*electronics*

MDPI

*Communication*

# Objective Video Quality Assessment Method for Object Recognition Tasks

**Mikołaj Leszczuk [1,*], Lucjan Janowski [1], Jakub Nawała [2] and Atanas Boev [3]**

[1] AGH University of Krakow, al. Adama Mickiewicza 30, 30-059 Kraków, Poland; qoe@agh.edu.pl
[2] Department of Electrical Electronic Engineering, University of Bristol, Bristol BS8 1QU, UK; jakub.nawala@bristol.ac.uk
[3] Huawei Technologies Dusseldorf GmbH, 40549 Düsseldorf, Germany; atanas.boev@huawei.com
* Correspondence: mikolaj.leszczuk@agh.edu.pl; Tel.: +48-12-617-3599

**Abstract:** In the field of video quality assessment for object recognition tasks, accurately predicting the impact of different quality factors on recognition algorithms remains a significant challenge. Our study introduces a novel evaluation framework designed to address this gap by focussing on machine vision rather than human perceptual quality metrics. We used advanced machine learning models and custom Video Quality Indicators to enhance the predictive accuracy of object recognition performance under various conditions. Our results indicate a model performance, achieving a mean square error (MSE) of 672.4 and a correlation coefficient of 0.77, which underscores the effectiveness of our approach in real-world scenarios. These findings highlight not only the robustness of our methodology but also its potential applicability in critical areas such as surveillance and telemedicine.

**Keywords:** video quality assessment; object recognition; TRVs (Target Recognition Videos); machine vision; random forest regressor; video quality indicators (VQIs); SRC (Source Reference Circuits); HRC (Hypothetical Reference Circuits); datasets; performance prediction



**Citation:** Leszczuk, M.; Janowski, L.; Nawała, J.; Boev, A. Objective Video Quality Assessment Method for Object Recognition Tasks. *Electronics* **2024**, *13*, 1750. https://doi.org/10.3390/electronics13091750

Academic Editors: Miguel García-Torres and Francisco A. Gómez Vela

Received: 22 March 2024
Revised: 18 April 2024
Accepted: 26 April 2024
Published: 1 May 2024



## 1. Introduction

The evaluation of video quality varies significantly between applications, highlighting a crucial divergence in the evaluation criteria. While entertainment videos prioritise viewer satisfaction, the quality assessment in Target Recognition Videos (TRVs) demands a focus on operational effectiveness. This is particularly critical in applications such as video surveillance, telemedicine, and fire safety, where accurate recognition of specific details can save lives. Existing quality predictors, primarily grounded in subjective evaluations, do not align with the intricate demands of recognition tasks. These conventional methods overlook critical aspects such as fluctuating lighting conditions, blurred motion, and obstructions, which play a pivotal role in applications such as surveillance and Automatic Number-Plate Recognition (ANPR) [1]. The shortcomings of these predictors in handling such variables can significantly impede their reliability and effectiveness, highlighting the need for our method, which is designed to robustly respond to these specific challenges. This discrepancy between subjective quality evaluations and the actual effectiveness of TRVs in practical scenarios underscores a significant lack in the present methodology to assess video quality in these situations [1].

Traditional quality assessment methods, such as Full Reference (FR) and No Reference (NR) metrics, while effective for conventional videos, fail to account for specific features crucial to TRVs' performance. These standard methods frequently ignore essential factors, such as the visibility of targets under diverse conditions, which are critical for precise recognition. Consequently, there is a pronounced deficiency in research, especially with respect to the objective assessment of TRVs for both manual and automated recognition tasks.

Despite the advances in video quality assessment, existing methodologies remain insufficiently tuned to the specific demands of TRVs, particularly in critical applications

such as surveillance and telemedicine. This misalignment poses significant challenges in ensuring operational effectiveness and reliability in environments where precision is paramount.

In the realm of autonomous driving, accurate recognition of objects in diverse driving environments is crucial for safety and efficiency. ITU-T Study Group 12 has recently launched a work item called 'P.obj-recog', which focusses on developing an object-recognition-rate-estimation model specifically for surveillance video in autonomous vehicles. This model evaluates the effectiveness of object recognition systems by considering various video and network parameters such as resolution, packet loss, and vehicle speed. Such advances are vital because they provide a structured approach to improving the reliability of object recognition systems, which are critical for the navigation of complex urban environments [2,3]. This initiative reflects the growing trend to integrate sophisticated machine learning algorithms to enhance the perceptual capabilities of automated systems, ensuring higher safety and operational efficiency in autonomous driving scenarios [4].

To better address these challenges, we propose the following research questions aimed at exploring and enhancing the methodology for assessing video quality in TRVs:

1. How can video quality assessment be effectively tailored to meet the specific needs of TRVs in high-stakes environments?
2. What role do advanced machine learning algorithms play in enhancing the feature detection capabilities necessary for TRVs?

We hypothesise that a method that integrates advanced machine learning algorithms and enhanced feature detection capabilities will significantly improve the precision and reliability of video quality assessments in TRVs, outperforming traditional methods in both objective and practical terms.

To address these gaps, our proposed method integrates advanced machine learning algorithms with enhanced feature detection capabilities, ensuring that the assessment of video quality in TRVs is both comprehensive and precise. This integration allows for a nuanced understanding of video content in varied environmental conditions, significantly improving the accuracy and reliability of quality assessments in critical recognition tasks.

In this letter paper, we endeavour to bridge the noted gap by presenting an objective evaluation methodology tailored for TRVs. This initiative is part of our ongoing effort, as the results presented are based on the methodology detailed in our article [5] for face recognition and are further elaborated in our article [1], where we transition from ANPR to object recognition. Our methodology is based on a dataset aimed at object recognition, addressing diverse real-world challenges, such as occlusion and inadequate lighting. Using this dataset, we design, develop, and evaluate a system equipped to predict the performance of machine vision algorithms by analysing the quality of the incoming TRVs. Our ultimate objective is to validate the possibility of creating precise models capable of forecasting the efficiency of TRV processing pipelines across a wide array of scenarios.

Our research introduces a notable departure from contemporary studies in our field. For instance, a pioneering approach by Shi et al. for assessing video quality in impaired conditions and their contribution to a relevant dataset [6] stand out. Their focus lies in the realm of public safety, leveraging their success in the NIST challenge, which diverges from our emphasis on object recognition tasks. Exploring the realm of TRV through the lens of selective laser melting process identification also highlights the varied applications of TRV, albeit distinct from our approach [7]. The domain of laparoscopic image quality, advanced by Khan et al., reflects the growing interest in the evaluation of TRV quality from a medical perspective [8]. An encrypted image database by Hofbauer et al. provides insight into image encryption distortions, yet it operates within a different context compared with our investigations [9].

The deployment of a cascaded deep neural network by Wu et al. to blindly predict image quality marks a significant advancement in the field, despite the utilisation of a different methodology in the use of TRV for specific tasks compared with our study [10]. Similarly, the methodologies proposed by Oszust for the blind image quality assessment [11] and

Mahankali's application of voxel-wise fMRI models for the evaluation of video quality without reference [12] demonstrate the breadth of innovative approaches being developed. These studies, while contributing valuable information, approach the challenge of quality assessment and TRV utilisation from perspectives different from ours.

Furthermore, our previous work, which offers a comprehensive review of objective methods to assess video quality in recognition tasks, underscores the ongoing evolution and the various approaches within this research domain [13]. This landscape of research illustrates not only the multiplicity of methodologies but also the broad spectrum of applications, from public safety to medical imaging, underscoring the unique positioning and contribution of our work within this field.

Finally, it is worth mentioning Larson and Chandler's work on FR image quality assessment, which offers insights into distortion measurement and its impact on perceived quality [14]. Similarly, Sheikh, Sabir, and Bovik's statistical evaluation provides a foundational understanding of FR image quality assessment algorithms, emphasising the importance of objective measurement in enhancing image processing techniques [15].

While our methodology aims to enhance video quality assessment for object recognition tasks by incorporating advanced machine learning and customised VQIs, it is distinct from existing approaches observed in the literature. For example, Lu et al. have developed an evaluation framework aimed primarily at enhancing video for human viewers, focussing on perceptual quality metrics that may not directly correlate with the performance of object recognition algorithms [16]. On the other hand, the SB-VQA framework uses a stack-based architecture to evaluate video enhancements through a combination of spatial and temporal features extracted through transformers [17]. Although innovative, this approach primarily addresses enhancements from a visual improvement standpoint without a direct focus on the nuanced requirements of object detection systems in varied operational environments.

Our approach diverges significantly by not only focussing on the perceptual aspects of video quality, but also predicting how various quality factors affect the accuracy and reliability of object recognition algorithms in real-world scenarios. This differentiation highlights the unique positioning of our research within the existing landscape and underscores the potential of our methodology to provide more relevant insights for applications requiring high precision in object recognition, such as surveillance and telemedicine.

Due to the concise nature of this 'letter paper', we do not delve more into a detailed analysis of how our work distinguishes itself in superiority compared with others' efforts within the realm of objective video quality assessment for recognition tasks. Instead, we direct the reader to our extensive surveys found in our papers [1,5,13]. These documents meticulously explore state-of-the-art methods for objective video quality assessment specifically tailored to recognition tasks, offering a broad perspective on the advancements and methodologies that underpin our current research. This approach allows us to focus on presenting our novel contributions without reiterating the extensive background covered in our previous work.

