

CQAD: An Image Quality Assessment Dataset for CCTV

Yujin Han¹, Taewan Kim^{1*}

Abstract

This paper presents the CCTV quality assessment dataset (CQAD), a novel image quality assessment (IQA) dataset specifically designed to reflect the complexities of real-world surveillance environments. Existing IQA datasets are largely based on natural images and fail to capture the diverse conditions and degradation characteristics commonly encountered in CCTV footage, such as varying illumination, fixed viewpoints, and environmental noise. To address this gap, CQAD comprises 120 reference images collected from actual surveillance cameras across a range of indoor and outdoor locations, captured under both day and night conditions. Each reference image was degraded using one of six common distortion types, and subjective quality ratings were obtained from 55 human participants using a mean opinion score (MOS) framework. Experimental analysis demonstrates that perceived image quality is affected not only by the type of distortion, but also by scene context, lighting conditions, and time of day. CQAD offers a valuable benchmark for the development of scene-aware IQA models, the evaluation of AI-based video analysis robustness, and the design of restoration techniques tailored to practical surveillance applications.

Key Words: CCTV Surveillance, Distortion Analysis, Image Quality Assessment, Mean Opinion Score, Subjective Quality Evaluation.

I. INTRODUCTION

Video surveillance plays a critical role in a wide range of domains, including public security, crime prevention, and behavioral analysis. With the rapid advancement of AI-based video analysis technologies, the importance of closed-circuit television (CCTV) footage has grown significantly [1-2]. Although modern CCTV systems often support high-resolution imaging, the quality of recorded video is frequently degraded in real-world environments due to factors such as storage compression, bandwidth limitations during network transmission, and the effects of long-term recording [3]. Such degradation impacts not only the visual clarity of footage but also the performance of AI systems that rely on accurate video interpretation [4-5].

For example, even within the same scene, variations in compression levels or the presence of noise can significantly affect the accuracy of AI tasks such as object detection, pose estimation, and abnormal behavior recognition. This implies that beyond perceptual image quality, there exists a *functional* quality dimension that determines whether AI systems can reliably extract meaningful information from video content. Therefore, in CCTV contexts, it is essential to assess video quality not solely based on human

visual perception, but from a practical standpoint specifically, whether the footage remains usable for AI-based analysis.

Despite the growing need for such assessments, most existing image quality assessment (IQA) research has focused on natural images or high-quality content. The datasets commonly used in these studies fail to capture the unique characteristics of CCTV footage, which typically includes challenging conditions such as low lighting, fixed camera positions, long and continuous recordings, low resolutions, and high-contrast scenes.

Fig. 1. illustrates how pose estimation accuracy fluctuates under different JPEG compression levels. While the algorithm performs reliably at a JPEG quality of 80, it generates fragmented skeletons at quality 60. Interestingly, the performance slightly improves at quality 40. This non-linear behavior highlights the inadequacy of conventional IQA datasets in predicting the practical usability of CCTV footage for AI applications.

Currently, there is no publicly available IQA dataset based on authentic CCTV recordings that reflect these real-world conditions. This absence constitutes a significant gap, particularly for applications in surveillance and security where reliable video quality assessment is critical. To ad-

Manuscript received May 18, 2025; Revised June 19, 2025; Accepted July 14, 2025. (ID No. JMIS-25M-05-015)

Corresponding Author (*): Taewan Kim, +82-0-940-4751, kimtwan21@dongduk.ac.kr

¹Division of Interdisciplinary Studies in Cultural Intelligence, Dongduk Women's University, Seoul, Korea, hanyujinius@gmail.com, kimtwan21@dongduk.ac.kr

