.60	58.17	11	3	4	4	2131	170	2894	61	83	41.21	37.80	80.00
26	8	715.00	1464.00	4310.00	19.90	32.80	14.20	41.20	95.30	57.53	11	3	4	4	2071	102	2954	63	89	40.14	37.90	79.80
30	8	678.00	1537.00	4347.00	18.70	30.60	13.50	38.90	88.50	54.04	11	3	4	4	1956	253	3069	56	84	37.41	32.80	79.70
34	8	639.00	1340.00	4386.00	18.30	32.30	12.70	36.80	93.80	52.86	11	3	3	5	1851	122	3174	51	80	35.65	33.40	79.60
38	8	503.00	1257.00	4522.00	14.80	28.60	10.00	33.40	95.70	49.52	11	1	5	5	1678	76	3347	58	86	31.14	30.70	78.90
42	8	420.00	1016.00	4605.00	13.00	29.20	8.40	27.30	96.10	42.52	11	0	6	5	1374	56	3651	54	87	25.59	25.20	77.00
46	8	250.00	602.00	4775.00	8.50	29.30	5.00	15.50	92.10	26.53	11	0	4	7	779	67	4246	43	79	16.66	13.30	73.90
<b>(b) MSR = 8</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	16	816.00	1542.00	4209.00	22.10	34.60	16.20	43.30	92.60	59.01	11	3	4	4	2178	174	2847	56	79	41.64	38.80	79.90
22	16	700.00	1675.00	4325.00	18.90	29.50	13.90	41.70	88.40	56.67	11	3	4	4	2095	274	2930	62	89	40.52	35.00	79.60
26	16	723.00	1418.00	4302.00	20.20	33.80	14.40	40.20	94.70	56.44	11	3	4	4	2021	114	3004	63	86	39.57	36.70	79.80
30	16	671.00	1703.00	4354.00	18.20	28.30	13.40	39.80	84.50	54.11	11	3	4	4	2000	368	3025	61	84	37.36	31.30	79.70
34	16	644.00	1489.00	4381.00	18.00	30.20	12.80	37.60	88.80	52.83	11	2	4	5	1889	238	3136	58	88	35.59	31.70	79.40
38	16	593.00	1169.00	4432.00	17.50	33.70	11.80	34.20	97.80	50.68	11	2	4	5	1718	38	3307	55	84	32.14	32.30	78.80
42	16	372.00	1088.00	4653.00	11.50	25.50	7.40	28.10	97.20	43.60	11	0	6	5	1413	41	3612	52	81	26.31	26.30	77.40
46	16	283.00	660.00	4742.00	9.50	30.00	5.60	17.50	93.60	29.49	11	0	4	7	877	60	4148	42	67	17.20	15.40	74.80
<b>(c) MSR = 16</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	32	809.00	1489.00	4216.00	22.10	35.20	16.10	43.10	94.50	59.20	11	3	4	4	2165	127	2860	60	82	41.77	39.40	79.80
22	32	706.00	1581.00	4319.00	19.30	30.90	14.00	42.00	92.50	57.77	11	3	4	4	2110	171	2915	59	82	40.91	37.40	79.90
26	32	718.00	1474.00	4307.00	19.90	32.80	14.30	40.40	92.90	56.31	11	3	4	4	2030	156	2995	61	90	39.65	36.10	79.60
30	32	689.00	1573.00	4336.00	18.90	30.50	13.70	39.50	87.90	54.51	11	3	4	4	1983	273	3042	62	83	38.09	32.80	79.90
34	32	574.00	1485.00	4451.00	16.20	27.90	11.40	37.10	90.70	52.66	11	3	5	4	1862	191	3163	56	77	35.64	32.10	79.40
38	32	663.00	1207.00	4362.00	19.20	35.50	13.20	35.10	94.60	51.20	11	2	4	5	1763	101	3262	47	77	31.73	32.10	78.40
42	32	418.00	985.00	4607.00	13.00	29.80	8.30	26.90	96.60	42.08	11	0	6	5	1350	47	3675	53	82	25.98	24.90	77.30
46	32	238.00	621.00	4787.00	8.10	27.70	4.70	16.40	96.60	28.04	11	0	3	8	824	29	4201	47	69	17.37	14.90	75.30
<b>(d) MSR = 32</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	64	809.00	1536.00	4216.00	22.00	34.50	16.10	43.10	92.70	58.84	11	3	4	4	2168	171	2857	57	81	41.80	38.60	79.90
22	64	701.00	1574.00	4324.00	19.20	30.80	14.00	41.90	92.80	57.73	11	3	4	4	2106	163	2919	62	87	40.86	37.40	79.90
26	64	756.00	1461.00	4269.00	20.90	34.10	15.00	40.60	92.30	56.39	11	2	5	4	2040	171	2985	60	85	39.58	36.00	79.90
30	64	675.00	1690.00	4350.00	18.30	28.50	13.40	39.70	84.50	54.02	11	2	5	4	1994	365	3031	65	84	37.35	31.10	79.80
34	64	679.00	1381.00	4346.00	19.20	33.00	13.50	38.10	93.30	54.11	11	2	4	5	1917	137	3108	50	85	35.54	34.40	79.20
38	64	592.00	1204.00	4433.00	17.40	33.00	11.80	34.40	96.70	50.75	11	2	4	5	1731	59	3294	49	76	32.27	32.30	78.70
42	64	521.00	988.00	4504.00	16.00	34.50	10.40	28.20	94.40	43.43	11	0	6	5	1419	84	3606	46	84	26.13	25.70	77.40
46	64	240.00	674.00	4785.00	8.10	26.30	4.80	16.90	93.60	28.63	11	0	5	6	850	58	4175	53	81	17.51	14.70	74.40
<b>(e) MSR = 64</b>																						

## B.6 RaceHorsesC



**Figure B.11:** Visualization of performance results on RaceHorsesC at different QP.



**Figure B.12:** Visualization of performance results on RaceHorsesC at different MSR.

**Table B.6:** Performance results on RaceHorsesC.

QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
Uncompressed	Uncompressed	930.00	260.00	696.00	66.20	78.20	57.20	70.10	96.40	81.17	13	4	8	1	1140	42	486	18	34	79.11	66.40	85.30

(a) Uncompressed Sequence

QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	8	861.00	298.00	765.00	62.00	74.30	53.00	68.10	96.30	79.78	13	4	7	2	1108	43	518	22	40	78.63	64.10	85.40
22	8	813.00	338.00	813.00	58.70	70.60	50.00	67.90	96.60	79.75	13	4	7	2	1104	39	522	23	41	78.05	64.10	85.40
26	8	864.00	269.00	762.00	62.80	76.30	53.10	66.90	96.60	79.05	13	4	7	2	1087	38	539	21	39	77.37	63.20	85.60
30	8	847.00	247.00	779.00	62.50	77.40	52.10	65.30	97.80	78.31	13	4	7	2	1062	24	564	21	34	75.62	62.50	86.00
34	8	853.00	227.00	773.00	63.20	79.00	52.50	64.70	98.10	77.97	13	4	7	2	1052	20	574	14	32	75.66	62.60	86.10
38	8	671.00	288.00	955.00	52.10	70.00	41.30	58.10	99.40	73.34	13	3	7	3	945	6	681	18	34	72.11	56.60	86.30
42	8	534.00	185.00	1092.00	45.70	74.30	32.80	43.20	98.70	60.10	13	1	8	4	702	9	924	18	39	60.41	41.50	84.80
46	8	276.00	111.00	1350.00	27.50	71.30	17.00	23.30	100.00	37.79	13	1	4	8	379	0	1247	19	28	39.83	22.10	83.60

(b) MSR = 8

QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	16	934.00	245.00	692.00	66.80	79.20	57.40	69.30	96.20	80.56	13	4	8	1	1127	44	499	21	36	78.72	65.30	85.40
22	16	919.00	264.00	707.00	65.60	77.70	56.50	69.60	96.30	80.80	13	4	8	1	1132	43	494	20	37	78.27	65.70	85.40
26	16	837.00	278.00	789.00	61.30	75.10	51.50	65.70	96.60	78.21	13	4	8	1	1069	38	557	20	37	78.19	62.20	85.70
30	16	890.00	212.00	736.00	65.40	80.80	54.70	65.90	97.90	78.77	13	3	8	2	1071	23	555	19	36	77.21	63.30	86.40
34	16	766.00	267.00	860.00	57.80	74.20	47.10	62.10	98.40	76.15	13	3	8	2	1009	16	617	16	40	74.49	60.10	86.30
38	16	733.00	232.00	893.00	56.80	76.00	45.10	58.40	99.20	73.52	13	3	7	3	949	8	677	21	36	71.87	56.60	86.30
42	16	499.00	260.00	1127.00	42.00	65.70	30.70	45.80	99.10	62.65	13	1	8	4	744	7	882	18	46	62.55	44.20	84.60
46	16	296.00	121.00	1330.00	29.10	71.00	18.20	25.00	99.50	39.96	13	1	3	9	407	2	1219	18	23	40.60	23.80	81.90