In conclusion, while existing methodologies provide some insight into video quality assessment, they do not address the specialised demands of TRVs in high-stakes environments like surveillance and telemedicine. Traditional video quality assessment methods often fail to accurately predict the operational effectiveness of TRVs due to their reliance on metrics designed for general viewing rather than specific recognition tasks. Our research aims to fill this crucial gap by developing an advanced machine learning algorithm that not only assesses but also enhances video quality specifically for the nuanced needs of TRVs. This innovative approach seeks to significantly improve both the accuracy and reliability of TRVs, ensuring better performance in critical applications.

The structure of this paper is as follows: in Section 2, we present the experimental framework; Sections 2.1 and 2.2 detail the collection of the corpus and the creation of degradation models, respectively; the experiments performed are explained in Sections 2.3 and 2.4; our results are shared in Section 3; and the article is summarised in Section 4.

## 2. Materials and Methods

This section presents the detailed methodology employed in our study. As depicted in Figure 1, the flowchart of our general methodology encompasses the essential elements of our research strategy. Our experimental setup is built upon a foundational dataset, known as Source Reference Circuits (SRCs, detailed in Section 2.1), along with various visual impairments, referred to as Hypothetical Reference Circuits (HRCs, detailed in Section 2.2). Each HRC applies a different kind of visual degradation to an SRC. We then analyse the altered video sequences using a computer vision library for object recognition (detailed in Section 2.3), in conjunction with a Video Quality Indicator (VQI, detailed in Section 2.4).

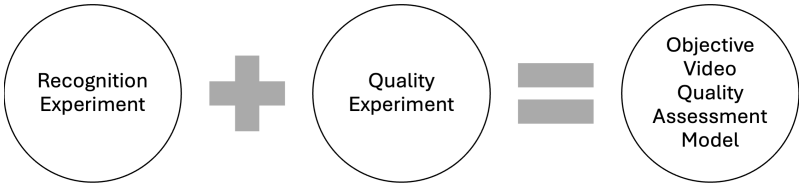


**Figure 1.** Flowchart illustrating our comprehensive methodology, detailing the interactions among the Recognition Experiment, the Quality Experiment, and the Objective Video Quality Assessment Model. This diagram provides an overview of how experimental components contribute to our overall objective of improving video quality assessment for TRVs.

To provide a visual representation of the intricate processes involved in our experimental framework, we included a detailed flowchart. This diagram (Figure 2) elucidates the sequence of steps from the initial video acquisition to the final stages of quality assessment and analysis. By delineating the interactions among the various components of our study, namely, the SRCs, the HRCs, the Recognition Experiment, and the Quality Experiment, it facilitates a deeper understanding of our methodological approach. This visual aid is crucial to understanding how each element contributes to the overarching objective of enhancing video quality assessment for object recognition tasks.

To further clarify the experimental procedures and aid in the reproducibility of our study, in Listing 1, we have provided a pseudocode representation of the workflow used in our experiments. This pseudocode outlines the sequential steps from video data loading to distortion application, recognition processing, and final quality assessment. By following these steps, researchers can replicate our experimental conditions and verify our findings. The pseudocode is detailed below and serves as a guide to navigate the complex interactions among the various components of our study.

**Listing 1.** Pseudocode for Video Quality Assessment Experiment.

```
Pseudocode: Video Quality Assessment Experiment

1. Start
2. Load video data
3. For each video frame:
   a. Apply Hypothetical Reference Circuits (HRCs) to introduce distortions
   b. For each distorted frame:
      i. Process frame through Object Recognition Tool
      ii. Record recognition results
4. Assess the quality of recognition for each distorted frame
   a. Compare the recognition results with ground truth
   b. Calculate quality metrics (e.g., accuracy, mean square error)
5. Aggregate results
6. Analyse the overall performance of the recognition system under various distortions
7. End
```

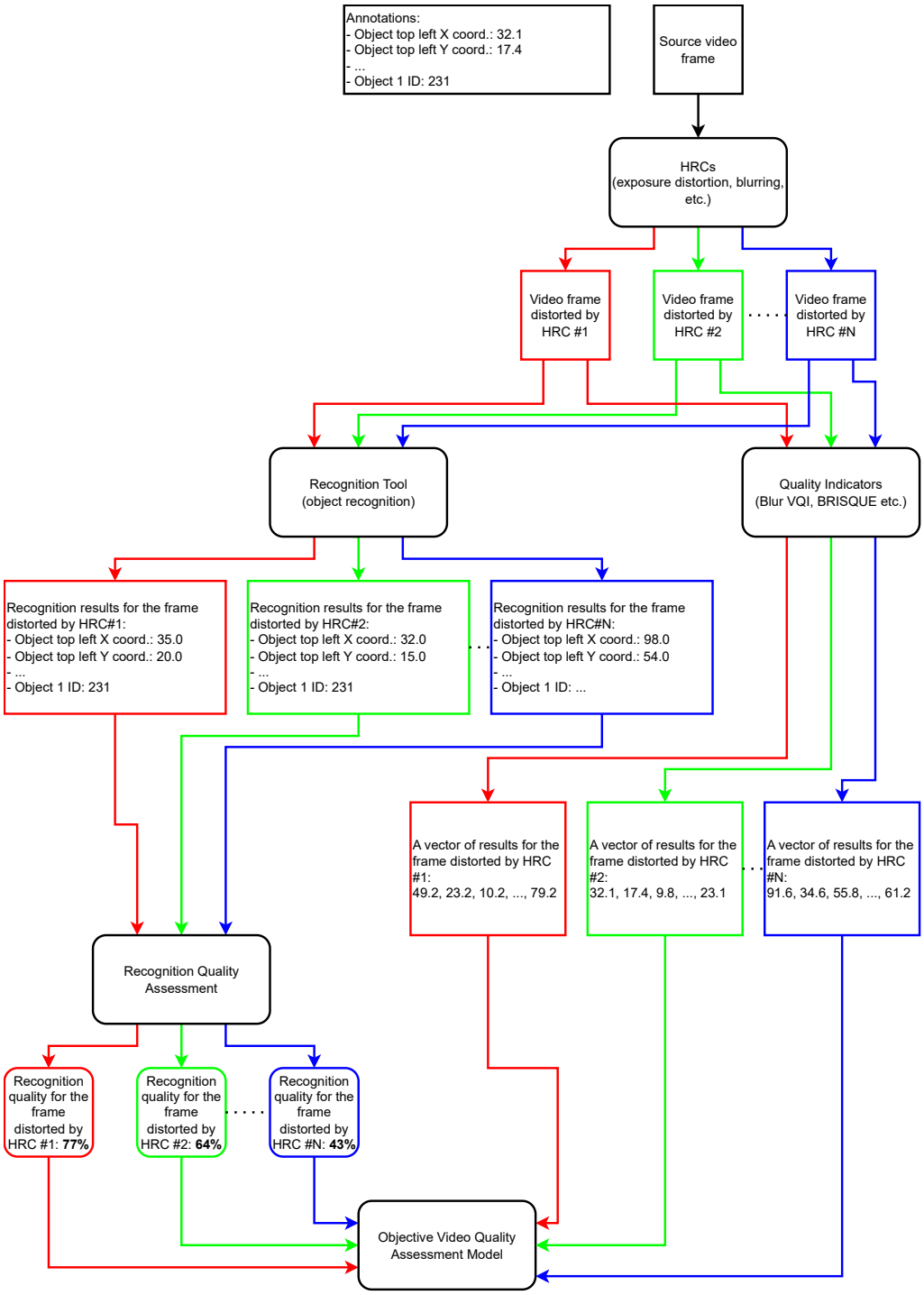


**Figure 2.** Flowchart illustrating the experimental workflow from source video acquisition through recognition tool processing to objective quality assessment. This visualisation helps to understand the sequential processing and quality assessment steps involved in the study.

The process presented in Figure 2 can be described using a series of equations. Starting from an image $S_i$, we use a transformation $H_j(S_i, l)$ that generates a new image $P_{ij}(l)$. $l$ represents the level of specific distortion. In addition, we have the recognition function $R()$ that returns the percentage of areas detected. The recognition is given by the following:

$$r_{ij} = R(P_{ij}(l)) = R(H_j(S_i, l)) \quad (1)$$

Finally, we have objective quality indicators $O_k()$ that indicate the level of a specific distortion $k$. The objective quality for a specific distortion $k$ and image $P_{ij}(l)$ is given by the following:

$$o_{ijk} = O_k(P_{ij}(l)) = O_k(H_j(S_i, l)) \quad (2)$$

Our model $M$ is a function that predicts $r_{ij}$ by $\hat{r}_{ij}$ as a function of $o_{ijk}$ for a specific set of $k$. Finding a sufficiently accurate function $M$, we are able, based on the values $o_{ij1}, \cdots, o_{ijK}$, to predict whether lack of detection is caused by lack of objects or by the low quality of the captured image. Note that the distortion level $l$ is unknown to our function $M$ since, in reality, we do not know how much motion blur or any other distortion was added.

### *2.1. Gathering of Existing Source Reference Circuits (SRCs)*

This subsection discusses the SRC selection and preparation process for the study. SRCs consist of various original video sequences selected to establish a comprehensive database with various characteristics. For the experimental design, a specific subset of the SRC library was chosen, informed by initial tests, the potential for additional training rounds, and a validation experiment for the model developed. This selection ensured that the single experimental cycle, including both the Recognition Experiment (Section 2.3) and the Quality Experiment (Section 2.4), would not exceed 1 week, considering the computation time required to process each frame.