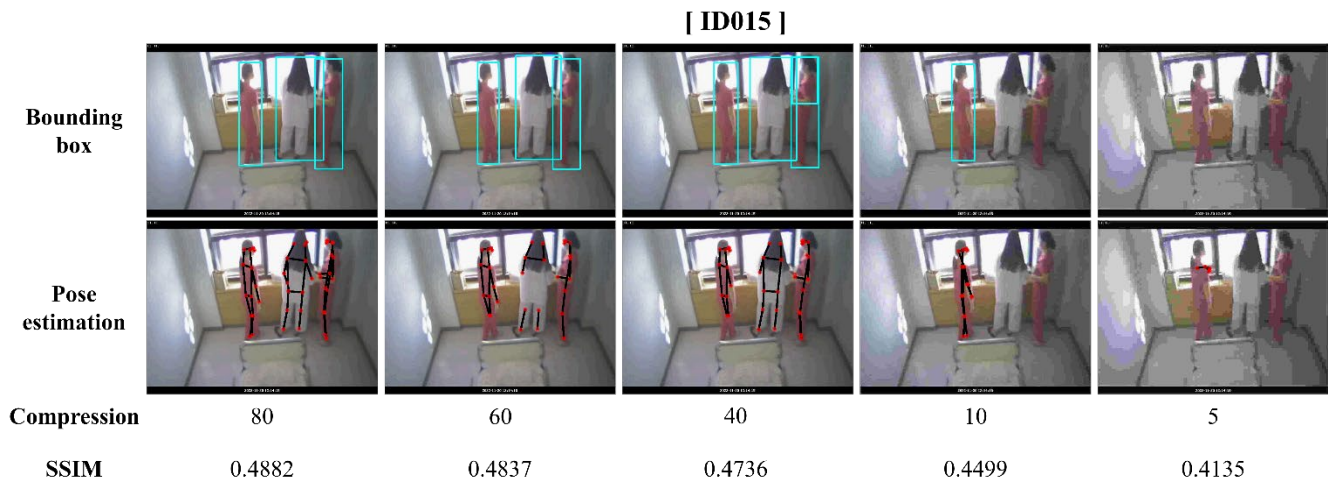


Fig. 1. Bounding box and pose estimation results under different compression levels of “ID015.jpg” of the proposed database.

dress this challenge, we introduce CQAD (CCTV quality assessment dataset), a new IQA dataset constructed from real-world CCTV footage captured in diverse surveillance settings. CQAD includes six types of common distortions and encompasses various scene types (e.g., indoor/outdoor, day/night). It also accounts for surveillance-specific features such as fixed viewpoints and repeated scene patterns characteristics often overlooked in existing datasets.

CQAD is designed to serve as a benchmark for developing quality prediction models tailored to surveillance scenarios. Furthermore, it provides a valuable resource for evaluating the robustness of AI-based video analysis systems and for designing restoration algorithms that reflect the constraints of real-world operational environments.

II. RELATED WORKS

2.1. IQA Datasets

IQA has been an active research area for decades, resulting in the development of numerous datasets designed to evaluate perceptual quality under control and synthetic distortion conditions. Notable examples include LIVE [6],

CSIQ [7], MDIQ [8], TID2013 [9], MDID [10], KADID-10k [11], and KonIQ-10k [12], as summarized in Table 1. These datasets have played a crucial role in the advancement of quality prediction models for natural images, offering a broad range of distortion types such as Gaussian noise, JPEG compression, and blurring. Although many of the commonly used IQA datasets were introduced several years ago, they continue to serve as standard benchmarks due to their methodological clarity and accessibility.

However, a major limitation shared across these datasets is their lack of relevance to real-world CCTV environments. The degradation characteristics in surveillance footage are often more complex, irregular, and context-dependent than those simulated in controlled datasets. Typical CCTV footage includes unique artifacts caused by long-term recording, environmental conditions (e.g., lighting variation, motion blur, low resolution), and hardware constraints, which are not adequately captured in existing IQA datasets.

2.2. CCTV Quality Assessment

Although several studies have investigated quality degradation in CCTV systems, most of them fall under the broader category of video quality assessment (VQA),

Table 1. Summary of IQA datasets including CQAD.

Database	LIVE	CSIQ	MDIQ	TID2013	MDID	KADID-10k	KonIQ-10k	CQAD
Year	2006	2010	2012	2013	2017	2019	2020	2025
# of ref. images	29	30	15	25	20	81	10,073	120
# of dist. images	779	866	405	3,000	1,600	10,125	N/A	720
Distortion types	5	6	3	24	5	25	Authentic	6
# of subjects	20–29	25	37	971	192	2,212	1,459	55
Range of scores	0–100	0–1	0–100	0–9	0–8	1–5	1–5	1-5
Image format	BMP	PNG	BMP	BMP	BMP	N/A	N/A	JPG
Subjective data type	DMOS	DMOS	DMOS	MOS	MOS	DCR	MOS	MOS

where the evaluation is conducted at the video sequence level rather than on individual frames or still images. These VQA approaches often focus on degradation caused by compression artifacts, packet loss during transmission, or bitrate adaptation.