(c) MSR = 16

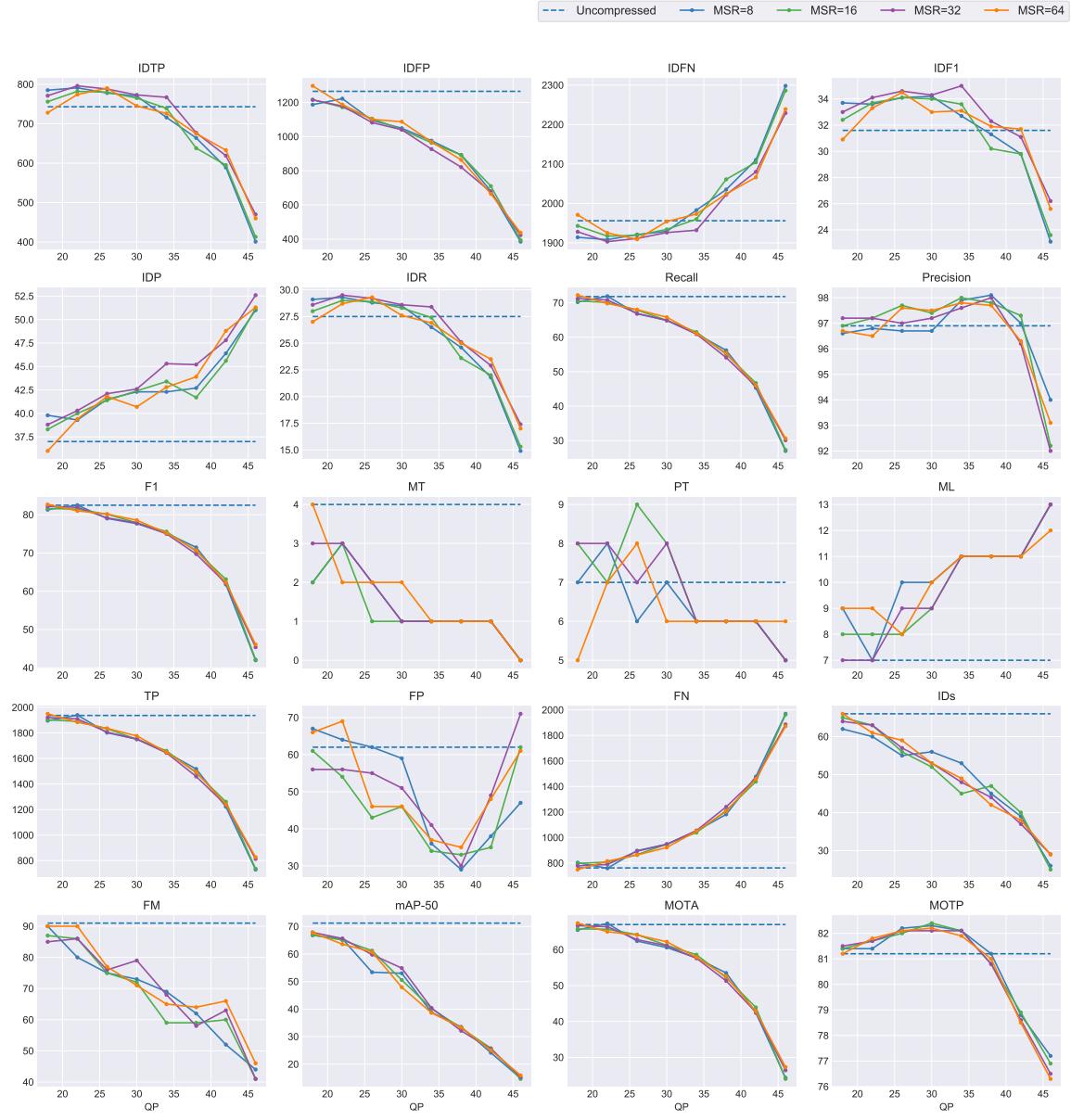
QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	32	900.00	258.00	726.00	64.80	77.70	55.40	68.00	96.20	79.68	13	4	8	1	1106	44	520	21	40	78.70	64.00	85.40
22	32	901.00	263.00	725.00	64.80	77.40	55.40	68.60	96.50	80.19	13	4	8	1	1115	41	511	18	38	77.87	64.90	85.40
26	32	898.00	243.00	728.00	65.10	78.70	55.20	67.60	97.00	79.67	13	4	7	2	1099	34	527	19	38	77.55	64.30	85.50
30	32	891.00	208.00	735.00	65.60	81.10	54.80	65.70	97.90	78.63	13	4	7	2	1068	23	558	19	38	76.15	63.10	86.10
34	32	947.00	139.00	679.00	70.00	87.20	58.20	65.30	98.40	78.50	13	3	8	2	1061	17	565	16	32	75.37	63.20	86.10
38	32	740.00	253.00	886.00	56.70	74.50	45.50	59.50	98.30	74.13	13	3	7	3	968	17	658	16	37	72.91	57.50	86.40
42	32	502.00	280.00	1124.00	41.80	64.20	30.90	46.90	98.60	63.56	13	1	9	3	763	11	863	16	41	62.24	45.30	84.50
46	32	264.00	165.00	1362.00	25.80	61.50	16.20	25.70	99.30	40.83	13	1	4	8	418	3	1208	22	26	41.56	24.20	83.80

(d) MSR = 32

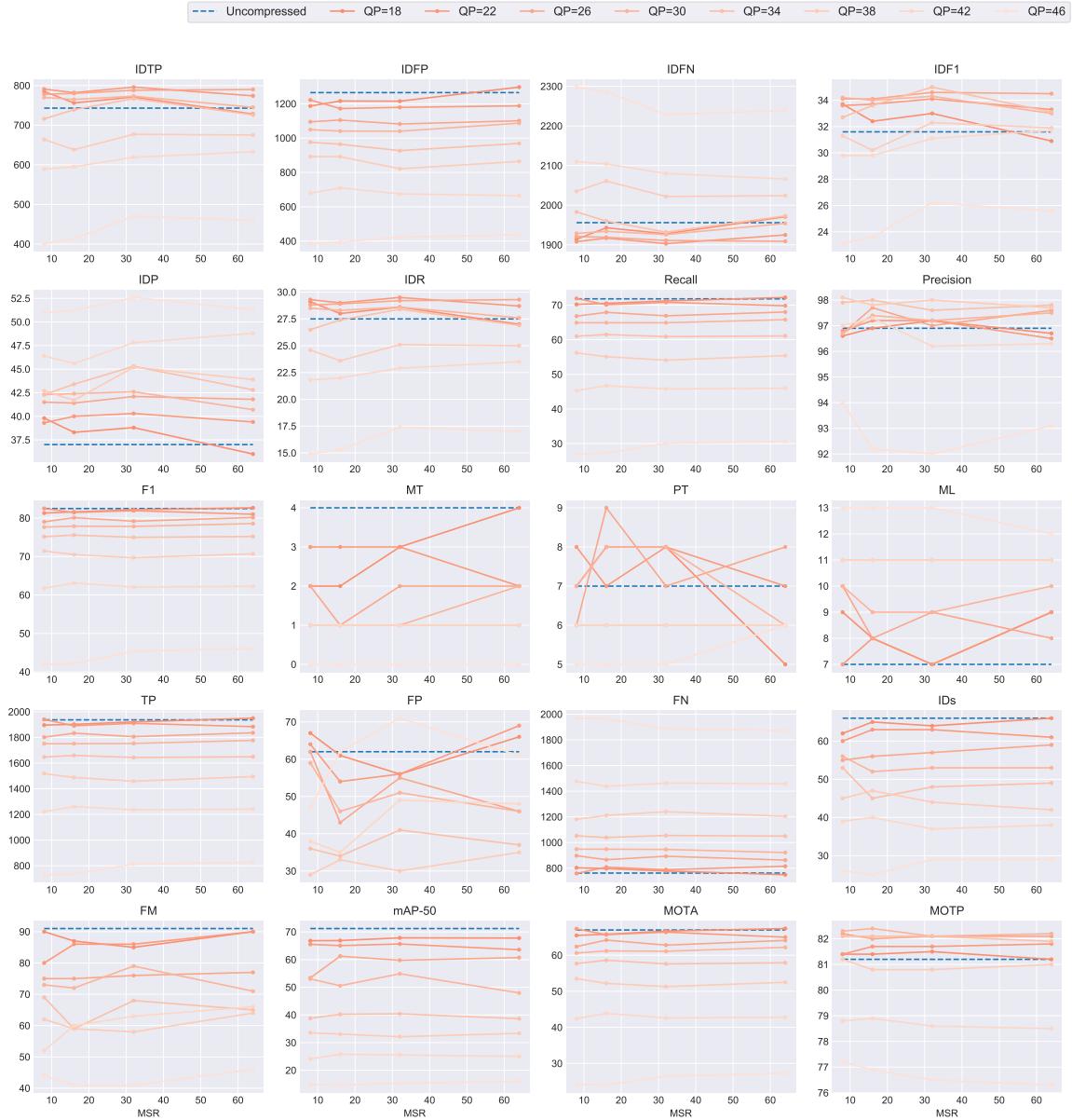
QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	64	913.00	266.00	713.00	65.30	77.40	56.20	69.30	96.20	80.56	13	4	8	1	1127	44	499	20	35	78.86	65.40	85.50
22	64	913.00	248.00	713.00	65.70	78.60	56.20	68.00	95.90	79.58	13	4	8	1	1106	47	520	22	35	77.81	63.80	85.60
26	64	862.00	288.00	764.00	62.30	75.00	53.00	68.10	96.90	79.99	13	4	7	2	1107	35	519	20	35	77.67	64.70	85.50
30	64	891.00	228.00	735.00	65.10	79.60	54.80	66.30	97.00	78.76	13	4	7	2	1078	33	548	19	38	76.67	63.10	85.90
34	64	852.00	222.00	774.00	63.30	79.30	52.40	64.80	98.80	78.27	13	3	8	2	1053	13	573	19	34	75.08	62.80	86.20
38	64	804.00	178.00	822.00	61.80	81.90	49.40	59.00	98.50	73.80	13	3	7	3	959	15	667	16	41	72.18	57.10	85.60
42	64	554.00	176.00	1072.00	47.20	75.90	34.10	44.00	99.00	60.92	13	1	8	4	715	7	911	16	36	61.17	42.60	85.10
46	64	275.00	152.00	1351.00	26.90	64.40	16.90	25.60	99.50	40.72	13	0	5	8	417	2	1209	23	32	43.52	24.10	82.30