The selection process considered the average processing time for each frame in both experiments to estimate the total number of frames that could be processed within a week. Based on these considerations, it was feasible to process video sequences from 120 unique SRC images within the given time frame. These were allocated as 80 for the initial training phase, 20 for the testing phase, and 20 for validation purposes, each image featuring at least one identifiable object. This method facilitated a practical and efficient experimental setup, striking a balance between comprehensive testing and the limitations of processing time and resource availability. The total count of HRCs, including the original SRC, is 65. Further details on the SRC collection and the specific selections for the experiment are provided in subsequent sections (Sections 2.1.1 and 2.1.2).

#### 2.1.1. The Object Recognition Set

In this study, we have integrated two primary source datasets for Source Reference Sequences (SRCs) ($S_i$): the nuScenes mini-database [18], available at http://www.nuscenes.org/, accessed on 22 April 2024, and a selection from the KITTI dataset [19], which can be found at http://www.cvlibs.net/datasets/kitti/, accessed on 22 April 2024.

The nuScenes dataset, contributed by Aptiv Autonomous Mobility, offers a resolution of 1600 × 900 (HD+). It encompasses v1.0-mini CAM_FRONT sweep images, including 4 scenes from both Boston and Singapore, totalling 1938 frames. Aptiv aims to advance public research on computer vision and autonomous driving by releasing a subset of their comprehensive data. The dataset is a large-scale public resource that includes 1000 driving scenes selected from Boston and Singapore, cities known for their dense traffic and challenging driving conditions. These scenes, each 20 s long, have been chosen to demonstrate a variety of driving manoeuvres, traffic situations, and unexpected behaviours, encouraging the development of safe driving technologies in complex urban settings.

As shown in Figure 3, the dataset includes statistics for the selected frames. The data collection effort spans multiple continents, facilitating the examination of computer vision algorithms' generalisability across different environments.

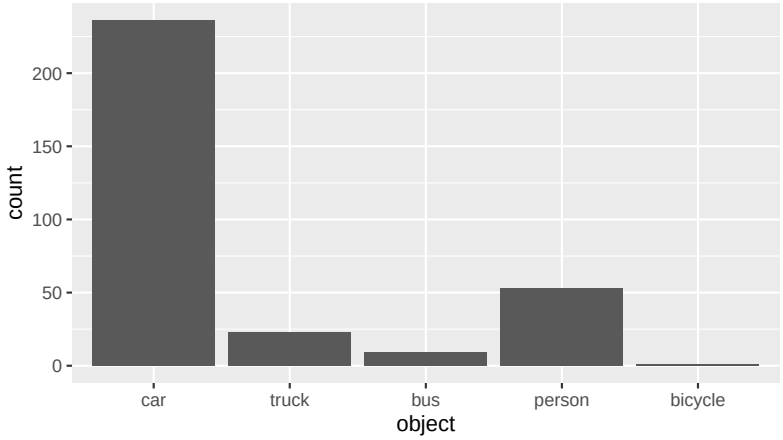


**Figure 3.** Statistical analysis of selected video frames from the nuScenes dataset, showcasing the diversity of scenes captured across urban settings in Boston and Singapore. This chart highlights the object type, critical for evaluating our video quality assessment methodology.

Aptiv annotates 23 object classes with accurate 3D bounding boxes across the dataset, adding object-level attributes such as visibility, activity, and pose. Following this overview, Figure 4 provides a visual example of a frame from the nuScenes dataset, highlighting the rich detail of the dataset and the variety of scenarios it encompasses.

**Figure 4.** Example frame from the nuScenes dataset depicting typical urban traffic conditions used for object detection testing. The frame demonstrates the application of our quality assessment techniques under realistic conditions.

This extensive dataset supports the goal of developing methods that ensure safety and efficiency in urban driving, highlighting the importance of diverse and comprehensive data in the advancement of autonomous driving technologies.

The KITTI dataset, originating from the Karlsruhe Institute of Technology and the Toyota Technological Institute at Chicago, offers a resolution of 1242 $\times$ 375. It is divided into three categories: 'City', 'Residential', and 'Road', totalling 7480 frames. This dataset uses the Annie-WAY autonomous driving platform to create challenging real-world computer vision benchmarks. The benchmarks span tasks like stereo, optical flow, visual odometer, 3D object detection, and 3D tracking, with ground truth provided by a Velodyne laser scanner and GPS localisation system.

In Figure 5, we present statistics for the selected frames from the KITTI dataset. This provides a glimpse into the depth of the dataset and the diversity of the scenarios it covers.

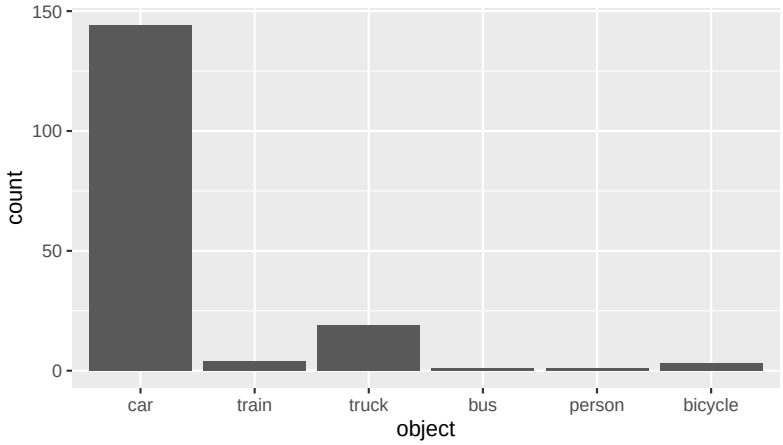


**Figure 5.** Statistical analysis of selected video frames from the KITTI dataset, showcasing the diversity of scenes captured across urban settings in Karlsruhe. This chart highlights the object type, critical for evaluating our video quality assessment methodology.

The KITTI dataset captures data by driving around Karlsruhe, in rural areas, and on highways, featuring scenarios with up to 15 cars and 30 pedestrians per image, thus offering a rich testing ground for computer vision algorithms in various real-world conditions. Figure 6 shows an example frame from the KITTI Vision Benchmark Suite, further illustrating the practical application of the dataset in the testing and improvement of computer vision systems.

**Figure 6.** Example frame from the KITTI Vision Benchmark Suite dataset depicting typical urban traffic conditions used for object detection testing. The frame demonstrates the application of our quality assessment techniques under realistic conditions.

By utilising these source datasets, nuScenes and KITTI, we establish a solid foundation for our object recognition set, providing a broad platform for the evaluation and enhancement of our computer vision models in diverse driving conditions and scenarios.

#### 2.1.2. The Object Recognition Subset

The set of selected SRC frames for object recognition is divided into a training set, a test set, and a validation set, in a ratio of 80 vs. 20 vs. 20, respectively. When selecting images for SRC sets, filtering was applied to ensure that only images with detection covering more than 10% of the video frame are included in the SRC sets.

Figure 7 presents a montage of selected SRC frames. For the full list, please refer to Appendix A.

**Figure 7.** Montage of selected SRC frames used in our object recognition experiments, illustrating the diversity of urban and rural scenes under various lighting and weather conditions. Each frame tests the robustness and adaptability of our video quality assessment and object recognition algorithms, visually representing the dataset's complexity and environmental variability.

### *2.2. Summary of Hypothetical Reference Circuits (HRC)*

Given the concise nature of this 'letter paper', we forgo an in-depth analysis of the development of Hypothetical Reference Circuits (HRCs). Instead, we offer an overview of the HRCs ($H(\cdot, l)$) used in our study, directing the reader to our previous work for comprehensive methodologies and insights. The HRCs outlined for this investigation include a spectrum of impairments encountered in the digital image acquisition process, crucial for video quality assessment in recognition tasks.

- Adjustments in photographic lighting to tackle the challenges of under-/over-exposure;
- Considerations of camera optics, specifically the effects of defocus (blur);
- Issues related to electronic sensors, such as Gaussian noise and motion blur;
- Processing artefacts, with a focus on JPEG compression.

For the application of HRCs, we selected FFmpeg [20] and ImageMagick [21] because of their comprehensive set of filters. FFmpeg was used to introduce under-/over-exposure and Gaussian noise distortions, while ImageMagick was used to add defocus effects, simulate motion blur, and implement JPEG compression.

The computational performance of these tools was evaluated under maximum load (all filters active), achieving a throughput of 439 frames per minute on a standard laptop equipped with an Intel i5 3317U processor and 16 GB of RAM. The equipment was sourced in Kraków, Poland.

The thresholds for various distortions applied through these tools are summarised in the following description. Typically, thresholds are set to identify the point at which recognition fails completely, usually the next-to-last step in our test sequence. For added safety, an additional step is included beyond this point.

- Under-Exposure: Applied via FFmpeg, parameter range from 0 to $-0.6$;
- Over-Exposure: Applied via FFmpeg, parameter range from 0 to 0.6;
- Defocus (Blur): Applied via ImageMagick, parameter range from 0 to 6;
- Gaussian Noise: Applied via FFmpeg, parameter range from 0 to 48;
- Motion Blur: Applied via ImageMagick, parameter range from 0 to 18
- JPEG Compression: Applied via ImageMagick, parameter range from 0 to 100.