Some efforts have been made to address surveillance-specific conditions. For instance, [13] proposed a deep learning-based VQA model tailored for surveillance applications, while [14] introduced a task-driven VQA framework that uses the performance of object detection algorithms as an indirect indicator of video quality. While promising, these approaches primarily emphasize task-oriented or global quality metrics and do not incorporate frame-level subjective quality annotations or Mean Opinion Score (MOS)-based datasets specifically designed for CCTV contexts.

Furthermore, none of the existing studies provide a publicly available image-level dataset that reflects the diverse and nuanced distortion patterns encountered in operational surveillance environments. As such, there remains a significant gap in the availability of datasets that capture the perceptual and functional quality of CCTV images, both from human and AI perspectives.

Moreover, while several in-the-wild video datasets have recently been proposed for surveillance applications, most of them target VQA tasks and do not provide frame-level annotations necessary for developing CCTV specific IQA models.

III. CQAD: A CCTV-SPECIFIC IQA DATASET

CCTV imagery exhibits characteristics that are fundamentally different from those of natural photographic images. Surveillance cameras are typically installed in fixed positions and operate continuously over long durations, capturing repeated scenes under diverse temporal and environmental conditions. Consequently, even footage recorded

at the same location can show substantial quality variation depending on factors such as time of day, weather, and lighting conditions.

Daytime recordings usually provide clearer visibility and more distinguishable objects. In contrast, nighttime footage often suffers from low illumination and increased visual artifacts due to the use of infrared (IR) lighting. IR illumination can interact with airborne particles or dust on the camera lens, resulting in scattered reflections and widespread visual noise conditions rarely represented in conventional IQA datasets. These environmental and physical influences introduce degradation patterns that differ significantly from the artificial distortions (e.g., Gaussian noise or JPEG compression) commonly used in traditional quality assessment studies.

Moreover, the operational context of CCTV use further distinguishes it from general-purpose photography. Surveillance footage is often evaluated not solely by visual clarity, but by its usability for critical applications such as real-time monitoring, anomaly detection, and forensic analysis. In such contexts, image quality must be considered from a practical and application-driven perspective.

Despite this, existing IQA datasets largely neglect the temporal, environmental, and functional constraints specific to surveillance video. Variations in lighting (e.g., day vs. night), setting (indoor vs. outdoor), and camera installation (e.g., fixed viewpoints) significantly affect perceived quality and system performance. Thus, a meaningful quality assessment framework for CCTV should extend beyond simple perceptual sharpness and incorporate context-aware criteria that reflect the operational goals of security and monitoring systems.

To fill this gap, we introduce CQAD (CCTV quality assessment dataset), which is specifically designed to capture the unique characteristics and challenges of real-world CCTV footage as shown in Fig. 2. The dataset includes samples recorded across a wide range of surveillance environments and times, encompassing both daytime and