(e) MSR = 64

## B.7 BasketballPass



**Figure B.13:** Visualization of performance results on BasketballPass at different QP.

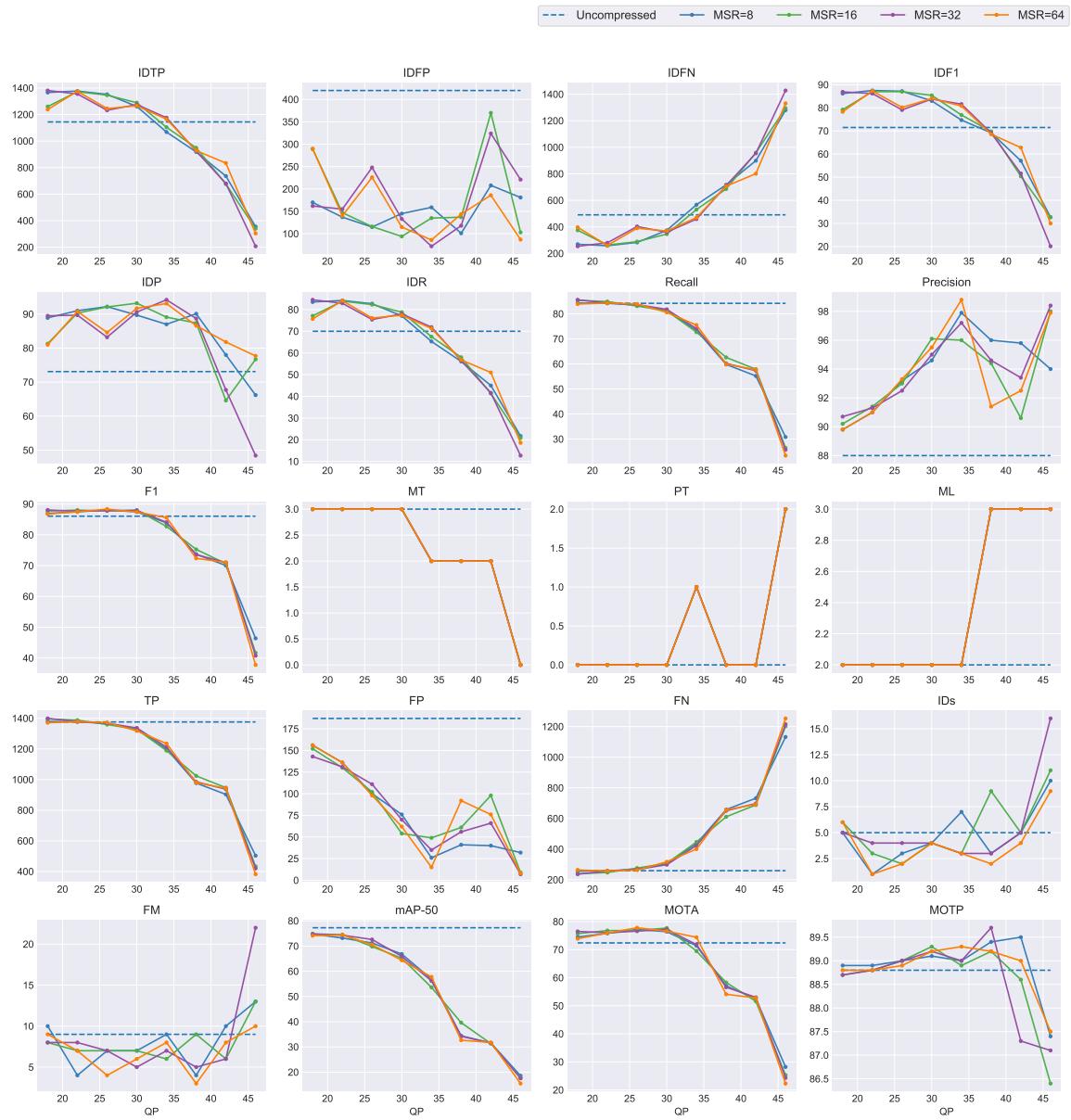


**Figure B.14:** Visualization of performance results on BasketballPass at different MSR.

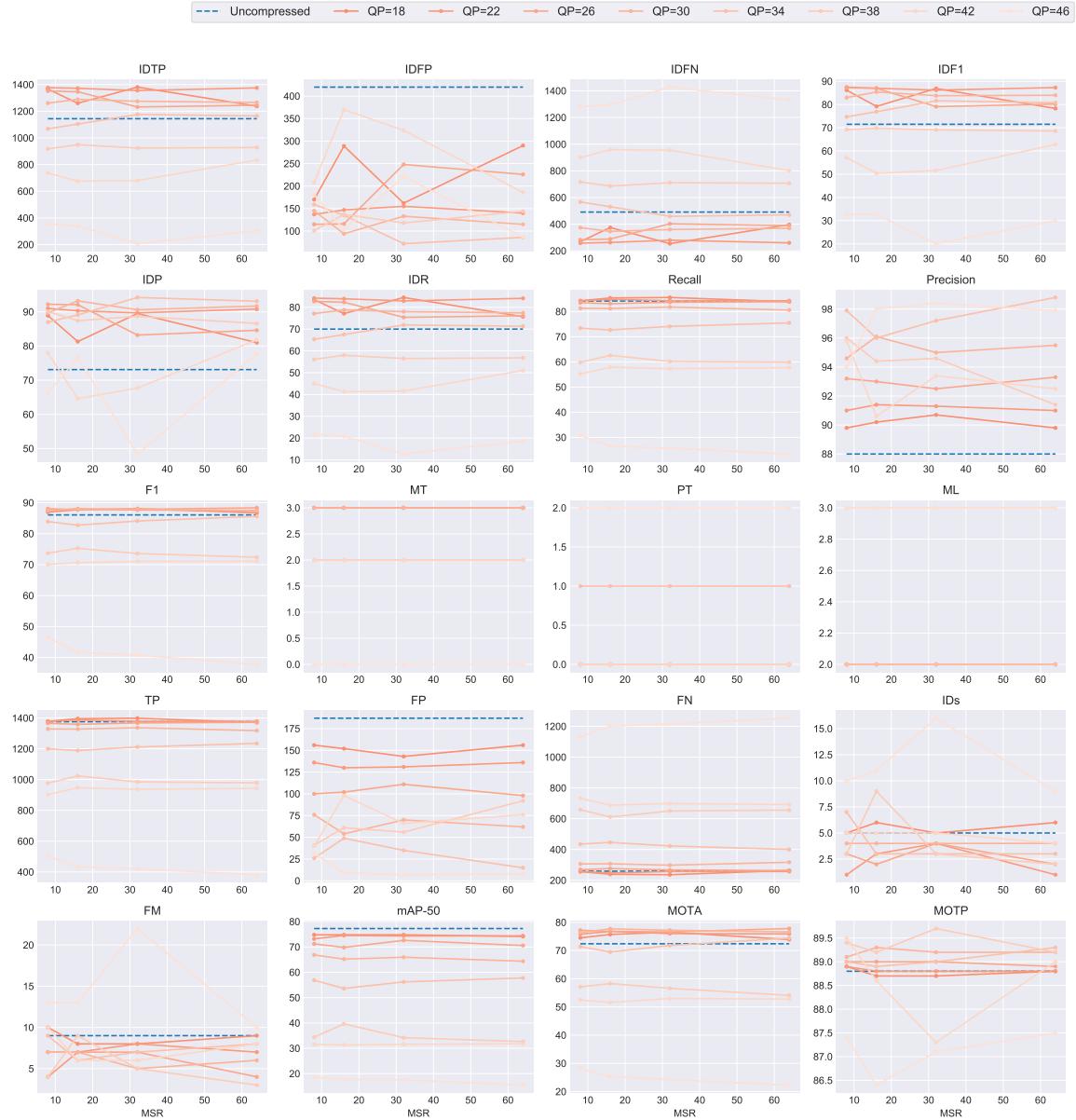
**Table B.7:** Performance results on BasketballPass.

QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
Uncompressed	Uncompressed	743.00	1265.00	1956.00	31.60	37.00	27.50	71.80	96.90	82.48	18	4	7	7	1937	62	762	66	91	71.22	67.00	81.20
<b>(a) Uncompressed Sequence</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	8	785.00	1187.00	1914.00	33.70	39.80	29.10	70.20	96.60	81.31	18	2	7	9	1896	67	803	62	90	66.81	65.50	81.40
22	8	791.00	1222.00	1908.00	33.60	39.30	29.30	71.90	96.80	82.51	18	3	8	7	1940	64	759	60	80	65.50	67.30	81.40
26	8	778.00	1095.00	1921.00	34.10	41.50	28.80	66.80	96.70	79.02	18	2	6	10	1802	62	897	55	75	53.38	62.40	82.20
30	8	770.00	1049.00	1929.00	34.20	42.30	28.50	64.90	96.70	77.67	18	1	7	10	1751	59	948	56	73	53.01	60.60	82.30
34	8	716.00	976.00	1983.00	32.70	42.30	26.50	61.00	97.90	75.17	18	1	6	11	1647	36	1052	53	69	38.80	57.70	82.10
38	8	664.00	892.00	2035.00	31.30	42.70	24.60	56.20	98.10	71.46	18	1	6	11	1518	29	1181	45	62	33.54	53.50	81.20
42	8	589.00	680.00	2110.00	29.80	46.40	21.80	45.30	97.00	61.76	18	1	6	11	1222	38	1477	39	52	24.14	42.40	78.80
46	8	401.00	385.00	2298.00	23.10	51.00	14.90	27.00	94.00	41.95	18	0	5	13	730	47	1969	26	44	14.91	24.30	77.20
<b>(b) MSR = 8</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	16	756.00	1216.00	1943.00	32.40	38.30	28.00	70.50	96.90	81.62	18	2	8	8	1902	61	797	65	87	66.87	65.80	81.40
22	16	782.00	1172.00	1917.00	33.70	40.00	29.00	70.10	97.20	81.46	18	3	7	8	1891	54	808	63	86	65.05	65.70	81.70
26	16	780.00	1105.00	1919.00	34.10	41.40	28.90	67.90	97.70	80.12	18	1	9	8	1833	43	866	56	75	61.21	64.20	82.00
30	16	765.00	1041.00	1934.00	34.00	42.40	28.30	64.90	97.40	77.90	18	1	8	9	1751	46	948	52	72	50.56	61.20	82.40
34	16	739.00	964.00	1960.00	33.60	43.40	27.40	61.50	98.00	75.57	18	1	6	11	1660	34	1039	45	59	40.22	58.60	82.10
38	16	638.00	892.00	2061.00	30.20	41.70	23.60	55.10	97.80	70.49	18	1	6	11	1488	33	1211	47	59	33.09	52.20	80.80
42	16	595.00	710.00	2104.00	29.80	45.60	22.00	46.70	97.30	63.11	18	1	6	11	1261	35	1438	40	60	25.78	43.90	78.90
46	16	413.00	394.00	2286.00	23.60	51.20	15.30	27.30	92.20	42.13	18	0	5	13	736	62	1963	25	41	14.65	24.00	76.90
<b>(c) MSR = 16</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	32	771.00	1215.00	1928.00	33.00	38.80	28.60	71.20	97.20	82.19	18	3	8	7	1921	56	778	64	85	67.85	66.70	81.50
22	32	796.00	1179.00	1903.00	34.10	40.30	29.50	70.80	97.20	81.93	18	3	8	7	1910	56	789	63	86	65.63	66.40	81.70
26	32	788.00	1082.00	1911.00	34.60	42.10	29.20	66.90	97.00	79.19	18	2	7	9	1806	55	893	57	76	59.75	62.80	82.10
30	32	773.00	1040.00	1926.00	34.30	42.60	28.60	64.90	97.20	77.83	18	1	8	9	1753	51	946	53	79	54.88	61.10	82.10
34	32	767.00	927.00	1932.00	35.00	45.30	28.40	60.90	97.60	75.00	18	1	6	11	1644	41	1055	48	68	40.40	57.60	82.10
38	32	677.00	821.00	2022.00	32.30	45.20	25.10	54.10	98.00	69.71	18	1	6	11	1459	30	1240	44	58	32.11	51.30	80.80
42	32	619.00	675.00	2080.00	31.10	47.80	22.90	45.80	96.20	62.06	18	1	6	11	1236	49	1463	37	63	25.54	42.60	78.60
46	32	470.00	423.00	2229.00	26.20	52.60	17.40	30.10	92.00	45.36	18	0	5	13	813	71	1886	29	41	15.31	26.40	76.50
<b>(d) MSR = 32</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	64	728.00	1297.00	1971.00	30.90	36.00	27.00	72.20	96.70	82.67	18	4	5	9	1950	66	749	66	90	67.78	67.40	81.20
22	64	774.00	1188.00	1925.00	33.30	39.40	28.70	69.80	96.50	81.01	18	2	7	9	1884	69	815	61	90	65.57	65.00	81.80
26	64	790.00	1101.00	1909.00	34.50	41.80	29.30	68.00	97.60	80.15	18	2	8	8	1836	46	863	59	77	60.72	64.10	82.10
30	64	745.00	1087.00	1954.00	33.00	40.70	27.60	65.80	97.50	78.57	18	2	6	10	1777	46	922	53	71	47.94	62.20	82.20
34	64	726.00	969.00	1973.00	33.10	42.80	26.90	61.10	97.80	75.21	18	1	6	11	1649	37	1050	49	65	38.64	57.90	81.90
38	64	675.00	864.00	2024.00	31.90	43.90	25.00	55.40	97.70	70.71	18	1	6	11	1495	35	1204	42	64	33.35	52.50	81.00
42	64	633.00	665.00	2066.00	31.70	48.80	23.50	46.00	96.30	62.26	18	1	6	11	1241	48	1458	38	66	25.03	42.80	78.50
46	64	460.00	437.00	2239.00	25.60	51.30	17.00	30.60	93.10	46.06	18	0	6	12	827	61	1872	29	46	15.94	27.30	76.30
<b>(e) MSR = 64</b>																						