For detailed methodologies on these HRCs and their implications for video quality assessment, consult our publications: Refs. [1,5]. These documents elaborate on the criteria for selecting HRCs, the decision to use specific camera models (including digital single-lens reflex cameras and basic pinhole camera models), and how these decisions affect the applicability of quality assessment methods in recognition scenarios.

### *2.3. Recognition Experiment*

Object detection in our study is performed using the YOLOv3 neural network, which is specifically trained on the comprehensive COCO database. This choice is crucial because it ensures robust object detection across a wide range of object classes pertinent to our study, particularly persons, cars, stop signs, traffic lights, trucks, and bicycles. The YOLOv3 model is renowned for its effectiveness in detecting objects with high precision in various lighting and occlusion conditions, making it highly suitable for evaluating video quality in TRVs. For a more in-depth understanding, we refer our readers to Redmon and Farhadi's work on YOLOv3 [22].

The network not only detects, but also accurately marks objects with bounding boxes, providing specificity scores that are critical for subsequent quality assessments. Object detection results are stored in a structured JSON format, enabling streamlined processing and analysis. An example of the KITTI database is provided in Listing 2 to illustrate the structure of the data and the level of detail captured by the detection algorithm.

**Listing 2.** Example of the recognition output file.

```
{ "000001.png":
    {"car":
      [-5, 186, 44, 70, 0.6787664890289307]
    },
 "000000.png":
   {"car":
     [669, 171, 56, 26, 0.8234646916389465]
   },
 "000002.png":
   {"car":
     [558, 186, 50, 17, 0.5980743765830994],
    "motorbike":
     [275, 228, 82, 54, 0.528740406036377]
   }
}
```

Since we know the bounder boxes for the source image and bounded boxes generated by YOLO, we are able to calculate $r_{ij}$ given by Equation (1). The $r_{ij}$ percentage of areas marked by the same object in the source dataset and by YOLO is compared with the percentage of the entire recognised part of the image.

Our approach leverages the OpenCV YOLO Object Detection tutorial (https://www.pyimagesearch.com/2018/11/12/yolo-object-detection-with-opencv/) as the basis for the code that extracts YOLO information, ensuring that our implementation adheres to established and effective practices. The robustness of YOLOv3, combined with its training in the diverse COCO dataset, addresses potential biases related to the distribution of training data. This is critical because it minimises the impact of data distribution shifts that could otherwise affect the performance evaluation in our quality assessment model. The tests indicate that one analyses 500 images in about 100 s. The KITTI database with object annotation contains 7480 images, which means that, for a single HRC, we can calculate the results for all KITTI frames in about 25 min, which is a reasonable time to obtain a significant amount of data points needed for modelling.

In Figures 8–15, the influence of different distortions on the number of objects detected for the selected values is shown.

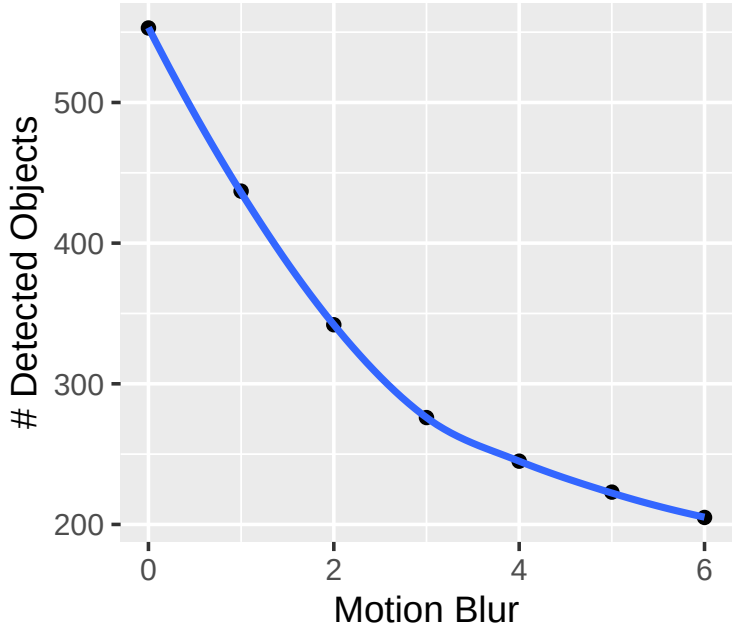


**Figure 8.** Object recognition performance with varying motion blur intensities, quantified in $\sigma$/degrees. This graph displays how motion blur impacts the accuracy of detecting various objects, providing insights into the algorithm's effectiveness in handling such environmental distortions.

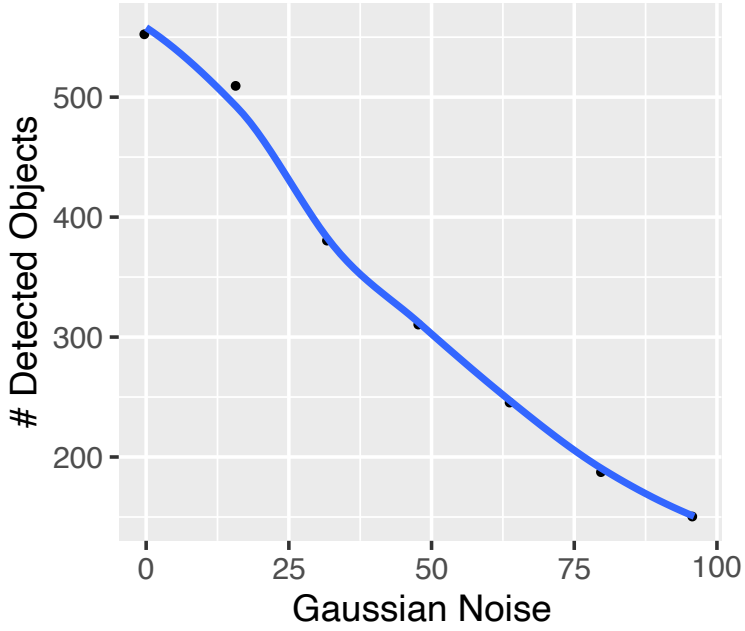


**Figure 9.** Object recognition performance with different levels of Gaussian noise, quantified in $\sigma$/pixels. This graph displays how Gaussian noise impacts the accuracy of detecting various objects, providing insights into the algorithm's effectiveness in handling such environmental distortions.

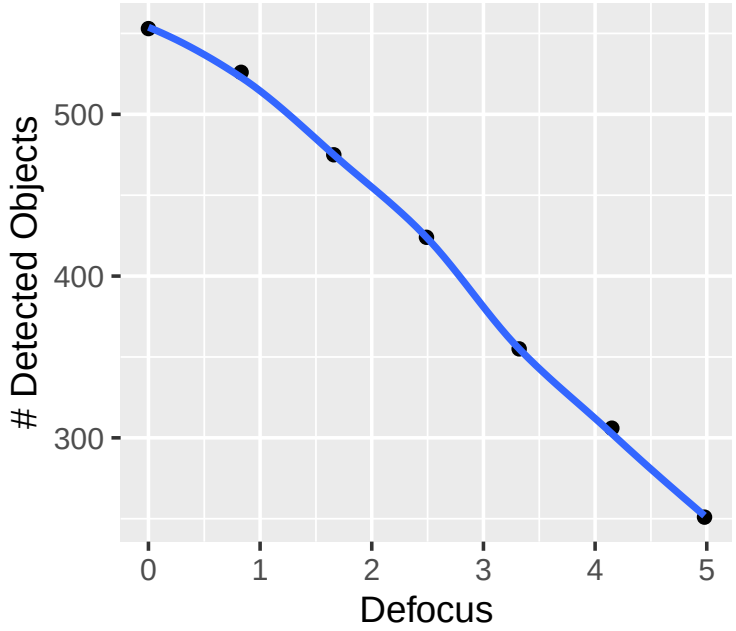


**Figure 10.** Object recognition performance with varying degrees of defocus, quantified in $\sigma$/pixels. This graph displays how defocus impacts the accuracy of detecting various objects, providing insights into the algorithm's effectiveness in handling such environmental distortions.

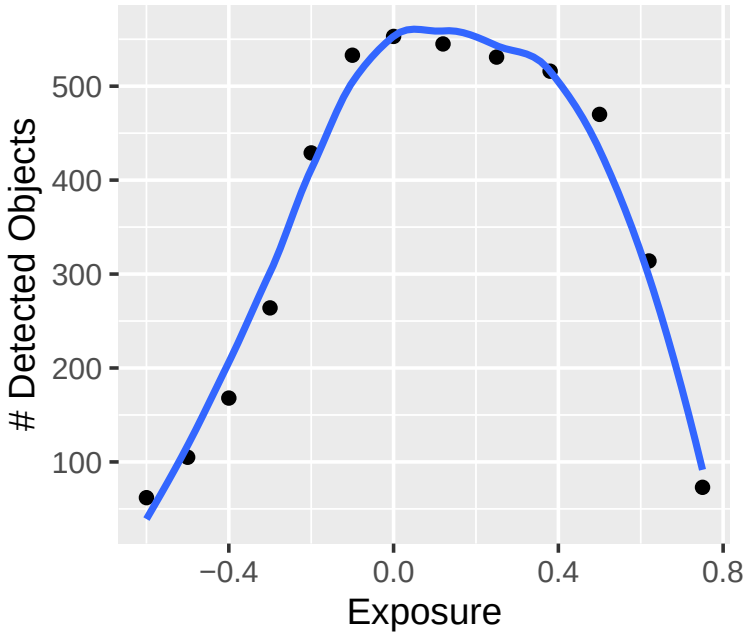


**Figure 11.** Object recognition performance under different exposure levels, measured in equivalent units (eq units). This graph displays how exposure impacts the accuracy of detecting various objects, providing insights into the algorithm's effectiveness in handling such environmental distortions.