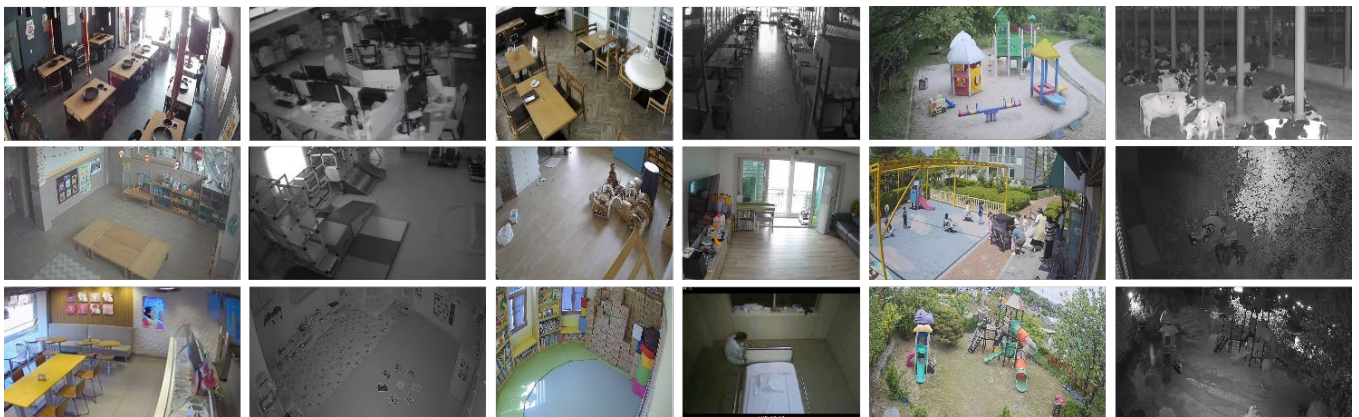


Fig. 2. Representative examples from the CQAD dataset.

nighttime scenes, as well as indoor and outdoor locations. It is constructed with the goal of enabling robust image quality prediction and system evaluation in practical surveillance scenarios.

3.1. Dataset Details

In this study, we present CQAD, a novel IQA dataset constructed from real CCTV footage captured in diverse surveillance environments. The source images were collected from a variety of indoor and outdoor locations, including hospitals, kindergartens, residential homes, retail stores, livestock facilities, alleyways, and building entrances, using operational surveillance systems. All data collection was conducted with prior consent from the respective institutions and in strict adherence to ethical guidelines. To ensure privacy protection, all images were manually reviewed to exclude any content containing personally identifiable information.

CQAD was specifically designed to capture the distinctive characteristics of CCTV image quality, incorporating variability across time, location, and environmental conditions. To reflect practical differences in lighting, camera angles, and background content, the dataset is categorized into four representative scene types of indoor-day (ID), indoor-night (IN), outdoor-day (OD), and outdoor-night (ON) as shown in Table 2.

A total of 120 reference images were selected from the collected footage. To simulate quality degradation commonly observed in real CCTV environments, each reference image was processed with six types of single distortions: contrast change (CC), gaussian blur (GB), Gaussian noise (GN), jpeg compression (JP), motion blur (MB), salt-and-pepper noise (SP).

Each image in the dataset was assigned to a unique identifier based on its scene type and applied distortion, allowing for precise indexing and traceability. The dataset is organized to support both qualitative distortion-wise analysis and the development of data-driven models for automated quality prediction. By reflecting real-world surveillance conditions and degradation types, CQAD aims to serve as a reliable benchmark for evaluating IQA algorithms in security-focused applications.

3.2. Subjective Evaluation using MOS

To quantitatively assess image quality and investigate

Table 2. Four scene types of CQAD.

	Indoor		Outdoor
Day	ID		OD
Night	IN		ON

perceptual patterns in the CQAD dataset, a MOS-based subjective evaluation was conducted. A total of 55 participants were recruited for the study, and each subject independently evaluated images under standardized and controlled viewing conditions to ensure consistency across sessions. Participants rated image quality using a five-point Likert scale:

- 5: Excellent quality
- 4: Good quality
- 3: Fair/average quality
- 2: Poor quality
- 1: Very poor quality

The set comprised both reference images and their distorted counterparts, randomly mixed to eliminate ordering effects. Each participant assessed a total of 840 images, covering a wide range of scene types and distortion categories. For each image, the final MOS was calculated by averaging the individual ratings across all valid responses. To enhance the reliability of the results, statistical outlier detection was applied, and anomalous scores defined based on deviation from group consensus were excluded according to predefined criteria.

The resulting MOS values provide a ground truth for analyzing perceptual tendencies across different distortion types and scene conditions. These scores also form the basis for benchmarking and training machine learning models focused on surveillance-oriented image quality prediction. The insights gained from this evaluation highlight how human observers perceive quality variations in CCTV imagery and underscore the limitations of general-purpose IQA methods in such specialized contexts.

