HIQA-DB: A Benchmark Dataset for Image Quality Assessment in Hospital Surveillance

Yujin Han and Taewan Kim
Data Science Major, Dongduk Women's University, South Korea
E-mail: hanyujinius@gmail.com, kimtwan21@dongduk.ac.kr

***Abstract*—Despite the growing use of intelligent video analytics in healthcare environments, the impact of image quality on algorithmic performance remains underexplored. In particular, hospital surveillance footage often exhibits highly uniform backgrounds (white walls, beds, and patient clothing) which complicates person detection, pose estimation, fall detection, and self-harm monitoring under degraded visual conditions. To address this, we present a new Hospital Image Quality Assessment Database (HIQA-DB), which contains 150 original images captured from real hospital Closed Circuit Television (CCTV) footage. Each reference image has four types of synthetically distorted versions, resulting in a total of 750 images. Subjective quality ratings collected from human evaluators reveal that compression artifacts degrade perceived quality the most and correlate strongly with performance drops in downstream analytics models. HIQA-DB provides a valuable benchmark for understanding and mitigating the effects of image quality on intelligent video analytics in clinical settings, and will be publicly released soon to support further research.**

I. Introduction

Ensuring patient safety and promptly detecting critical behaviors are paramount in sensitive environments such as psychiatric hospitals. High-risk behaviors including self-harm, escape-attempts, and fall-down can lead to severe consequences if not recognized in time, significantly increasing the burden on medical staff [1]. To mitigate these risks, hospitals have widely deployed CCTV-based monitoring systems. However, automatic detection and assessment of abnormal behaviors remain challenging in practice. Recent advances in deep learning have led to notable progress in human detection, pose estimation, and abnormal behavior recognition from video streams [2]. Despite this progress, real-world psychiatric hospital CCTV footage presents unique challenges because inconsistent video quality caused by illumination changes, unconventional camera angles, and inherent image degradations frequently degrade the performance of these algorithms. Models trained on public datasets or controlled experimental settings often fail to generalize to hospital surveillance environments. Moreover, there is a clear lack of datasets that adequately capture the specific characteristics of psychiatric hospital scenes while accounting for diverse image quality conditions. This absence hampers the objective evaluation and improvement of detection, pose estimation, and abnormal behavior detection models in realistic settings. In addition, the highly uniform backgrounds typical of psychiatric wards—such as white walls, beds, and patient clothing—further reduce the discriminative visual cues necessary for robust detection and recognition.

TABLE I: Performance of human detection, pose estimation, and behavior detection across different QA levels.

QA Level	Human Detection (YOLO v9)	Pose Estimation (HRNet)	Self-harm	Escape-attempt	Fall-down
0.5	0.6143	0.5562	0.4821	0.4987	0.4492
0.4	0.5237	0.4685	0.3714	0.3849	0.3328
0.3	0.4178	0.3571	0.2639	0.2742	0.2275
0.2	0.3924	0.3345	0.2531	0.2642	0.2125

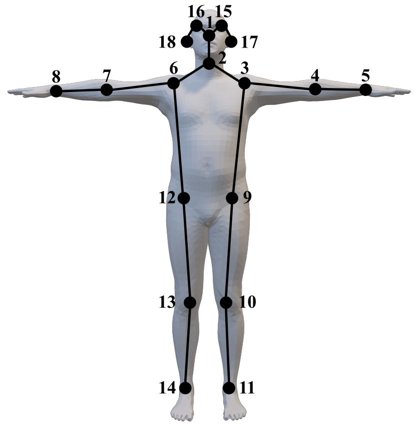
To address these limitations, we quantitatively analyze how the performance of three key computer vision tasks—detection, pose estimation, and abnormal behavior recognition—varies under different image degradation types. Additionally, we introduce the HIQA-DB, a novel dataset constructed from real psychiatric hospital CCTV footage. HIQA-DB comprises 750 images, including 150 originals and 600 synthetically distorted images with four representative degradations: Contrast Change (CC), Gaussian Blur (GB), motion Blur (MB), and JPEG Compression (JC). This controlled setup enables systematic evaluation of how image quality impacts algorithmic performance. Furthermore, we conduct subjective quality evaluation by collecting Mean Opinion Scores (MOS) for each degradation level. The rest of this paper is organized as follows. Section 2 presents quantitative results and discussions on the impact of image quality on detection, pose estimation, and abnormal behavior detection, supported by illustrative tables. Section 3 details the composition of HIQA-DB and reports the MOS analysis. Finally, Section 4 concludes the paper and outlines directions for future work.

II. Motivation

In psychiatric hospital CCTV environments, image quality degradation can significantly undermine the reliability and stability of AI-based monitoring systems. However, the degree to which such degradations lead to measurable performance loss remains underexplored, and systematic analyses are scarce. This section examines how image quality affects recognition and analysis performance, highlighting limitations of existing approaches and motivating the construction of HIQA-DB.