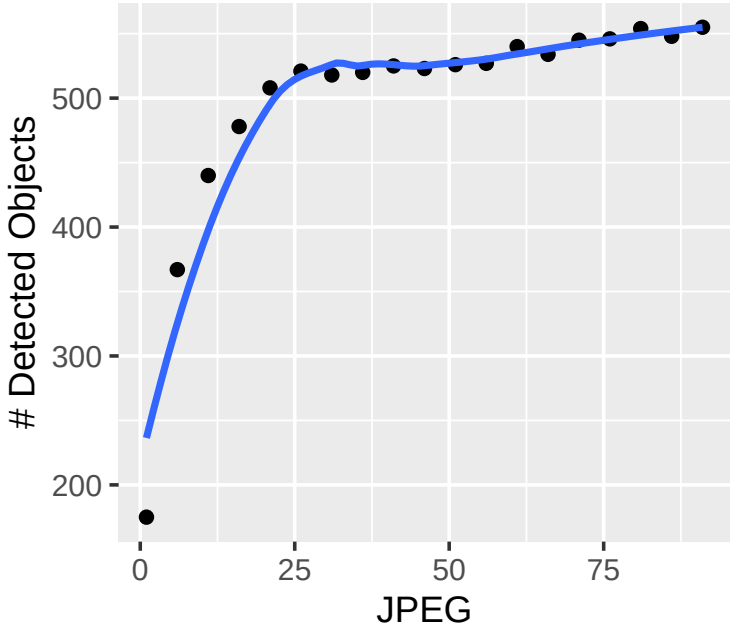


**Figure 12.** Object recognition performance under various levels of JPEG compression, measured in quality units. This graph displays how JPEG compression impacts the accuracy of detecting various objects, providing insights into the algorithm's effectiveness in handling such environmental distortions.

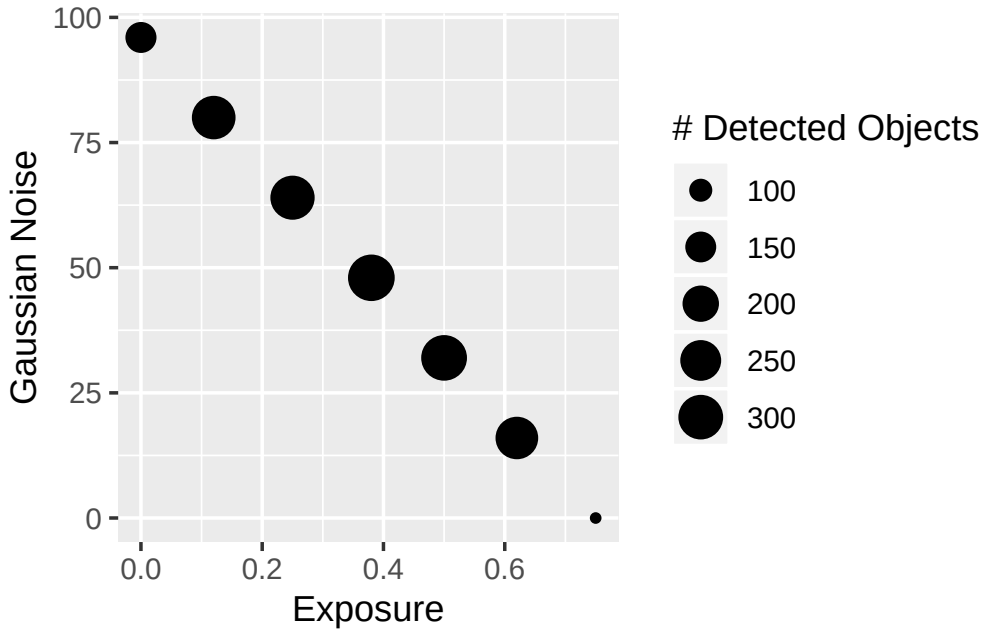


**Figure 13.** Object recognition performance under combined Gaussian noise and exposure, measured in $\sigma$/pixels and equivalent units, respectively. This graph displays how these conditions impact the accuracy of detecting various objects, providing insights into the algorithm's effectiveness in handling such environmental distortions.

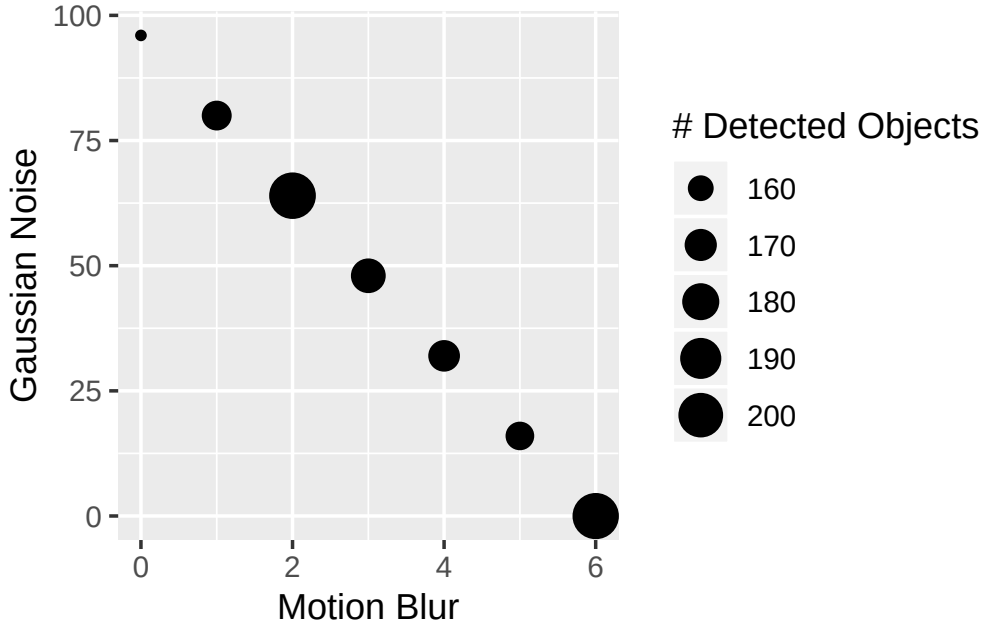


**Figure 14.** Object recognition performance under combined Gaussian noise and motion blur, measured in $\sigma$/pixels and $\sigma$/degrees, respectively. This graph displays how these conditions impact the accuracy of detecting various objects, providing insights into the algorithm's effectiveness in handling such environmental distortions.

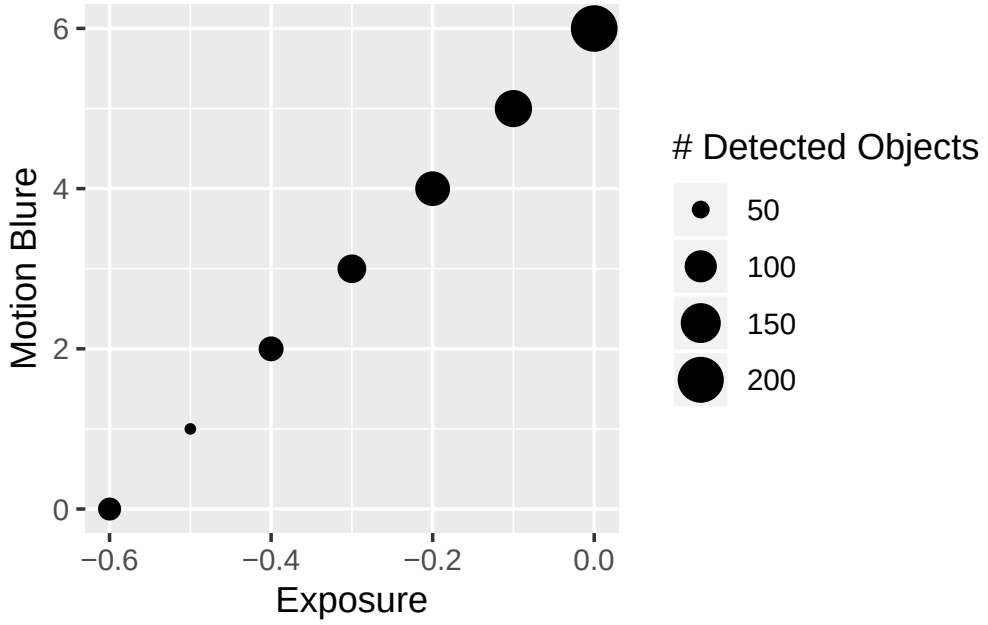


**Figure 15.** Object recognition performance under combined exposure levels and motion blur, measured in equivalent units and $\sigma$/degrees, respectively. This graph displays how these conditions impact the accuracy of detecting various objects, providing insights into the algorithm's effectiveness in handling such environmental distortions.

### *2.4. Quality Experiment*

In alignment with the scope of this 'letter paper', detailed discussions on the Quality Experiment are not included. However, to furnish a comprehensive understanding, we summarise the core aspects and specifically delineate the Video Quality Indicators (VQIs) utilised. For exhaustive methodologies, detailed analyses, and the rationale behind these experiments, we direct our readers to our extensive publications: Refs. [1,5].