3.3. Analysis of Subjective Quality Score

Using the collected MOS, we analyzed perceptual quality trends across different distortion types and environmental conditions. Figs. 3 and Fig. 4 present the average MOS and standard deviation for each distortion type under vary-

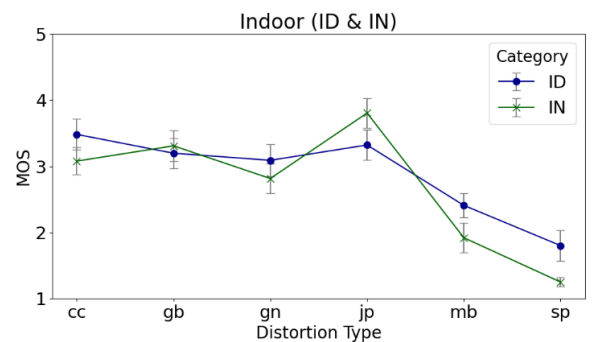


Fig. 3. Average MOS for each distortion in indoor.

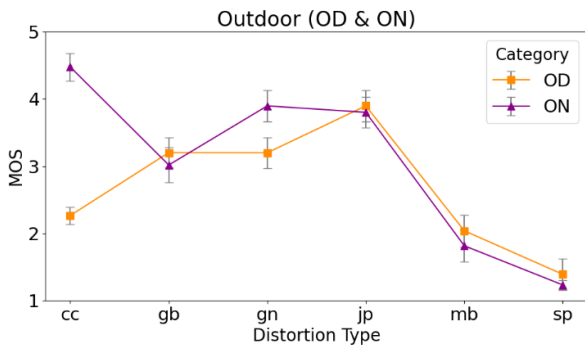


Fig. 4. Average MOS for each distortion in outdoor.

ing lighting conditions in indoor and outdoor scenes, respectively.

3.3.1. Indoor Analysis

As illustrated in Fig. 3, most distortion types in indoor scenes received higher MOS values during daytime compared to nighttime. This discrepancy is primarily attributed to enhanced ambient lighting during the day, which improves visual clarity. In contrast, nighttime footage often suffers from reduced illumination, the activation of IR lighting, and increased visual noise, all of which contribute to diminished perceived image quality.

Among the distortions, salt-and-pepper noise (SP) exhibited the most pronounced quality degradation under night conditions, with an average MOS of approximately 1.5, the lowest among all distortion types. Conversely, Gaussian blur (GB) and Gaussian noise (GN) showed relatively minor differences between day and night, suggesting these distortions are less influenced by lighting conditions. Interestingly, for jpeg compression (JP), nighttime conditions resulted in slightly higher MOS scores than daytime. This may indicate that under low-light conditions, compression artifacts are less visually disruptive or are masked by the overall degradation in image clarity.

3.3.2. Outdoor Analysis

In contrast, Figure 4 reveals a different trend for outdoor scenes. Notably, some distortion types received higher MOS scores at night than during the day. For example, contrast change (CC) achieved an average MOS of 4.5 at night, significantly outperforming its daytime counterpart. This may be due to the inherently low contrast of night scenes, where moderate contrast adjustments can enhance visual clarity.

JPEG compression (JP) also maintained relatively high MOS scores in both lighting conditions, indicating that compression artifacts had limited impact on perceived quality in outdoor contexts. However, more severe distortions such as motion blur (MB) and salt-and-pepper noise (SP)

consistently received low MOS values, regardless of lighting condition. SP again had the lowest MOS under outdoor night conditions, highlighting its substantial effect on object recognizability and overall scene interpretability.

These results demonstrate that perceived image quality in CCTV footage is influenced not only by the type of distortion, but also by scene context, lighting conditions, and time of day. Even when the same distortion is applied, its impact on perceptual quality can vary significantly based on environmental factors. This underscores the context-sensitive nature of subjective evaluation in surveillance applications. Therefore, when developing CCTV-specific image quality assessment models, it is essential to move beyond distortion-centric approaches. Instead, models should incorporate scene-aware and time-sensitive features to more accurately reflect human perception under realistic operational conditions.

IV. CONCLUSION

In this study, we introduced CQAD, a novel IQA dataset specifically designed for surveillance and security applications. Unlike conventional IQA datasets that are largely composed of natural images under controlled conditions, CQAD is constructed from real-world CCTV footage captured in diverse operational environments, spanning different scene types (indoor/outdoor) and lighting conditions (day/night).