## B.8 BlowingBubbles



**Figure B.15:** Visualization of performance results on BlowingBubbles at different QP.

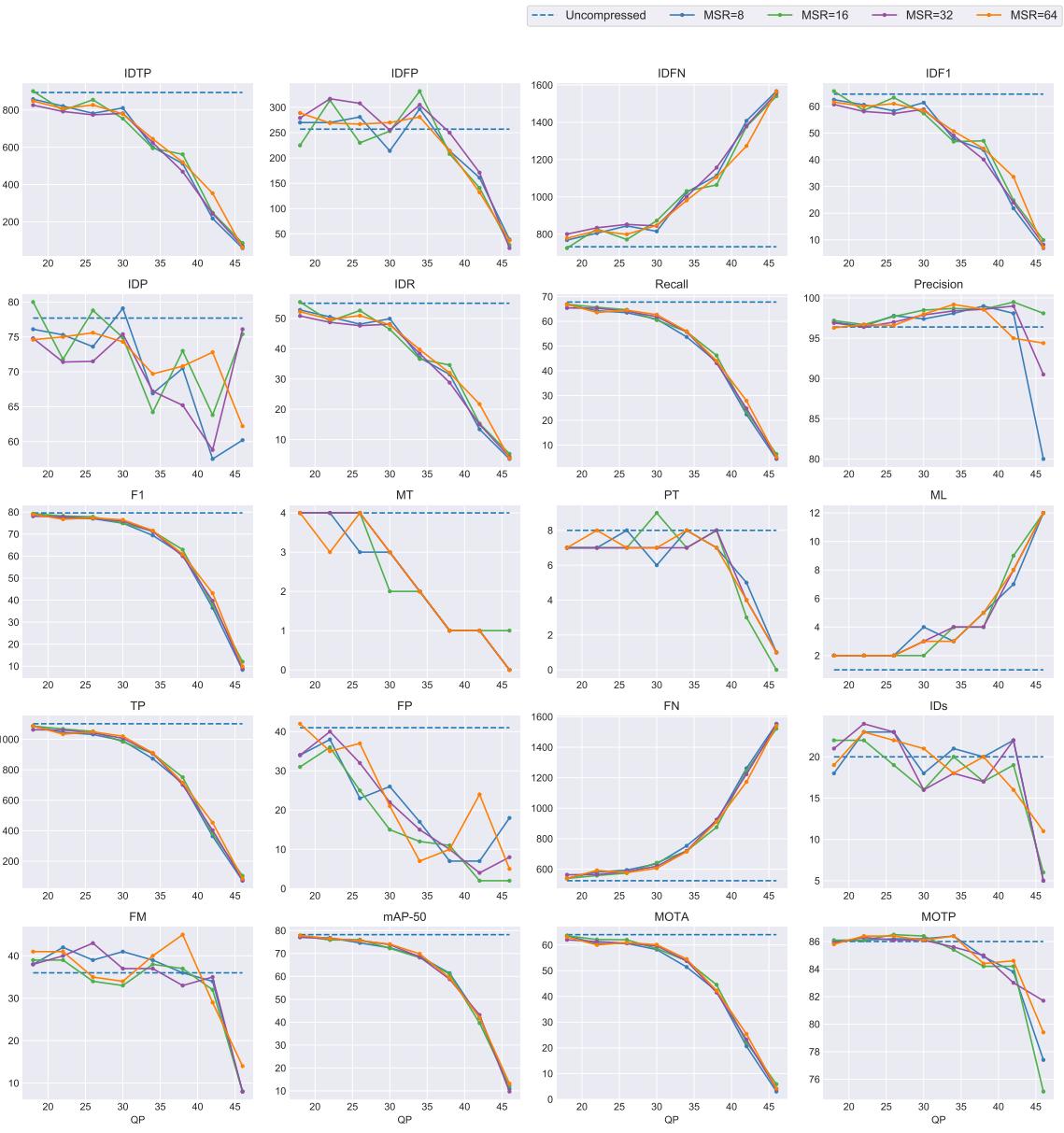


**Figure B.16:** Visualization of performance results on BlowingBubbles at different MSR.

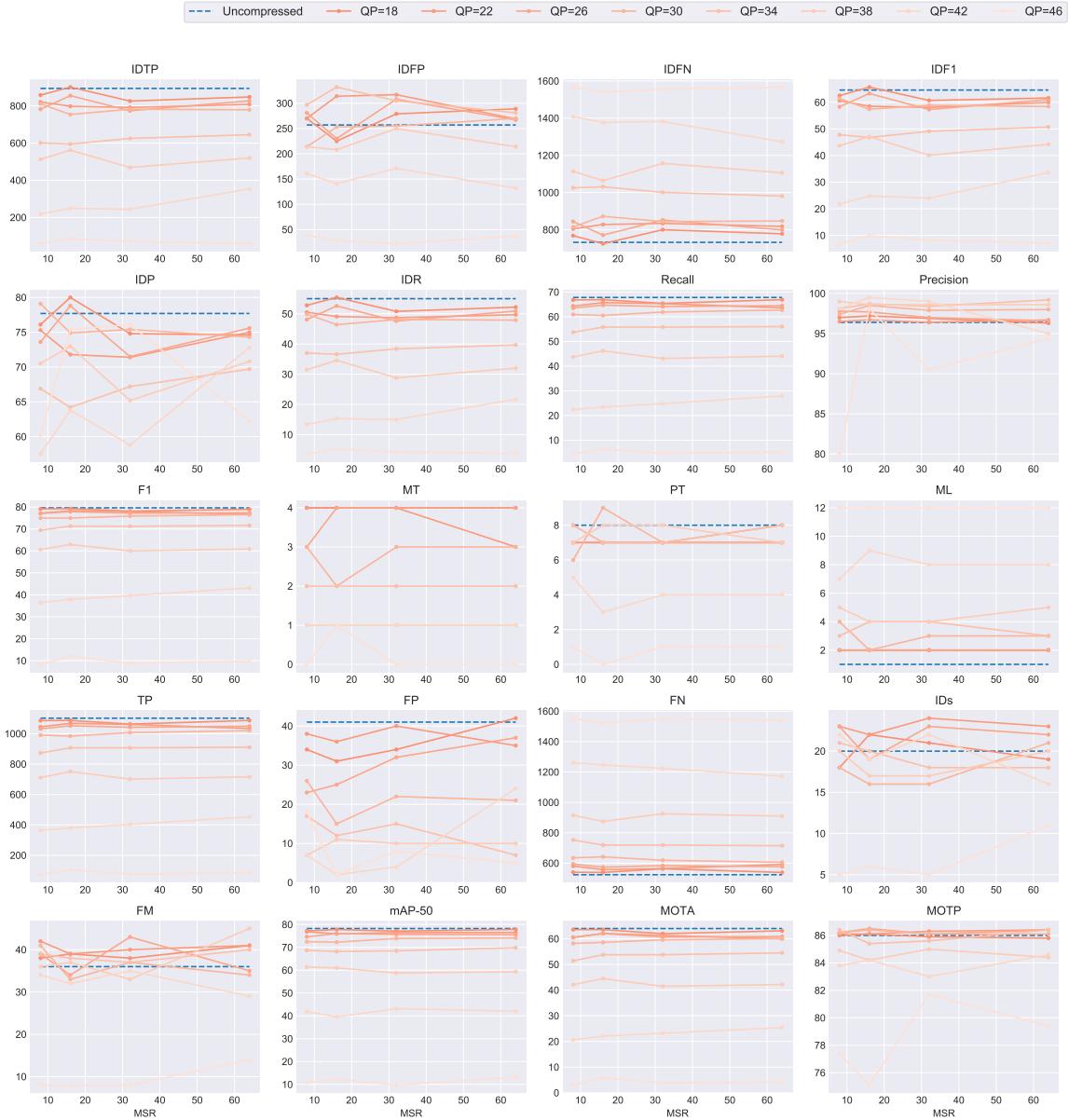
**Table B.8:** Performance results on BlowingBubbles.

QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
Uncompressed	Uncompressed	1144.00	420.00	491.00	71.50	73.10	70.00	84.20	88.00	86.06	5	3	0	2	1376	187	259	5	9	77.27	72.40	88.80
<b>(a) Uncompressed Sequence</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	8	1366.00	170.00	269.00	86.20	88.90	83.50	84.30	89.80	86.96	5	3	0	2	1379	156	256	5	10	74.77	74.50	88.90
22	8	1377.00	137.00	258.00	87.50	91.00	84.20	84.20	91.00	87.47	5	3	0	2	1377	136	258	1	4	73.19	75.80	88.90
26	8	1352.00	115.00	283.00	87.20	92.20	82.70	83.50	93.20	88.08	5	3	0	2	1366	100	269	3	7	71.23	77.20	89.00
30	8	1261.00	145.00	374.00	83.00	89.70	77.10	81.30	94.60	87.45	5	3	0	2	1329	76	306	4	7	66.85	76.40	89.10
34	8	1068.00	159.00	567.00	74.70	87.00	65.30	73.40	97.90	83.90	5	2	1	2	1200	26	435	7	9	56.87	71.40	89.00
38	8	918.00	101.00	717.00	69.20	90.10	56.10	59.80	96.00	73.69	5	2	0	3	977	41	658	3	4	34.42	57.10	89.40
42	8	736.00	208.00	899.00	57.10	78.00	45.00	55.20	95.80	70.04	5	2	0	3	903	40	732	5	10	31.57	52.50	89.50
46	8	355.00	181.00	1280.00	32.70	66.20	21.70	30.80	94.00	46.40	5	0	2	3	503	32	1132	10	13	18.57	28.20	87.40
<b>(b) MSR = 8</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	16	1260.00	289.00	375.00	79.20	81.30	77.10	85.40	90.20	87.73	5	3	0	2	1396	152	239	6	8	74.80	75.70	88.70
22	16	1372.00	147.00	263.00	87.00	90.30	83.90	84.90	91.40	88.03	5	3	0	2	1388	130	247	3	7	74.53	76.80	88.80
26	16	1346.00	116.00	289.00	87.00	92.10	82.30	83.10	93.00	87.77	5	3	0	2	1359	102	276	2	7	69.80	76.80	89.00
30	16	1289.00	94.00	346.00	85.40	93.20	78.80	81.20	96.10	88.02	5	3	0	2	1328	54	307	4	7	65.20	77.70	89.30
34	16	1104.00	135.00	531.00	76.90	89.10	67.50	72.70	96.00	82.74	5	2	1	2	1189	49	446	3	6	53.63	69.50	88.90
38	16	949.00	137.00	686.00	69.80	87.40	58.00	62.60	94.40	75.28	5	2	0	3	1024	61	611	9	9	39.61	58.30	89.20
42	16	676.00	370.00	959.00	50.40	64.60	41.30	57.90	90.60	70.65	5	2	0	3	947	98	688	5	6	31.33	51.60	88.60
46	16	340.00	103.00	1295.00	32.70	76.70	20.80	26.50	98.00	41.72	5	0	2	3	433	9	1202	11	13	17.79	25.30	86.40
<b>(c) MSR = 16</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	32	1381.00	162.00	254.00	86.90	89.50	84.50	85.60	90.70	88.08	5	3	0	2	1399	143	236	5	8	74.79	76.50	88.70
22	32	1355.00	155.00	280.00	86.20	89.70	82.90	84.30	91.30	87.66	5	3	0	2	1378	131	257	4	8	74.33	76.00	88.80
26	32	1232.00	248.00	403.00	79.10	83.20	75.40	83.70	92.50	87.88	5	3	0	2	1368	111	267	4	7	72.60	76.60	89.00
30	32	1275.00	133.00	360.00	83.80	90.60	78.00	81.80	95.00	87.91	5	3	0	2	1337	70	298	4	5	65.94	77.20	89.20
34	32	1176.00	72.00	459.00	81.60	94.20	71.90	74.10	97.20	84.09	5	2	1	2	1212	35	423	3	7	56.21	71.80	89.00
38	32	924.00	118.00	711.00	69.10	88.70	56.50	60.20	94.60	73.58	5	2	0	3	985	56	650	3	5	34.20	56.60	89.70
42	32	680.00	324.00	955.00	51.60	67.70	41.60	57.30	93.40	71.03	5	2	0	3	937	66	698	5	6	31.56	53.00	87.30
46	32	207.00	221.00	1428.00	20.10	48.40	12.70	25.70	98.40	40.76	5	0	2	3	420	7	1215	16	22	17.55	24.30	87.10
<b>(d) MSR = 32</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	64	1238.00	290.00	397.00	78.30	81.00	75.70	83.90	89.80	86.75	5	3	0	2	1371	156	264	6	9	74.13	73.90	88.80
22	64	1375.00	140.00	260.00	87.30	90.80	84.10	84.30	91.00	87.52	5	3	0	2	1378	136	257	1	7	74.40	75.90	88.80
26	64	1245.00	226.00	390.00	80.20	84.60	76.10	83.90	93.30	88.35	5	3	0	2	1372	98	263	2	4	70.57	77.80	88.90
30	64	1266.00	115.00	369.00	84.00	91.70	77.40	80.60	95.50	87.42	5	3	0	2	1318	62	317	4	6	64.36	76.60	89.20
34	64	1165.00	86.00	470.00	80.80	93.10	71.30	75.50	98.80	85.59	5	2	1	2	1235	15	400	3	8	57.76	74.40	89.30
38	64	928.00	144.00	707.00	68.60	86.60	56.80	59.90	91.40	72.37	5	2	0	3	979	92	656	2	3	32.66	54.10	89.20
42	64	834.00	186.00	801.00	62.80	81.80	51.00	57.70	92.50	71.07	5	2	0	3	943	76	692	4	8	31.78	52.80	89.00
46	64	304.00	87.00	1331.00	30.00	77.70	18.60	23.40	97.90	37.77	5	0	2	3	382	8	1253	9	10	15.49	22.30	87.50
<b>(e) MSR = 64</b>																						

## B.9 RaceHorsesD



**Figure B.17:** Visualization of performance results on RaceHorsesD at different QP.



**Figure B.18:** Visualization of performance results on RaceHorsesD at different MSR.

**Table B.9:** Performance results on RaceHorsesD.

QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
Uncompressed	Uncompressed	894.00	257.00	732.00	64.60	77.70	55.00	67.80	96.40	79.61	13	4	8	1	1102	41	524	20	36	78.23	64.00	86.00

(a) Uncompressed Sequence

QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	8	858.00	270.00	768.00	62.50	76.10	52.80	66.80	97.00	79.12	13	4	7	2	1086	34	540	18	38	77.19	63.60	86.00
22	8	821.00	270.00	805.00	60.60	75.30	50.50	64.30	96.50	77.18	13	4	7	2	1045	38	581	23	42	76.83	60.50	86.10
26	8	782.00	281.00	844.00	58.30	73.60	48.10	63.50	97.80	77.00	13	3	8	2	1032	23	594	23	39	74.61	60.60	86.20
30	8	811.00	214.00	815.00	61.40	79.10	49.90	60.90	97.40	74.94	13	3	6	4	991	26	635	18	41	72.49	58.20	86.20
34	8	601.00	297.00	1025.00	47.80	66.90	37.00	53.70	98.10	69.41	13	2	8	3	873	17	753	21	39	68.69	51.40	86.40
38	8	512.00	214.00	1114.00	43.70	70.50	31.50	43.70	99.00	60.63	13	1	7	5	711	7	915	20	36	61.38	42.10	84.90
42	8	218.00	161.00	1408.00	21.80	57.50	13.40	22.40	98.10	36.47	13	1	5	7	364	7	1262	22	34	41.83	20.60	83.80
46	8	59.00	39.00	1567.00	6.90	60.20	3.60	4.40	80.00	8.34	13	0	1	12	72	18	1554	5	8	10.88	3.00	77.40

(b) MSR = 8

QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	16	901.00	225.00	725.00	65.70	80.00	55.40	66.90	97.20	79.25	13	4	7	2	1087	31	539	22	39	77.80	63.60	86.10
22	16	798.00	314.00	828.00	58.50	71.80	49.10	65.70	96.70	78.24	13	4	7	2	1068	36	558	22	39	75.90	62.10	86.10
26	16	855.00	230.00	771.00	63.30	78.80	52.60	64.70	97.70	77.85	13	4	7	2	1052	25	574	19	34	76.01	62.00	86.50
30	16	754.00	253.00	872.00	57.40	74.90	46.40	60.50	98.50	74.96	13	2	9	2	984	15	642	16	33	72.26	58.60	86.40
34	16	595.00	332.00	1031.00	46.80	64.20	36.60	55.80	98.70	71.29	13	2	7	4	907	12	719	20	38	68.19	53.80	85.40
38	16	562.00	208.00	1064.00	47.10	73.00	34.60	46.20	98.60	62.92	13	1	8	4	751	11	875	17	37	61.00	44.50	84.20
42	16	249.00	141.00	1377.00	24.80	63.80	15.30	23.40	99.50	37.89	13	1	3	9	380	2	1246	19	32	39.59	22.10	84.20
46	16	86.00	28.00	1540.00	9.90	75.40	5.30	6.40	98.10	12.02	13	1	0	12	104	2	1522	6	8	12.17	5.90	75.10

(c) MSR = 16

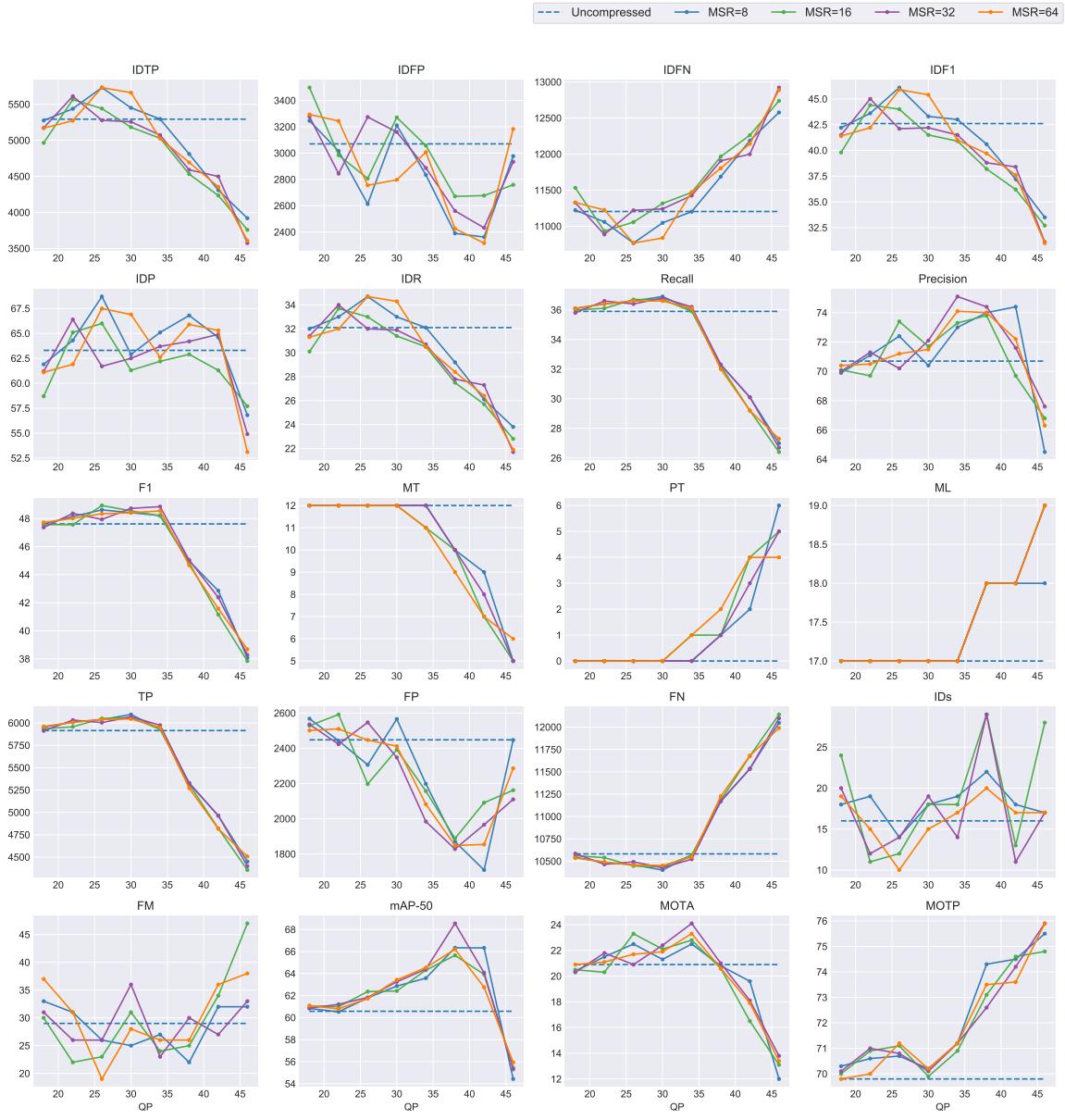
QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	32	826.00	279.00	800.00	60.70	74.80	50.80	65.40	96.90	78.09	13	4	7	2	1063	34	563	21	38	77.09	62.00	85.90
22	32	792.00	317.00	834.00	58.50	71.40	48.70	65.30	96.40	77.86	13	4	7	2	1061	40	565	24	40	76.43	61.30	86.30
26	32	774.00	308.00	852.00	57.30	71.50	47.60	64.10	97.00	77.19	13	4	7	2	1042	32	584	23	43	75.69	60.70	86.10
30	32	782.00	255.00	844.00	58.90	75.40	48.10	61.90	97.90	75.84	13	3	7	3	1007	22	619	16	37	73.86	59.60	86.10
34	32	625.00	305.00	1001.00	49.10	67.20	38.40	55.80	98.40	71.22	13	2	7	4	907	15	719	18	37	68.47	53.80	85.60
38	32	469.00	250.00	1157.00	40.10	65.20	28.80	43.10	98.60	59.98	13	1	8	4	701	10	925	17	33	58.77	41.50	85.00
42	32	244.00	171.00	1382.00	24.00	58.80	15.00	24.80	99.00	39.66	13	1	4	8	403	4	1223	22	35	43.15	23.20	83.00
46	32	70.00	22.00	1556.00	8.20	76.10	4.30	4.70	90.50	8.94	13	0	1	12	76	8	1550	5	8	9.69	3.90	81.70