The Quality Experiment aims to evaluate the effectiveness and computational efficiency of various VQIs, identifying those best suited for real-time video quality assessment in automated systems, including but not limited to object recognition. This experiment involves the following:

- Breaking down each video into individual frames;
- Applying a comprehensive set of 19 VQIs to each frame;
- Documentation of execution times for each VQI;
- Aggregation of the results into a comprehensive metrics vector.

The Quality Experiment's primary focus is on individual video frames, with an exception for the Temporal Activity (TA) Video Quality Indicator (VQI), which is excluded due to its unique characteristics in assessing temporal aspects of video sequences.

For the sake of clarity, the VQIs are categorised into 'All Metrics' and 'Our (AGH) Metrics'. 'All Metrics' encompasses a broad spectrum of VQIs developed by various research groups worldwide, providing a wide range of perspectives on video quality assessment. Conversely, 'Our (AGH) Metrics' refers specifically to the VQIs developed by the AGH University of Krakow team, focused on particular aspects of video quality that are pertinent to our research interests and projects.

Below is a list of the VQIs employed, categorised by their source:

Our (AGH) Metrics [1]:

- AGH VQI: Blockiness;
- AGH VQI: Blur;
- AGH VQI: Contrast;
- AGH VQI: Exposure;

- AGH VQI: Noise;
- AGH VQI: Spatial Activity;
- AGH VQI: Temporal Activity.

All Metrics:

- LIVE VQI: BIQI [23];
- LIVE VQI: BRISQUE [24];
- LIVE VQI: NIQE [25];
- LIVE VQI: OG-IQA [26];
- LIVE VQI: FFRIQUEE [27];
- LIVE VQI: IL-NIQE [28];
- UMIACS VQI: CORNIA [29];
- BUPT VQI: HOSA [30].

The distinction between 'Our (AGH) Metrics' and 'All Metrics' allows for a nuanced analysis of video quality, catering to both general assessment frameworks and specific scenarios relevant to our research focus. This bifurcation enables the targeted exploration of quality aspects that are crucial for the performance of specialised systems, such as ANPR, under various conditions.

For a thorough understanding of the methodology behind the selection and application of these VQIs, including the computational frameworks and the execution strategy employed, we encourage our readers to refer to the cited papers. These publications provide detailed insights into the experimental design, execution nuances, and analytical perspectives that underpin our approach to advance video quality assessment.

## 3. Results

The modelling using a random forest regressor proved to be the most effective.

Using all metrics, we have obtained MSE: 672.4 and correlation: 0.77 for the test set (not used in the training process or in the validation of initial models). The scatter plot obtained for the predicted and actual values for the test set is shown in Figure 16.

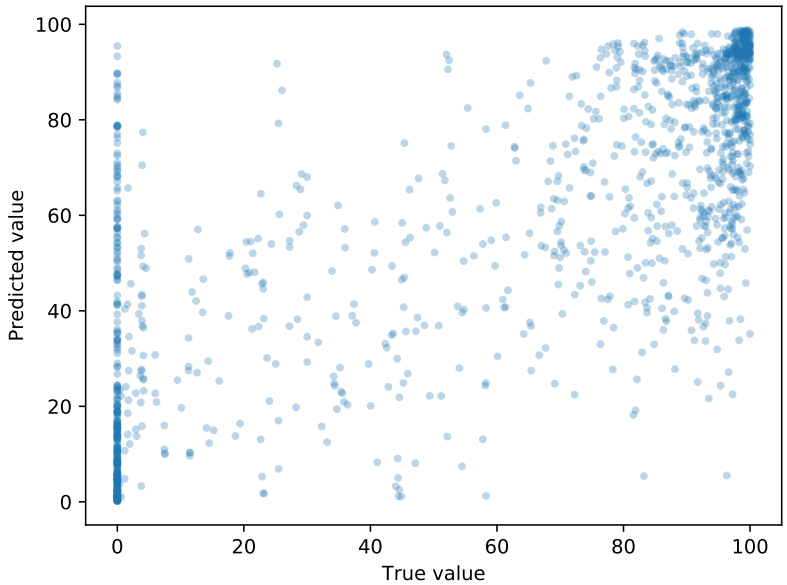


**Figure 16.** Scatter plot showing the correlation between the predicted and actual percentages of detected areas, using all metrics. Each point represents one observation, with semi-transparency used to indicate overlapping data points and to highlight the distribution density.

Using only our metrics, we have obtained MSE: 722.1 and correlation: 0.75. The scatter plot obtained for the predicted and actual values for the test set is shown in Figure 17.

The two models perform similarly, with the slightly worst result for only AGH metrics, which is expected. We can see (by the darker region) that the models work reasonably well

by classifying the most 100 to 100 groups and the most 0 to 0. It is probably the reason why models such as the neural network did not work so well.

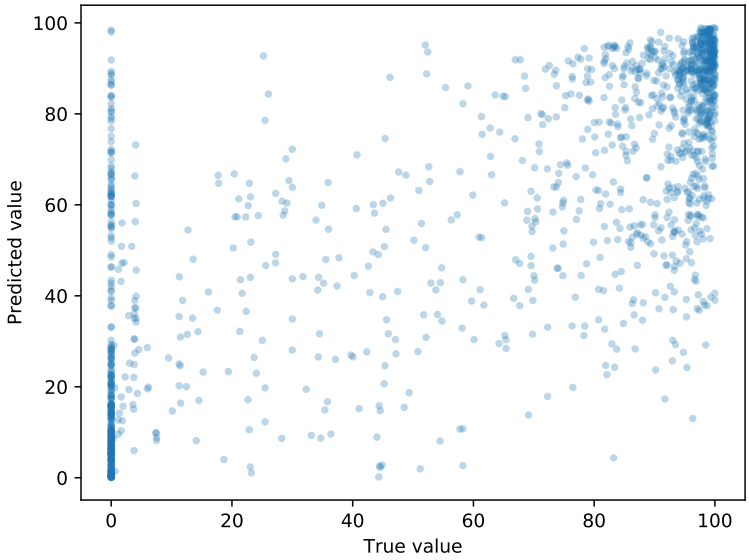


**Figure 17.** Scatter plot showing the correlation between the predicted and actual percentages of detected areas, using only our AGH metrics. Points are semi-transparent to show the overlap and distribution density of the data points effectively.

## 4. Conclusions

Our study represents a significant advancement in the objective assessment of video quality for TRVs, specifically designed to address the intricacies of their operational demands. The implementation of a comprehensive evaluation system, rooted in a meticulously curated object recognition dataset, coupled with a random forest regression model, has led to notable results. We achieved a mean square error of 672.4 and a correlation of 0.77 in all metrics, figures that not only substantiate the predictive competence of the model, but also set a new benchmark for assessing TRV processing efficiency.

These statistical results are of considerable significance. Our findings lay the groundwork for the enhancement of systems that rely on precise object recognition, such as video surveillance and telemedicine, potentially revolutionising their operational effectiveness.

Recognising the limitations inherent in our study, particularly the model's dependence on a specific dataset and image distortion types, we foresee an extensive avenue of exploration in enhancing generalisability. Our future work will focus on enlarging the scope of our dataset and integrating various machine learning algorithms, moving beyond the confines of the random forest regression model. The anticipated expansion not only will refine our predictive framework but also is expected to bolster the model's resilience and its broader applicability to varied real-world scenarios.

The challenge of computational resource constraints and the limited diversity of SRCs tested have been acknowledged as areas that need attention. To this end, our subsequent efforts will be channelled towards amassing a broader array of datasets, engaging with a wider range of recognition systems, and boosting the computational efficacy of our model training and evaluation phases. These steps will contribute to a more comprehensive and thorough validation of our evaluation system.

In addition, we are poised to explore the impact of additional environmental factors on recognition performance. Future iterations of our work will discuss how varying digital camera settings, noise levels, and lighting conditions affect object detection accuracy. Preliminary results have shown that these factors are critical to the adaptability of recognition algorithms, signalling a promising trajectory for research aimed at solidifying the robustness of recognition systems in operationally diverse environments.

In culmination, our ongoing research trajectory includes plans to collaborate with industrial and academic partners, with the aim of testing and refining our methodologies across a spectrum of real-world applications. Such collaborations will not only enhance the practical relevance of our work but also drive innovation in the domain of object recognition for TRVs.

Ultimately, this paper has sought to bridge the gap between the current state of video quality assessment and the emerging needs of advanced recognition systems. We trust that our methodical approach, underscored by concrete statistical evidence and a defined path forward, will invigorate future research and development in this rapidly evolving field.

**Author Contributions:** Introduction, M.L. and A.B.; Acquisition of the Existing SRC, M.L.; Preparation of HRC, M.L. and A.B.; Recognition Experiment, M.L. and J.N.; Quality Experiment, L.J. and J.N.; Results, M.L. and L.J.; Conclusions, M.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received funding from the Huawei Innovation Research Programme (HIRP). This work was supported by the Polish Ministry of Science and Higher Education with the subvention funds of the Faculty of Computer Science, Electronics and Telecommunications of AGH University.

**Data Availability Statement:** The following supporting information can be downloaded at https://qoe.agh.edu.pl/, including Video Indicators and the databases.