The dataset includes six types of single distortions applied to selected reference images, and perceptual quality was evaluated through a MOS-based subjective assessment involving human participants. Our analysis revealed that perceived image quality varies significantly depending on the distortion type, scene context, time of recording, and illumination, confirming that quality perception in surveillance scenarios is highly context-dependent and shaped by environmental realism.

CQAD fills a critical gap in the current landscape of IQA research by providing a benchmark tailored to the unique challenges of CCTV imagery. Beyond conventional perceptual evaluation, the dataset offers a platform for advancing task-aware and utility-driven quality assessment particularly for applications involving object detection, behavior recognition, and security monitoring. By capturing realistic degradation patterns commonly encountered in surveillance footage, CQAD supports the development of scene-aware, application-oriented, and automated IQA models. We believe this dataset will serve as a valuable resource for both academic research and practical deployments in video analytics and surveillance system design.

ACKNOWLEDGEMENT

This study was supported by the Dongduk Women's University grant.

REFERENCES

- [1] B. Kwon and T. Kim, "Toward an online continual learning architecture for intrusion detection of video surveillance," *IEEE Access*, vol. 10, pp. 89732-89744, 2022.
- [2] H. Song, J. Kang, and T. Kim, "Continual learning based personalized abnormal behavior recognition alarm system", *APSIPA Transactions on Signal and Information Processing*, vol. 13, no. 1, 2024.
- [3] T. Kim, C. Kang, Y. Kim, and S. Yang, "AI camera: Real-time license plate number recognition on device", in *IEEE International Conference on Consumer Electronics (ICCE 2022)*, Las Vegas, NV, 7-9, Jan. 2022.
- [4] S. Ai and J. Kwon, "Extreme low-light image enhancement for surveillance cameras using attention U-Net," *Sensors*, vol. 20, no. 2, p. 495, 2020.
- [5] N. Chahine, S. Calarasanu, D. Garcia-Civiero, T. Cayla, S. Ferradans, and J. Ponce, "An image quality assessment dataset for portraits," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9968-9978, 2023.
- [6] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440-3451, 2006.
- [7] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, pp. 011006-1–011006-12, 2010.
- [8] A. Lahoulou, A. Bouridane, E. Viennet, and M. Haddadi, "Full-reference image quality metrics performance evaluation over image quality databases," *Arabian Journal for Science and Engineering*, vol. 38, no. 9, pp. 2327-2356, 2013.
- [9] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, and J. Astola, et al., "Color image database TID2013: Peculiarities and preliminary results," in *European Workshop on Visual Information Processing (EUVIP)*, pp. 106-111, 2013.
- [10] W. Sun, F. Zhou, and Q. Liao, "MDID: A multiply distorted image database for image quality assessment," *Pattern Recognition*, vol. 61, pp. 153-168, 2017.
- [11] H. Lin, V. Hosu, and D. Saupe, "KADID-10k: A large-scale artificially distorted IQA database," in *Proceedings of the Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1-3, 2019.
- [12] V. Hosu, H. Lin, T. Sziranyi, and D. Saupe, "KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment," *IEEE Transactions on Image Processing*, vol. 29, pp. 4041-4056, 2020.
- [13] D. Ding, J. Tong, and L. Kong, "A deep learning approach for quality enhancement of surveillance video," *Journal of Intelligent Transportation Systems*, vol. 24, no. 3, pp. 304-314, 2020.
- [14] J. Nightingale, Q. Wang, and C. Grecos, "Task-oriented quality assessment and adaptation in real-time mission critical video streaming applications," *Real-Time Image and Video Processing 2015*, vol. 9400, pp. 174-182, 2015.

AUTHORS



Yujin Han is a fourth-year student in the Division of Interdisciplinary Studies in Cultural Intelligence at Dongduk Women's University, Seoul, South Korea. Currently, she is pursuing her B.S. degree in Data Science. Her research interests include computer vision and machine learning.



Taewan Kim received the B.S., M.S., and Ph.D. degrees in electrical and electronic engineering from Yonsei University, Seoul, South Korea, in 2008, 2010, and 2015, respectively. From 2015 to 2021, he was with the Vision AI Laboratory, SK Telecom, Seoul. In 2022, he joined as a faculty with the Division of Future Convergence (Data Science Major), Dongduk Women's University, Seoul, where he is currently an Assistant Professor. His research interests include computer vision and machine learning including continual and online learning.