(d) MSR = 32

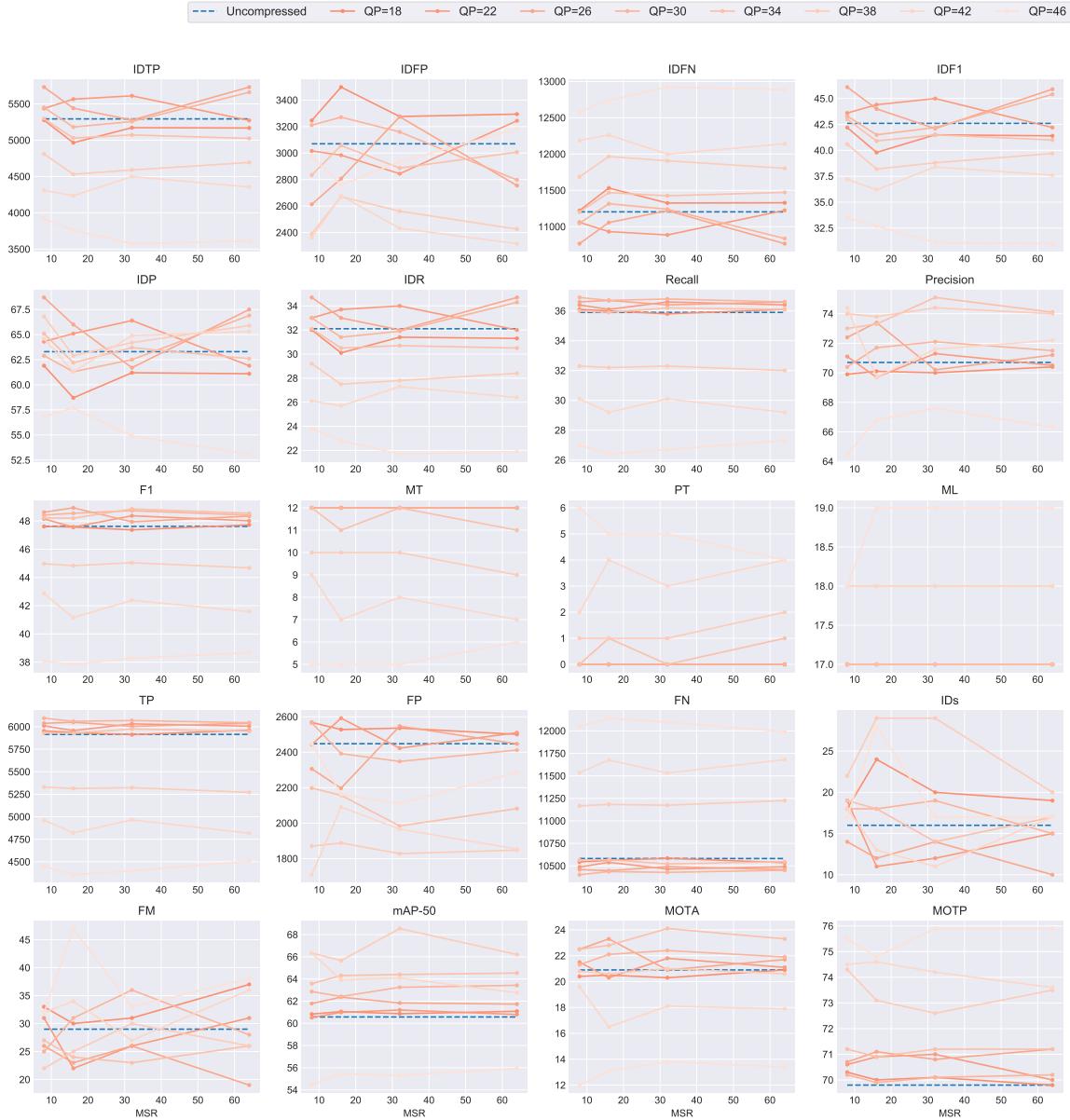
QP	MSR	IDTP	IDFP	IDFN	IDFI	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	64	848.00	289.00	778.00	61.60	74.60	52.20	66.90	96.30	78.95	13	4	7	2	1087	42	539	19	41	77.85	63.10	85.80
22	64	808.00	269.00	818.00	60.00	75.00	49.70	63.60	96.70	76.73	13	3	8	2	1034	35	592	23	41	76.52	60.00	86.40
26	64	827.00	267.00	799.00	61.00	75.60	50.90	64.50	96.60	77.35	13	4	7	2	1049	37	577	22	35	75.55	60.90	86.40
30	64	779.00	270.00	847.00	58.40	74.30	47.90	62.70	98.00	76.47	13	3	7	3	1020	21	606	21	34	74.09	60.10	86.10
34	64	645.00	281.00	981.00	50.70	69.70	39.70	56.00	99.20	71.59	13	2	8	3	911	7	715	18	40	69.83	54.50	86.40
38	64	520.00	214.00	1106.00	44.20	70.80	32.00	44.00	98.60	60.85	13	1	7	5	716	10	910	20	45	59.33	42.20	84.40
42	64	353.00	132.00	1273.00	33.60	72.80	21.70	27.90	95.00	43.13	13	1	4	8	453	24	1173	16	29	42.03	25.40	84.60
46	64	61.00	37.00	1565.00	7.10	62.20	3.80	5.20	94.40	9.86	13	0	1	12	85	5	1541	11	14	13.18	4.20	79.40