**Conflicts of Interest:** Author Atanas Boev was employed by the company Huawei Technologies Dusseldorf GmbH. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Appendix A. The Object Recognition Subset Files

Below, please find the list of selected SRC frames for object recognition:

```
KITTI/City/2011_09_26_drive_0009_sync/image_02/data/0000000149.png
KITTI/City/2011_09_26_drive_0009_sync/image_02/data/0000000297.png
KITTI/City/2011_09_26_drive_0011_sync/image_02/data/0000000208.png
KITTI/City/2011_09_26_drive_0048_sync/image_02/data/0000000005.png
KITTI/City/2011_09_26_drive_0048_sync/image_02/data/0000000011.png
KITTI/City/2011_09_26_drive_0051_sync/image_02/data/0000000255.png
KITTI/City/2011_09_26_drive_0059_sync/image_02/data/0000000187.png
KITTI/City/2011_09_26_drive_0091_sync/image_02/data/0000000303.png
KITTI/City/2011_09_26_drive_0093_sync/image_02/data/0000000144.png
KITTI/City/2011_09_26_drive_0093_sync/image_02/data/0000000289.png
KITTI/Residential/2011_09_26_drive_0019_sync/image_02/data/0000000467.png
KITTI/Residential/2011_09_26_drive_0022_sync/image_02/data/0000000070.png
KITTI/Residential/2011_09_26_drive_0022_sync/image_02/data/0000000131.png
KITTI/Residential/2011_09_26_drive_0022_sync/image_02/data/0000000170.png
KITTI/Residential/2011_09_26_drive_0022_sync/image_02/data/0000000242.png
KITTI/Residential/2011_09_26_drive_0022_sync/image_02/data/0000000401.png
KITTI/Residential/2011_09_26_drive_0022_sync/image_02/data/0000000422.png
KITTI/Residential/2011_09_26_drive_0022_sync/image_02/data/0000000512.png
KITTI/Residential/2011_09_26_drive_0022_sync/image_02/data/0000000533.png
KITTI/Residential/2011_09_26_drive_0023_sync/image_02/data/0000000031.png
KITTI/Residential/2011_09_26_drive_0023_sync/image_02/data/0000000178.png
KITTI/Residential/2011_09_26_drive_0023_sync/image_02/data/0000000205.png
KITTI/Residential/2011_09_26_drive_0023_sync/image_02/data/0000000219.png
KITTI/Residential/2011_09_26_drive_0023_sync/image_02/data/0000000237.png
KITTI/Residential/2011_09_26_drive_0023_sync/image_02/data/0000000248.png
KITTI/Residential/2011_09_26_drive_0023_sync/image_02/data/0000000272.png
KITTI/Residential/2011_09_26_drive_0023_sync/image_02/data/0000000305.png
KITTI/Residential/2011_09_26_drive_0023_sync/image_02/data/0000000371.png
KITTI/Residential/2011_09_26_drive_0035_sync/image_02/data/0000000059.png
KITTI/Residential/2011_09_26_drive_0035_sync/image_02/data/0000000066.png
KITTI/Residential/2011_09_26_drive_0035_sync/image_02/data/0000000115.png
KITTI/Residential/2011_09_26_drive_0036_sync/image_02/data/0000000056.png
KITTI/Residential/2011_09_26_drive_0036_sync/image_02/data/0000000147.png
KITTI/Residential/2011_09_26_drive_0036_sync/image_02/data/0000000150.png
KITTI/Residential/2011_09_26_drive_0036_sync/image_02/data/0000000254.png
KITTI/Residential/2011_09_26_drive_0036_sync/image_02/data/0000000268.png
KITTI/Residential/2011_09_26_drive_0036_sync/image_02/data/0000000790.png
KITTI/Residential/2011_09_26_drive_0039_sync/image_02/data/0000000106.png
KITTI/Residential/2011_09_26_drive_0039_sync/image_02/data/0000000158.png
KITTI/Residential/2011_09_26_drive_0039_sync/image_02/data/0000000175.png
KITTI/Residential/2011_09_26_drive_0039_sync/image_02/data/0000000198.png
```

```
KITTI/Residential/2011_09_26_drive_0039_sync/image_02/data/0000000211.png
KITTI/Residential/2011_09_26_drive_0046_sync/image_02/data/0000000062.png
KITTI/Residential/2011_09_26_drive_0046_sync/image_02/data/0000000063.png
KITTI/Residential/2011_09_26_drive_0061_sync/image_02/data/0000000123.png
KITTI/Residential/2011_09_26_drive_0061_sync/image_02/data/0000000405.png
KITTI/Residential/2011_09_26_drive_0064_sync/image_02/data/0000000041.png
KITTI/Residential/2011_09_26_drive_0064_sync/image_02/data/0000000053.png
KITTI/Residential/2011_09_26_drive_0064_sync/image_02/data/0000000108.png
KITTI/Residential/2011_09_26_drive_0064_sync/image_02/data/0000000190.png
KITTI/Residential/2011_09_26_drive_0064_sync/image_02/data/0000000191.png
KITTI/Residential/2011_09_26_drive_0064_sync/image_02/data/0000000344.png
KITTI/Residential/2011_09_26_drive_0086_sync/image_02/data/0000000566.png
KITTI/Residential/2011_09_26_drive_0087_sync/image_02/data/0000000039.png
KITTI/Road/2011_09_26_drive_0028_sync/image_02/data/0000000075.png
KITTI/Road/2011_09_26_drive_0029_sync/image_02/data/0000000216.png
KITTI/Road/2011_09_26_drive_0052_sync/image_02/data/0000000022.png
KITTI/Road/2011_09_26_drive_0052_sync/image_02/data/0000000026.png
KITTI/Road/2011_09_26_drive_0052_sync/image_02/data/0000000052.png
KITTI/Road/2011_09_26_drive_0052_sync/image_02/data/0000000073.png
nuScenes/samples/CAM_FRONT/n008-2018-08-01-15-16-36-0400__CAM_FRONT__1533151606012404.jpg
nuScenes/samples/CAM_FRONT/n008-2018-08-01-15-16-36-0400__CAM_FRONT__1533151610412404.jpg
nuScenes/samples/CAM_FRONT/n008-2018-08-01-15-16-36-0400__CAM_FRONT__1533151611412404.jpg
nuScenes/samples/CAM_FRONT/n008-2018-08-01-15-16-36-0400__CAM_FRONT__1533151611862404.jpg
nuScenes/samples/CAM_FRONT/n008-2018-08-01-15-16-36-0400__CAM_FRONT__1533151612362404.jpg
nuScenes/samples/CAM_FRONT/n008-2018-08-01-15-16-36-0400__CAM_FRONT__1533151621012404.jpg
nuScenes/samples/CAM_FRONT/n008-2018-08-01-15-16-36-0400__CAM_FRONT__1533151621912404.jpg
nuScenes/samples/CAM_FRONT/n008-2018-08-01-15-16-36-0400__CAM_FRONT__1533151622412404.jpg
nuScenes/samples/CAM_FRONT/n008-2018-08-27-11-48-51-0400__CAM_FRONT__1535385093162404.jpg
nuScenes/samples/CAM_FRONT/n008-2018-08-27-11-48-51-0400__CAM_FRONT__1535385096862404.jpg
nuScenes/samples/CAM_FRONT/n008-2018-08-27-11-48-51-0400__CAM_FRONT__1535385097362404.jpg
nuScenes/samples/CAM_FRONT/n008-2018-08-27-11-48-51-0400__CAM_FRONT__1535385107412404.jpg
nuScenes/samples/CAM_FRONT/n008-2018-08-28-16-43-51-0400__CAM_FRONT__1535489301512404.jpg
nuScenes/samples/CAM_FRONT/n008-2018-08-28-16-43-51-0400__CAM_FRONT__1535489302512404.jpg
nuScenes/samples/CAM_FRONT/n008-2018-08-28-16-43-51-0400__CAM_FRONT__1535489303912404.jpg
nuScenes/samples/CAM_FRONT/n008-2018-08-28-16-43-51-0400__CAM_FRONT__1535489305912404.jpg
nuScenes/samples/CAM_FRONT/n008-2018-08-28-16-43-51-0400__CAM_FRONT__1535489307412404.jpg
nuScenes/samples/CAM_FRONT/n008-2018-08-28-16-43-51-0400__CAM_FRONT__1535489308362404.jpg
nuScenes/samples/CAM_FRONT/n008-2018-08-28-16-43-51-0400__CAM_FRONT__1535489308862404.jpg
nuScenes/samples/CAM_FRONT/n008-2018-08-28-16-43-51-0400__CAM_FRONT__1535489311862404.jpg
nuScenes/samples/CAM_FRONT/n008-2018-08-30-15-16-55-0400__CAM_FRONT__1535657118612404.jpg
nuScenes/samples/CAM_FRONT/n008-2018-08-30-15-16-55-0400__CAM_FRONT__1535657119612404.jpg
nuScenes/samples/CAM_FRONT/n008-2018-08-30-15-16-55-0400__CAM_FRONT__1535657120112404.jpg
nuScenes/samples/CAM_FRONT/n015-2018-07-24-11-22-45+0800__CAM_FRONT__1532402928112460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-07-24-11-22-45+0800__CAM_FRONT__1532402929162460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-07-24-11-22-45+0800__CAM_FRONT__1532402935662460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-07-24-11-22-45+0800__CAM_FRONT__1532402937162460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-07-24-11-22-45+0800__CAM_FRONT__1532402937662460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-07-24-11-22-45+0800__CAM_FRONT__1532402938162460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-07-24-11-22-45+0800__CAM_FRONT__1532402938612460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-07-24-11-22-45+0800__CAM_FRONT__1532402939112460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-07-24-11-22-45+0800__CAM_FRONT__1532402943162460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-07-24-11-22-45+0800__CAM_FRONT__1532402944662460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-07-24-11-22-45+0800__CAM_FRONT__1532402945162460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-07-24-11-22-45+0800__CAM_FRONT__1532402945662460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-07-24-11-22-45+0800__CAM_FRONT__1532402946262460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-10-02-10-50-40+0800__CAM_FRONT__1538448763512460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-10-08-15-36-50+0800__CAM_FRONT__1538984240912467.jpg
nuScenes/samples/CAM_FRONT/n015-2018-10-08-15-36-50+0800__CAM_FRONT__1538984242412460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-10-08-15-36-50+0800__CAM_FRONT__1538984242912460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-10-08-15-36-50+0800__CAM_FRONT__1538984243412460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-10-08-15-36-50+0800__CAM_FRONT__1538984243912460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-10-08-15-36-50+0800__CAM_FRONT__1538984244412460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-10-08-15-36-50+0800__CAM_FRONT__1538984244912460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-10-08-15-36-50+0800__CAM_FRONT__1538984246912460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-11-21-19-38-26+0800__CAM_FRONT__1542800847912460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-11-21-19-38-26+0800__CAM_FRONT__1542800848912460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-11-21-19-38-26+0800__CAM_FRONT__1542800849412460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-11-21-19-38-26+0800__CAM_FRONT__1542800849912460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-11-21-19-38-26+0800__CAM_FRONT__1542800850412460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-11-21-19-38-26+0800__CAM_FRONT__1542800851412460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-11-21-19-38-26+0800__CAM_FRONT__1542800851912460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-11-21-19-38-26+0800__CAM_FRONT__1542800852412460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-11-21-19-38-26+0800__CAM_FRONT__1542800852912460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-11-21-19-38-26+0800__CAM_FRONT__1542800853412460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-11-21-19-38-26+0800__CAM_FRONT__1542800853912460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-11-21-19-38-26+0800__CAM_FRONT__1542800854912460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-11-21-19-38-26+0800__CAM_FRONT__1542800855412460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-11-21-19-38-26+0800__CAM_FRONT__1542800989412460.jpg
nuScenes/samples/CAM_FRONT/n015-2018-11-21-19-38-26+0800__CAM_FRONT__1542800991912460.jpg
```