A. Performance Degradation Analysis

We quantitatively analyzed how human detection, pose estimation, and abnormal behavior detection performance varies



(a) Definition of the skeleton joint numbering used for pose estimation.

	Self-harm		Escape-attempt	Fall-down	
	Hit Head	**Injure Arms/Legs**	**Climb Window**	**Up→Down**	**Sideways**
Position Algorithm	$W_c \leq W_p$ $20 \leq \theta_{SEWL} \leq 90$ or $20 \leq \theta_{SEWR} \leq 90$ $\theta_{SEWL_c} \leq \theta_{SEWL_p}$ or $\theta_{SEWR_c} \leq \theta_{SEWR_p}$	$W_c \geq W_p$ $\theta_{SEWL} \leq 180$ or $\theta_{SEWR} \leq 180$	$\theta_{HSEL} \geq 70$ or $\theta_{HSER} \geq 70$ $\theta_{HNAL} \leq 30$ or $\theta_{HNAR} \leq 30$	$APy_p + \alpha \leq APy_c$	$APx_c + \beta \leq APx_p$ or $APx_p + \beta \leq APx_c$

(b) Rules of the position algorithm applied to define self-harm, escape-attempt, and fall-down behaviors. For self-harm, W_c and W_p indicate the wrist y-coordinates in the current and previous frames, respectively. For fall-down detection, APy and APx represent the average y and x positions of all skeleton joints. Angles are denoted as θ followed by the abbreviations of connected joints (e.g., θ_{HNAR}), and constants α and β are set to 10% of the image height and width.

Fig. 1: Illustration of (a) the skeleton numbering system and (b) the position algorithm criteria used for behavior classification.

across different quality levels. For human detection, we employed YOLO v9 [3]; for pose estimation, we used High-Resolution Network (HRNet) [4]. Quality levels were defined by segmenting BRISQUE [5] scores into four intervals. As shown in Table I, all tasks exhibit consistent performance degradation as quality decreases. At the highest quality level (0.5), detection performance remains relatively stable, but at the lowest level (0.2), accuracy for self-harm, escape-attempts, and fall-down drops below 0.3. Similarly, human detection and pose estimation also degrade steeply at low quality levels. These results confirm that even state-of-the-art models fail to maintain reliable performance under real-world hospital CCTV conditions. Figure 1 illustrates the skeleton numbering and behavior classification criteria [6] used for evaluation.

B. Limitations of Existing Databases

Existing datasets have been widely used to evaluate algorithm performance under degraded image conditions [7]. However, most are constructed from natural images or controlled settings, lacking the distinct characteristics of psychiatric hospital environments. Also, Many datasets focus on high-quality imagery with fixed camera positions and stable lighting, making them unsuitable for assessing combined degradations common in real-world CCTV footage [8]. Furthermore, they inadequately represent the sharp performance decline in detection, pose estimation, and abnormal behavior detection observed in low-quality hospital images. These limitations hinder both objective evaluation and practical deployment of robust monitoring systems in healthcare settings. To address these challenges, we propose HIQA-DB, a dataset reflecting diverse real-world degradations in psychiatric hospital CCTV footage, enabling systematic analysis and benchmarking.

III. Hospital Image Quality Dataset

This section introduces the overall composition and characteristics of the newly constructed HIQA-DB. The database is based on videos captured in real psychiatric hospital CCTV environments and was designed to evaluate the impact of image quality on the performance of computer vision tasks. Subsequently, the composition of the data, the representative degradation conditions applied, and the results of subjective quality assessment are presented step by step.

A. Development and Composition of HIQA-DB

In this study, we propose HIQA-DB, which is designed to systematically reflect various types of image degradation that can occur in psychiatric hospital CCTV environments and to enable a more realistic evaluation of detection and recognition models applicable to real-world clinical settings. Most publicly available datasets are composed of images collected from natural scenes or controlled experimental environments and therefore have limitations in capturing frequent quality degradation factors such as lighting variations, diverse camera viewpoints, and complex indoor structures observed in actual hospital surveillance footage.