(e) MSR = 64

## B.10 FourPeople



**Figure B.19:** Visualization of performance results on FourPeople at different QP.



**Figure B.20:** Visualization of performance results on FourPeople at different MSR.

**Table B.10:** Performance results on FourPeople.

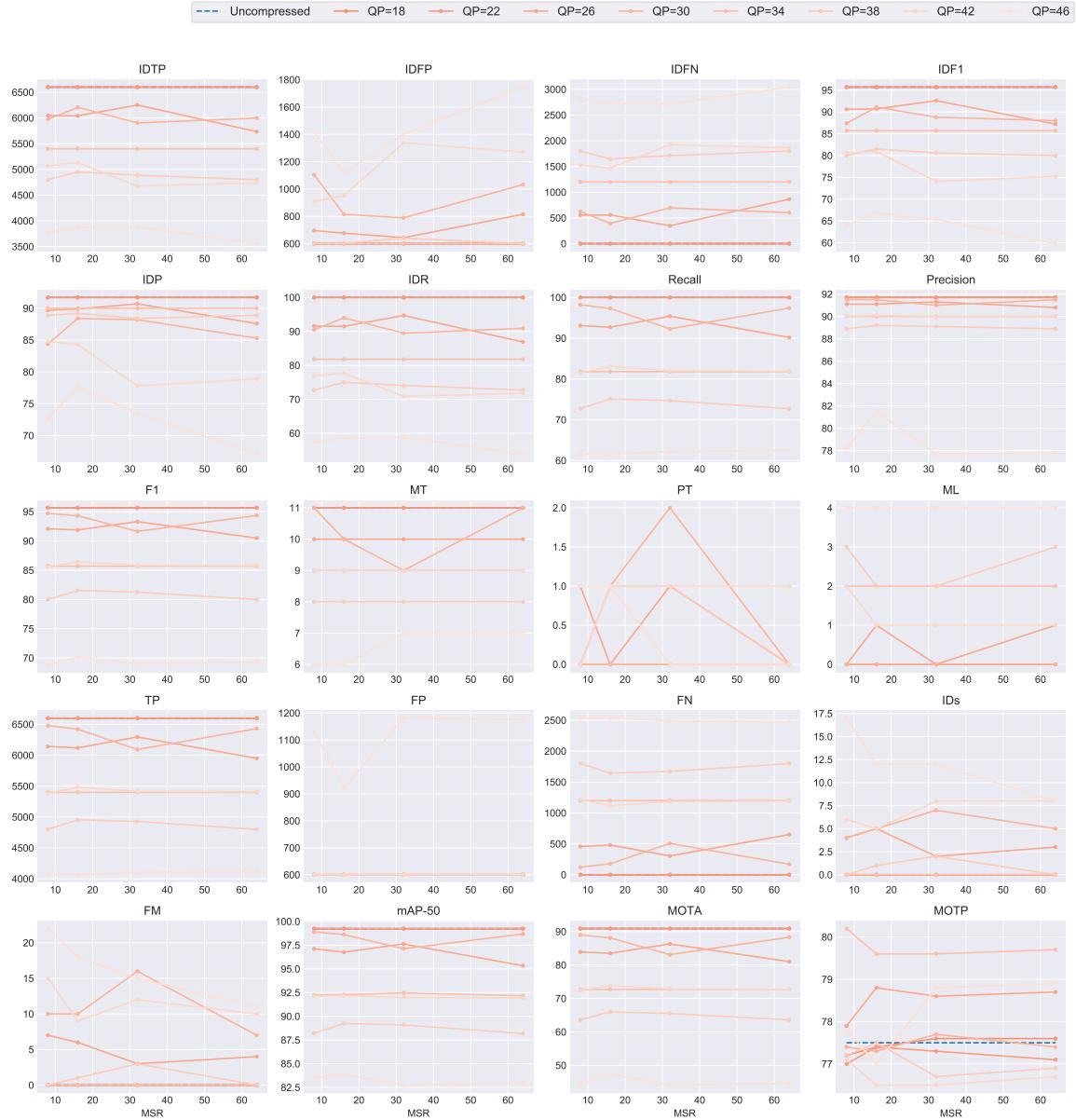
QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
Uncompressed	Uncompressed	5293.00	3070.00	11204.00	42.60	63.30	32.10	35.90	70.70	47.62	29	12	0	17	5915	2448	10582	16	29	60.58	20.90	69.80
<b>(a) Uncompressed Sequence</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	8	5275.00	3246.00	11222.00	42.20	61.90	32.00	36.10	69.90	47.61	29	12	0	17	5935	2568	10544	18	33	60.83	20.40	70.30
22	8	5437.00	3016.00	11060.00	43.60	64.30	33.00	36.40	71.10	48.15	29	12	0	17	6010	2443	10487	19	31	60.53	21.50	70.60
26	8	5731.00	2613.00	10766.00	46.10	68.70	34.70	36.60	72.40	48.62	29	12	0	17	6038	2306	10459	14	26	61.79	22.50	70.70
30	8	5450.00	3211.00	11047.00	43.30	62.90	33.00	36.90	70.40	48.42	29	12	0	17	6095	2566	10402	18	25	62.86	21.30	70.20
34	8	5295.00	2834.00	11202.00	43.00	65.10	32.10	36.00	73.00	48.22	29	12	0	17	5931	2198	10566	19	27	63.58	22.50	71.20
38	8	4810.00	2390.00	11687.00	40.60	66.80	29.20	32.30	74.00	44.97	29	10	1	18	5330	1870	11167	22	22	66.34	20.80	74.30
42	8	4309.00	2362.00	12188.00	37.20	64.60	26.10	30.10	74.40	42.86	29	9	2	18	4962	1709	11535	18	32	66.33	19.60	74.50
46	8	3920.00	2977.00	12577.00	33.50	56.80	23.80	27.00	64.50	38.07	29	5	6	18	4450	2447	12047	17	32	54.43	12.00	75.50
<b>(b) MSR = 8</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	16	4965.00	3498.00	11532.00	39.80	58.70	30.10	36.00	70.10	47.57	29	12	0	17	5935	2528	10562	24	30	61.06	20.50	70.00
22	16	5565.00	2984.00	10932.00	44.40	65.10	33.70	36.10	69.70	47.56	29	12	0	17	5957	2592	10540	11	22	60.99	20.30	70.90
26	16	5440.00	2807.00	11057.00	44.00	66.00	33.00	36.70	73.40	48.93	29	12	0	17	6050	2197	10447	12	23	62.36	23.30	71.10
30	16	5181.00	3271.00	11316.00	41.50	61.30	31.40	36.70	71.70	48.55	29	12	0	17	6060	2392	10437	18	31	62.42	22.10	69.90
34	16	5029.00	3057.00	11468.00	40.90	62.20	30.50	35.90	73.30	48.20	29	11	1	17	5929	2157	10568	18	24	64.31	22.80	70.90
38	16	4531.00	2671.00	11966.00	38.20	62.90	27.50	32.20	73.80	44.84	29	10	1	18	5314	1888	11183	29	25	65.65	20.60	73.10
42	16	4235.00	2677.00	12262.00	36.20	61.30	25.70	29.20	69.70	41.16	29	7	4	18	4821	2091	11676	13	34	63.93	16.50	74.60
46	16	3759.00	2759.00	12738.00	32.70	57.70	22.80	26.40	66.80	37.84	29	5	5	19	4356	2162	12141	28	47	55.45	13.10	74.80
<b>(c) MSR = 16</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	32	5172.00	3276.00	11325.00	41.50	61.20	31.40	35.80	70.00	47.37	29	12	0	17	5912	2536	10585	20	31	60.86	20.30	70.10
22	32	5611.00	2844.00	10886.00	45.00	66.40	34.00	36.60	71.30	48.37	29	12	0	17	6032	2423	10465	12	26	61.20	21.80	71.00
26	32	5277.00	3274.00	11220.00	42.10	61.70	32.00	36.40	70.20	47.94	29	12	0	17	6004	2547	10493	14	26	61.84	20.90	70.80
30	32	5257.00	3160.00	11240.00	42.20	62.50	31.90	36.80	72.10	48.73	29	12	0	17	6069	2348	10428	19	36	63.25	22.40	70.10
34	32	5072.00	2887.00	11425.00	41.50	63.70	30.70	36.20	75.10	48.85	29	12	0	17	5975	1984	10522	14	23	64.40	24.10	71.20
38	32	4590.00	2561.00	11907.00	38.80	64.20	27.80	32.30	74.40	45.04	29	10	1	18	5323	1828	11174	29	30	68.55	21.00	72.60
42	32	4499.00	2432.00	11998.00	38.40	64.90	27.30	30.10	71.60	42.38	29	8	3	18	4965	1966	11532	11	27	64.07	18.10	74.20
46	32	3575.00	2934.00	12922.00	31.10	54.90	21.70	26.70	67.60	38.28	29	5	5	19	4399	2110	12098	17	33	55.31	13.80	75.90
<b>(d) MSR = 32</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	64	5168.00	3294.00	11329.00	41.40	61.10	31.30	36.10	70.40	47.73	29	12	0	17	5961	2501	10536	19	37	61.09	20.90	69.80
22	64	5272.00	3244.00	11225.00	42.20	61.90	32.00	36.40	70.50	48.01	29	12	0	17	6006	2510	10491	15	31	60.80	21.10	70.00
26	64	5731.00	2755.00	10766.00	45.90	67.50	34.70	36.60	71.20	48.35	29	12	0	17	6039	2447	10458	10	19	61.73	21.70	71.20
30	64	5660.00	2797.00	10837.00	45.40	66.90	34.30	36.60	71.50	48.42	29	12	0	17	6045	2412	10452	15	28	63.43	21.90	70.20
34	64	5025.00	3007.00	11472.00	41.00	62.60	30.50	36.10	74.10	48.55	29	11	1	17	5950	2082	10547	17	26	64.53	23.30	71.20
38	64	4693.00	2426.00	11804.00	39.70	65.90	28.40	32.00	74.00	44.68	29	9	2	18	5271	1848	11226	20	26	66.20	20.60	73.50
42	64	4355.00	2316.00	12142.00	37.60	65.30	26.40	29.20	72.20	41.58	29	7	4	18	4817	1854	11680	17	36	62.76	17.90	73.60
46	64	3610.00	3183.00	12887.00	31.00	53.10	21.90	27.30	66.30	38.68	29	6	4	19	4507	2286	11990	17	38	55.93	13.40	75.90

**(e) MSR = 64**

## B.11 Johnny



**Figure B.21:** Visualization of performance results on Johnny at different QP.



**Figure B.22:** Visualization of performance results on Johnny at different MSR.

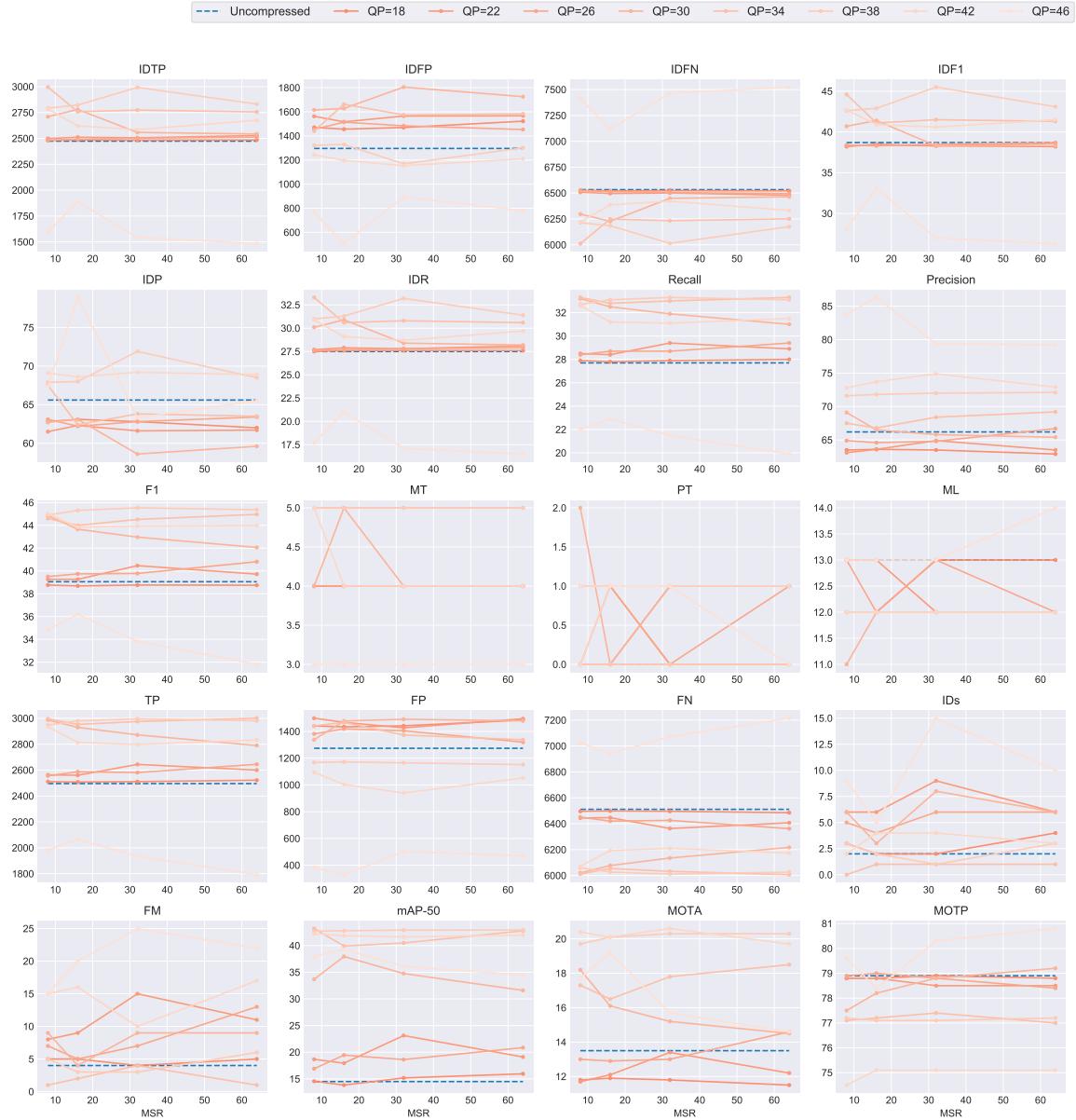
**Table B.11:** Performance results on Johnny.

QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
Uncompressed	Uncompressed	6600.00	600.00	0.00	95.70	91.70	100.00	100.00	91.70	95.67	11	11	0	0	6600	600	0	0	0	99.24	90.90	77.50
<b>(a) Uncompressed Sequence</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	8	6600.00	600.00	0.00	95.70	91.70	100.00	100.00	91.70	95.67	11	11	0	0	6600	600	0	0	0	99.25	90.90	77.20
22	8	6600.00	600.00	0.00	95.70	91.70	100.00	100.00	91.70	95.67	11	11	0	0	6600	600	0	0	0	99.24	90.90	77.00
26	8	6047.00	696.00	553.00	90.60	89.70	91.60	93.10	91.10	92.09	11	10	1	0	6143	600	457	4	7	97.11	83.90	77.90
30	8	5975.00	1104.00	625.00	87.40	84.40	90.50	98.20	91.50	94.73	11	11	0	0	6479	600	121	4	10	98.90	89.00	77.40
34	8	5400.00	600.00	1200.00	85.70	90.00	81.80	81.80	90.00	85.70	11	9	0	2	5400	600	1200	0	0	92.24	72.70	80.20
38	8	4800.00	600.00	1800.00	80.00	88.90	72.70	72.70	88.90	79.99	11	8	0	3	4800	600	1800	0	0	88.21	63.60	77.20
42	8	5073.00	909.00	1527.00	80.60	84.80	76.90	81.50	90.00	85.54	11	9	0	2	5382	600	1218	6	15	92.13	72.40	77.10
46	8	3782.00	1417.00	2818.00	64.10	72.70	57.30	61.60	78.20	68.91	11	6	1	4	4067	1132	2533	17	22	83.50	44.20	77.80
<b>(b) MSR = 8</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	16	6600.00	600.00	0.00	95.70	91.70	100.00	100.00	91.70	95.67	11	11	0	0	6600	600	0	0	0	99.23	90.90	77.40
22	16	6600.00	600.00	0.00	95.70	91.70	100.00	100.00	91.70	95.67	11	11	0	0	6600	600	0	0	0	99.25	90.90	77.40
26	16	6042.00	677.00	558.00	90.70	89.90	91.50	92.70	91.10	91.89	11	10	0	1	6119	600	481	5	6	96.76	83.50	78.80
30	16	6207.00	815.00	393.00	91.10	88.40	94.00	97.30	91.50	94.31	11	10	1	0	6422	600	178	5	10	98.62	88.10	77.30
34	16	5400.00	600.00	1200.00	85.70	90.00	81.80	81.80	90.00	85.70	11	9	0	2	5400	600	1200	0	0	92.30	72.70	79.60
38	16	4953.00	602.00	1647.00	81.50	89.20	75.00	75.10	89.20	81.54	11	8	1	2	4955	600	1645	1	1	89.24	66.00	77.50
42	16	5130.00	953.00	1470.00	80.90	84.30	77.70	83.10	90.10	86.46	11	9	1	1	5483	600	1117	5	9	92.18	73.90	76.50
46	16	3873.00	1116.00	2727.00	66.80	77.60	58.70	61.60	81.50	70.17	11	6	1	4	4066	923	2534	12	18	83.87	47.40	77.00
<b>(c) MSR = 16</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	32	6600.00	600.00	0.00	95.70	91.70	100.00	100.00	91.70	95.67	11	11	0	0	6600	600	0	0	0	99.23	90.90	77.60
22	32	6600.00	600.00	0.00	95.70	91.70	100.00	100.00	91.70	95.67	11	11	0	0	6600	600	0	0	0	99.22	90.90	77.30
26	32	6252.00	644.00	348.00	92.60	90.70	94.70	95.40	91.30	93.30	11	10	1	0	6296	600	304	2	3	97.63	86.30	78.60
30	32	5904.00	789.00	696.00	88.80	88.20	89.50	92.30	91.00	91.65	11	9	2	0	6093	600	507	7	16	97.12	83.10	77.70
34	32	5400.00	600.00	1200.00	85.70	90.00	81.80	81.80	90.00	85.70	11	9	0	2	5400	600	1200	0	0	92.46	72.70	79.60
38	32	4888.00	640.00	1712.00	80.60	88.40	74.10	74.70	89.10	81.27	11	8	1	2	4928	600	1672	2	3	89.09	65.50	76.70
42	32	4677.00	1338.00	1923.00	74.10	77.80	70.90	82.00	90.00	85.81	11	9	1	1	5415	600	1185	8	12	91.99	72.80	76.50
46	32	3886.00	1399.00	2714.00	65.40	73.50	58.90	62.20	77.70	69.09	11	7	0	4	4104	1181	2496	12	15	82.75	44.10	78.80
<b>(d) MSR = 32</b>																						
QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	64	6600.00	600.00	0.00	95.70	91.70	100.00	100.00	91.70	95.67	11	11	0	0	6600	600	0	0	0	99.24	90.90	77.60
22	64	6600.00	600.00	0.00	95.70	91.70	100.00	100.00	91.70	95.67	11	11	0	0	6600	600	0	0	0	99.25	90.90	77.10
26	64	5735.00	815.00	865.00	87.20	87.60	86.90	90.20	90.80	90.50	11	10	0	1	5950	600	650	3	4	95.34	81.00	78.70
30	64	598.00	1032.00	602.00	88.00	85.30	90.90	97.40	91.50	94.36	11	11	0	0	6430	600	170	5	7	98.68	88.30	77.40
34	64	5400.00	600.00	1200.00	85.70	90.00	81.80	81.80	90.00	85.70	11	9	0	2	5400	600	1200	0	0	92.17	72.70	79.70
38	64	4800.00	600.00	1800.00	80.00	88.90	72.70	72.70	88.90	79.99	11	8	0	3	4800	600	1800	0	0	88.18	63.60	76.90
42	64	4738.00	1270.00	1862.00	75.20	78.90	71.80	81.90	90.00	85.76	11	9	1	1	5408	600	1192	8	10	91.92	72.70	76.70
46	64	3557.00	1746.00	3043.00	59.80	67.10	53.90	62.50	77.80	69.32	11	7	0	4	4127	1176	2473	8	11	82.92	44.60	78.90
<b>(e) MSR = 64</b>																						

## B.12 KristenAndSara



**Figure B.23:** Visualization of performance results on KristenAndSara at different QP.



**Figure B.24:** Visualization of performance results on KristenAndSara at different MSR.

**Table B.12:** Performance results on KristenAndSara.

QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
Uncompressed	Uncompressed	2474.00	1296.00	6533.00	38.70	65.60	27.50	27.70	66.20	39.06	17	4	0	13	2496	1274	6511	2	4	14.50	13.50	78.90

(a) Uncompressed Sequence

QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	8	2481.00	1471.00	6526.00	38.30	62.80	27.50	27.90	63.50	38.77	17	4	1	12	2511	1441	6496	3	5	14.55	11.80	78.80
22	8	2499.00	1563.00	6508.00	38.20	61.50	27.70	28.50	63.10	39.27	17	4	0	13	2563	1499	6444	6	8	18.67	11.70	78.80
26	8	2483.00	1455.00	6524.00	38.40	63.10	27.60	28.40	64.90	39.51	17	4	0	13	2556	1382	6451	5	7	16.91	13.00	78.90
30	8	2710.00	1615.00	6297.00	40.70	62.70	30.10	33.20	69.10	44.85	17	4	2	11	2987	1338	6020	6	9	33.68	18.20	77.50
34	8	2996.00	1441.00	6011.00	44.60	67.50	33.30	33.30	67.50	44.60	17	5	0	12	2996	1441	6011	0	1	43.19	17.30	77.10
38	8	2793.00	1322.00	6214.00	42.60	67.90	31.00	32.70	71.60	44.90	17	5	0	12	2947	1168	6060	3	5	42.70	19.70	77.20
42	8	2786.00	1243.00	6221.00	42.70	69.10	30.90	32.60	72.80	45.03	17	5	0	12	2935	1094	6072	2	15	42.22	20.40	74.50
46	8	1595.00	770.00	7412.00	28.10	67.40	17.70	22.00	83.80	34.85	17	3	1	13	1982	383	7025	9	15	37.95	17.70	79.60

(b) MSR = 8

QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	16	2486.00	1456.00	6521.00	38.40	63.10	27.60	27.80	63.60	38.69	17	4	1	12	2508	1434	6499	2	5	13.84	11.90	78.80
22	16	2511.00	1517.00	6496.00	38.50	62.30	27.90	28.40	63.60	39.27	17	4	0	13	2560	1468	6447	6	9	17.94	12.10	78.80
26	16	2492.00	1514.00	6515.00	38.30	62.20	27.70	28.70	64.60	39.74	17	4	1	12	2587	1419	6420	4	5	19.46	12.90	79.00
30	16	2780.00	1628.00	6227.00	41.40	63.10	30.90	32.50	66.50	43.66	17	5	0	12	2930	1478	6077	3	4	37.95	16.10	78.20
34	16	2758.00	1663.00	6249.00	41.10	62.40	30.60	32.80	66.80	44.00	17	5	0	12	2953	1468	6054	1	2	39.91	16.50	77.20
38	16	2823.00	1329.00	6184.00	42.90	68.00	31.30	33.10	71.80	45.31	17	5	0	12	2980	1172	6027	2	3	42.79	20.10	77.10
42	16	2621.00	1197.00	6386.00	40.90	68.60	29.10	31.20	73.70	43.84	17	4	1	12	2814	1004	6193	4	16	41.85	20.10	75.10
46	16	1888.00	506.00	7119.00	33.10	78.90	21.00	22.90	86.30	36.20	17	3	1	13	2065	329	6942	5	20	39.29	19.20	78.40

(c) MSR = 16

QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	32	2483.00	1470.00	6524.00	38.30	62.80	27.60	27.90	63.50	38.77	17	4	0	13	2511	1442	6496	2	4	15.18	11.80	78.90
22	32	2506.00	1565.00	6501.00	38.30	61.60	27.80	29.40	64.90	40.47	17	4	1	12	2644	1427	6363	9	15	23.13	13.40	78.50
26	32	2502.00	1484.00	6505.00	38.50	62.80	27.80	28.70	64.80	39.78	17	4	0	13	2581	1405	6426	6	7	18.62	13.00	78.80
30	32	2558.00	1805.00	6449.00	38.30	58.60	28.40	31.90	65.80	42.97	17	4	1	12	2872	1491	6135	8	9	34.75	15.20	78.80
34	32	2774.00	1574.00	6233.00	41.50	63.80	30.80	33.00	68.40	44.52	17	5	0	12	2975	1373	6032	1	4	40.48	17.80	77.40
38	32	2993.00	1168.00	6014.00	45.50	71.90	33.20	33.30	72.00	45.54	17	5	0	12	2996	1165	6011	1	3	42.89	20.30	77.10
42	32	2585.00	1153.00	6422.00	40.60	69.20	28.70	31.10	74.90	43.95	17	4	1	12	2798	940	6209	4	10	41.62	20.60	75.10
46	32	1543.00	894.00	7464.00	27.00	63.30	17.10	21.50	79.40	33.84	17	3	1	13	1935	502	7072	15	25	36.09	15.70	80.30

(d) MSR = 32

QP	MSR	IDTP	IDFP	IDFN	IDF1	IDP	IDR	Recall	Precision	F1	GT	MT	PT	ML	TP	FP	FN	IDs	FM	mAP-50	MOTA	MOTP
18	64	2485.00	1522.00	6522.00	38.20	62.00	27.60	28.00	62.90	38.75	17	4	0	13	2522	1485	6485	4	5	15.95	11.50	78.80
22	64	2528.00	1567.00	6479.00	38.60	61.70	28.10	28.90	63.50	39.72	17	4	1	12	2600	1495	6407	6	11	19.11	12.20	78.50
26	64	2512.00	1453.00	6495.00	38.70	63.40	27.90	29.40	66.70	40.81	17	4	1	12	2645	1320	6362	6	13	20.86	14.60	78.40
30	64	2543.00	1726.00	6464.00	38.30	59.60	28.20	31.00	65.40	42.06	17	4	1	12	2790	1479	6217	6	9	31.60	14.50	79.20
34	64	2756.00	1583.00	6251.00	41.30	63.50	30.60	33.30	69.20	44.96	17	5	0	12	3001	1338	6006	1	1	42.77	18.50	77.00
38	64	2832.00	1301.00	6175.00	43.10	68.50	31.40	33.10	72.10	45.37	17	5	0	12	2981	1152	6026	3	6	42.92	20.30	77.20
42	64	2675.00	1210.00	6332.00	41.50	68.90	29.70	31.50	72.90	43.99	17	4	1	12	2833	1052	6174	3	17	41.95	19.70	75.10
46	64	1483.00	780.00	7524.00	26.30	65.50	16.50	19.90	79.20	31.81	17	3	0	14	1792	471	7215	10	22	34.50	14.60	80.80

(e) MSR = 64