## References


1. Leszczuk, M.; Janowski, L.; Nawała, J.; Zhu, J.; Wang, Y.; Boev, A. Objective Video Quality Assessment and Ground Truth Coordinates for Automatic License Plate Recognition. *Electronics* **2023**, *12*, 4721. [CrossRef]
2. ITU-T Study Group 12. *LS about New Work Item P.Obj-Recognition: Object-Recognition-Rate-Estimation Model in Surveillance Video of Autonomous Driving, 2023*; Ref.: SG12-TD311; ITU-T Study Group 12: Geneva, Switzerland, 2023,
3. NTT. *Draft Terms of Reference (ToR) P.obj-recog* ; Contribution SG12-Cn; International Telecommunication Union: San Mateo, CA, USA, 2023.
4. NTT. *Draft Test Plan of P.obj-recog*; Contribution SG12-Cn; International Telecommunication Union: San Mateo, CA, USA, 2023.
5. Leszczuk, M.; Janowski, L.; Nawała, J.; Boev, A. Objective video quality assessment method for face recognition tasks. *Electronics* **2022**, *11*, 1167. [CrossRef]
6. Shi, H.; Liu, C. An Innovative Video Quality Assessment Method and An Impairment Video Dataset. In Proceedings of the 2021 IEEE International Conference on Imaging Systems and Techniques (IST), Kaohsiung, Taiwan, 24–26 August 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–6.
7. Xing, W.; Lyu, T.; Chu, X.; Rong, Y.; Lee, C.G.; Sun, Q.; Zou, Y. Recognition and classification of single melt tracks using deep neural network: A fast and effective method to determine process windows in selective laser melting. *J. Manuf. Process.* **2021**, *68*, 1746–1757. [CrossRef]
8. Khan, Z.A.; Beghdadi, A.; Cheikh, F.A.; Kaaniche, M.; Pelanis, E.; Palomar, R.; Fretland, Å.A.; Edwin, B.; Elle, O.J. Towards a video quality assessment based framework for enhancement of laparoscopic videos. In Proceedings of the Medical Imaging 2020: Image Perception, Observer Performance, and Technology Assessment, Houston, TX, USA, 15–20 February 2020; International Society for Optics and Photonics: Bellingham, WA, USA, 2020; Volume 11316, p. 113160P.
9. Hofbauer, H.; Autrusseau, F.; Uhl, A. To recognize or not to recognize—A database of encrypted images with subjective recognition ground truth. *Inf. Sci.* **2021**, *551*, 128–145. [CrossRef]
10. Wu, J.; Ma, J.; Liang, F.; Dong, W.; Shi, G.; Lin, W. End-to-end blind image quality prediction with cascaded deep neural network. *IEEE Trans. Image Process.* **2020**, *29*, 7414–7426. [CrossRef]
11. Oszust, M. Local feature descriptor and derivative filters for blind image quality assessment. *IEEE Signal Process. Lett.* **2019**, *26*, 322–326. [CrossRef]
12. Mahankali, N.S.; Raghavan, M.; Channappayya, S.S. No-Reference Video Quality Assessment Using Voxel-wise fMRI Models of the Visual Cortex. *IEEE Signal Process. Lett.* **2021**. [CrossRef]
13. Kawa, K.; Leszczuk, M.; Boev, A. Survey on the state-of-the-art methods for objective video quality assessment in recognition tasks. In Proceedings of the Multimedia Communications, Services and Security: 10th International Conference, MCSS 2020, Kraków, Poland, 8–9 October 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 332–350.
14. C. Larson, E.; Chandler, D. Most apparent distortion: Full-reference image quality assessment and the role of strategy. *J. Electron. Imaging* **2010**, *19*, 011006. [CrossRef]
15. Sheikh, H.R.; Sabir, M.F.; Bovik, A.C. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans. Image Process.* **2006**, *15*, 3440–3451. [CrossRef]
16. Lu, J.; Zou, B.; Cheng, Z.; Pu, S.; Zhou, S.; Niu, Y.; Wu, F. Object-qa: Towards high reliable object quality assessment. *arXiv* **2020**, arXiv:2005.13116.
17. Huang, D.J.; Kao, Y.T.; Chuang, T.H.; Tsai, Y.C.; Lou, J.K.; Guan, S.H. Sb-vqa: A stack-based video quality assessment framework for video enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 1613–1622.
18. Caesar, H.; Bankiti, V.; Lang, A.H.; Vora, S.; Liong, V.E.; Xu, Q.; Krishnan, A.; Pan, Y.; Baldan, G.; Beijbom, O. nuScenes: A multimodal dataset for autonomous driving. *arXiv* **2019**, arXiv:1903.11027.
19. Geiger, A.; Lenz, P.; Stiller, C.; Urtasun, R. Vision meets Robotics: The KITTI Dataset. *Int. J. Robot. Res. (IJRR)* **2013**, *32*, 0278364913491297. [CrossRef]
20. FFmpeg. FFmpeg. 2019. Available online: https://ffmpeg.org/ (accessed on 4 June 2019).
21. ImageMagick Studio LLC. *ImageMagick: Convert, Edit, Or Compose Bitmap Images*; ImageMagick Studio LLC: Lost Springs, WY, USA, 2011.
22. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
23. Xue, W.; Mou, X.; Zhang, L.; Bovik, A.C.; Feng, X. Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features. *IEEE Trans. Image Process.* **2014**, *23*, 4850–4862. [CrossRef] [PubMed]
24. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.* **2012**, *21*, 4695–4708. [CrossRef] [PubMed]
25. Yang, C.; He, Q.; An, P. Unsupervised blind image quality assessment via joint spatial and transform features. *Sci. Rep.* **2023**, *13*, 10865. [CrossRef] [PubMed]
26. Liu, L.; Hua, Y.; Zhao, Q.; Huang, H.; Bovik, A.C. Blind image quality assessment by relative gradient statistics and adaboosting neural network. *Signal Process. Image Commun.* **2016**, *40*, 1–15. [CrossRef]
27. Ghadiyaram, D.; Bovik, A.C. Perceptual quality prediction on authentically distorted images using a bag of features approach. *J. Vis.* **2017**, *17*, 32–32. [CrossRef] [PubMed]

28. Zhang, L.; Zhang, L.; Bovik, A.C. A feature-enriched completely blind image quality evaluator. *IEEE Trans. Image Process.* **2015**, *24*, 2579–2591. [CrossRef] [PubMed]
29. Ye, P.; Kumar, J.; Kang, L.; Doermann, D. Unsupervised feature learning framework for no-reference image quality assessment. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 1098–1105.
30. Xu, J.; Ye, P.; Li, Q.; Du, H.; Liu, Y.; Doermann, D. Blind image quality assessment based on high order statistics aggregation. *IEEE Trans. Image Process.* **2016**, *25*, 4444–4457. [CrossRef] [PubMed]