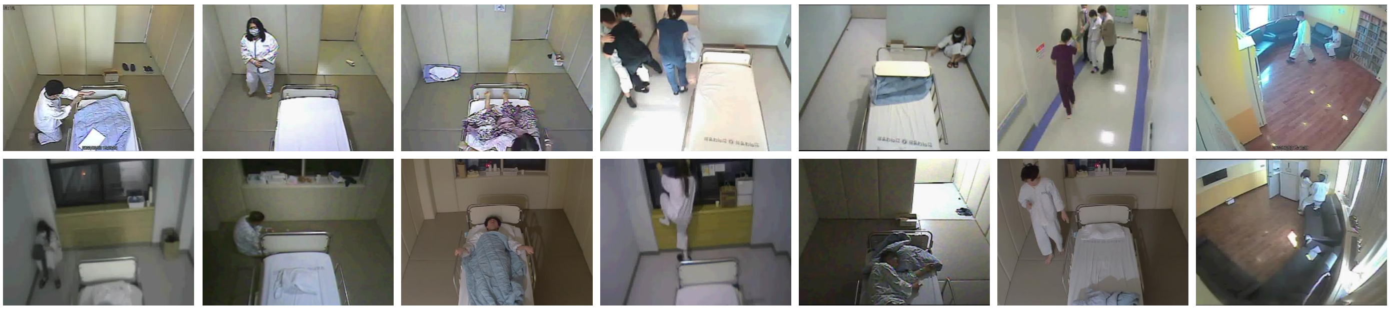


Fig. 2: Representative examples from the proposed HIQA-DB dataset.

TABLE II: Comparison of MOS scores and performance in human detection, pose estimation, and behavior detection under different degradation conditions.

Degradation	MOS	Human Detection	Pose Estimation	Behavior Detection
Contrast change (CC)	3.1	0.81	0.78	0.75
Gaussian blur (GB)	2.6	0.65	0.60	0.58
Motion blur (MB)	2.4	0.59	0.55	0.50
JPEG compression (JC)	2.2	0.55	0.52	0.48

To address these constraints and comprehensively assess the robustness of algorithms, all data were collected from operational psychiatric hospital CCTV systems.HIQA-DB consists of 150 images, all sourced from CCTV recordings captured within psychiatric wards. Each image was augmented with four representative types of controlled distortion to facilitate quantitative analysis of algorithmic performance under varying quality conditions. Figure 2 presents example images included in the database, illustrating a range of viewpoints, lighting conditions, and scene complexities. This composition is intended to mitigate bias toward specific scenarios and to provide a more precise benchmark for evaluating model performance in realistic surveillance environments.

Additionally, this study obtained official approval for data use from the hospital administration. Before collection and utilization, all patients and staff involved were fully informed about the research objectives and data usage plans, and written consent was obtained through a formal consent process. All necessary administrative procedures were implemented to ensure ethical and legal compliance, thereby securing the reliability and appropriateness of the dataset for research purposes.

B. Impact of Typical Distortions in Hospital CCTV

In this study, four representative types of degradation frequently observed in real CCTV environments were consistently applied to all collected images to simulate realistic conditions. The selected degradations of CC, GB, MB, and JC reflect common forms of image quality deterioration occurring during the installation and operation of CCTV systems.

First, CC simulates situations where contrast is significantly reduced due to lighting variations or instability in automatic camera exposure. These conditions can degrade visibility even when the subject remains largely stationary and often reduce the reliability of object detection. Overall, the impact of CC was relatively moderate, but it still introduced ambiguity in recognizing fine-grained actions. GB was applied to model blurring caused by lens contamination, defocus, or the resolution limitations of the system. This degradation disperses silhouettes and blurs contours, which significantly undermines the accuracy of pose estimation and keypoint recognition. In particular, hospital CCTV environments are more vulnerable to GB because of variations in camera positioning and distances to subjects. This vulnerability is further amplified in psychiatric single rooms, where cameras are mounted high to ensure unobstructed monitoring of the entire patient area, resulting in greater distances and smaller subject size in the frame. MB occurs when subjects move rapidly or the camera itself shakes, strongly affecting the detection of sudden patient behaviors or escape-attempts. In videos with MB, the edges of objects appear elongated, and motion direction and shape become distorted. MOS evaluation showed that these videos were perceived as having low visual quality, suggesting that detection models may experience severe performance degradation in actual emergency scenarios. Lastly, JC was introduced to simulate compression artifacts arising during data storage and transmission. In hospital CCTV systems, high levels of compression are often employed to reduce storage requirements and network bandwidth, which frequently results in noise and the loss of fine details. In this study, JC also had a pronounced negative impact on perceived visual quality and led to a substantial drop in recognition accuracy.

As summarized in Table II, JC and MB caused the most significant declines in both subjective quality and model performance across all recognition tasks. In this study, human detection refers to locating individuals within the frame using bounding boxes, while behavior detection denotes recognizing high-risk actions—such as self-harm, escape-attempt, and fall-down—by analyzing body joint positions and motion patterns. JC yielded the lowest MOS of only 2.2, with human detection and pose estimation accuracies of 0.55 and 0.52, respectively. MB similarly exhibited a low MOS of 2.4 and reduced behavior detection performance to 0.50. In contrast,

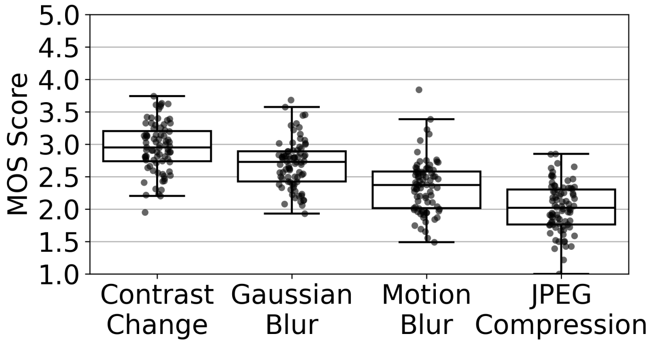


Fig. 3: Distribution of MOS for each degradation type applied to the dataset. MOS was rated on a 5-point scale from 1 to 5.

CC maintained a relatively higher MOS of 3.1 and caused only moderate decreases in recognition accuracy. This quantitative analysis elucidates the hierarchical impact of each degradation on algorithm performance and suggests that combining image restoration and enhancement techniques is essential to achieve robustness in real-world settings. Accordingly, the diverse degradation conditions and associated performance results included in this database can serve as a valuable reference for evaluating and improving model resilience in future studies.

C. Subjective Quality Assessment

Figure 3 visualizes the distribution of MOS scores under each degradation condition. To establish a reliable subjective quality benchmark, this assessment was conducted independently of the algorithm performance analysis. A total of 57 participants rated each image on a five-point scale ranging from 1 (very poor quality) to 5 (very high quality). All evaluations were performed individually in controlled environments to minimize bias.

The results showed that JC and MB yielded the lowest MOS scores. CC maintained a relatively high MOS, suggesting that contrast reduction alone had a less detrimental impact on perceived image quality. GB and MB led to lower scores due to the loss of contours and detail, while JC was rated lowest overall, primarily because of block noise and fine texture loss. Overall, these findings confirm that hospital CCTV footage is highly vulnerable to various distortions, particularly JC and MB, which substantially degrade perceived quality. This analysis underscores the necessity of integrating restoration and enhancement techniques and provides important context for interpreting quantitative performance metrics.

IV. Conclusions

In this study, we systematically analyzed the impact of various image degradations in psychiatric hospital CCTV environments on the performance of computer vision tasks and newly constructed the HIQA-DB for this purpose. HIQA-DB consists of original images captured from real CCTV footage, as well as images consistently processed with four representative types of degradation. Using state-of-the-art models such as YOLO v9 and HRNet, we quantitatively evaluated the changes in detection and recognition performance across different quality levels. The results confirmed that as image quality deteriorates, overall performance decreases sharply. In particular, JC and MB conditions showed the most significant declines in both the MOS and algorithm accuracy. HIQA-DB can serve as a valuable resource for evaluating the robustness of computer vision models in environments similar to psychiatric hospital CCTV and for developing methods to restore image quality and improve recognition performance. Furthermore, the experimental results presented in this study are expected to contribute to establishing a foundation for enhancing the reliability and efficiency of real-world medical monitoring systems. The proposed dataset will be publicly released soon to support further research.

Acknowledgment

This work was supported by the Grant of the Korea Health Technology Research and Development Project through the Korea Health Industry Development Institute (KHIDI), Ministry of Health and Welfare, Republic of Korea, under Grant HI22C1884.

References

[1] H. Song, J. Kang, and T. Kim, “Real-time abnormal behavior recognition for patient monitoring in hospitals,” in *2024 IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2024, pp. 1–8. doi: 10.1109/AVSS61716.2024.10672574.

[2] T. Kim, “Ai feedback architecture of video surveillance system,” in *2023 International Conference on Electronics, Information, and Communication (ICEIC)*, 2023, pp. 1–4. doi: 10.1109/ICEIC57457.2023.10049874.

[3] C.-Y. Wang, I.-H. Yeh, and H.-Y. Mark Liao, “Yolov9: Learning what you want to learn using programmable gradient information,” in *European conference on computer vision*, Springer, 2024, pp. 1–21.

[4] J. Wang, K. Sun, T. Cheng, *et al.*, “Deep high-resolution representation learning for visual recognition,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 10, pp. 3349–3364, 2020.

[5] A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-reference image quality assessment in the spatial domain,” *IEEE Transactions on image processing*, vol. 21, no. 12, pp. 4695–4708, 2012.

[6] Y. Han and T. Kim, “New abnormal behavior detection for patient surveillance system,” in *2024 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2024, pp. 1–5. doi: 10.1109/APSIPAASC63619.2025.10849331.

[7] Z. Chen, T. Jiang, and Y. Tian, “Quality assessment for comparing image enhancement algorithms,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 3003–3010.

[8] S. Athar and Z. Wang, “A comprehensive performance evaluation of image quality assessment algorithms,” *Ieee Access*, vol. 7, pp. 140 030–140 070, 2019.