IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 29 November 2023, accepted 29 December 2023, date of publication 10 January 2024,
date of current version 19 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3352146

RESEARCH ARTICLE

Balancing Privacy and Accuracy: Exploring the Impact of Data Anonymization on Deep Learning Models in Computer Vision

JUN HA LEE AND SU JEONG YOU

AI Robot Research and Development Department, Korea Institute of Industrial Technology, Sangnok-gu, Ansan-si, Gyeonggi-do 15588, South Korea

Corresponding author: Su Jeong You (sjyou21@kitech.re.kr)

This work was supported by the Institute of Information and Communication Planning and Evaluation (IITP) Grant funded by the Korean Government [Ministry of Science and ICT (MSIT)], Development of Personal Information Processing Technology for Data Collection/Utilization for Autonomous Driving, under Grant 2021-0-01062.

ABSTRACT Computer vision has become indispensable in various applications, including autonomous driving, medical imaging, security and surveillance, robotics, and pattern recognition. In recent years, the quality of training data has emerged as a critical factor for ensuring effectiveness in real-world scenarios. However, the increasing stringency of privacy regulations in various regions necessitates careful handling of collected images for computer vision. Personal information within images is typically anonymized by applying anonymization patterns to remove it. Empirical findings underscore the significant influence of data quality on the training of deep learning models. Striking the right balance between privacy and recognition performance becomes paramount. Therefore, it is essential to understand how the anonymization of image datasets affects deep learning model performance. In this paper, we thoroughly analyze the effects of different anonymization techniques on the performance of deep learning-based models in computer vision tasks, with a particular emphasis on presenting a model-centric perspective, such as the type of deep learning model and the number of parameters. We aim to provide valuable insights and guidelines for selecting the optimal level of anonymization that strikes a balance between recognition accuracy and privacy protection.

INDEX TERMS Computer vision, privacy, anonymization, recognition performance, model-centric perspective.

I. INTRODUCTION

Collecting and storing vast quantities of visual data is essential to developing robust and effective computer vision algorithms [1], [2], [3]. However, obtaining explicit consent for all collected data presents significant challenges throughout the process, and there are also doubts about the practicality of obtaining consent from the data subjects. For this reason, many countries have implemented data privacy regulations that protect the rights of data subjects through legislative means. The European Union (EU) replaced its 1995 data protection directive with the General Data Protection Regulation (GDPR) [4] in May 2018. The GDPR forbids the processing of personal information solely on implicit consent and bolsters the decision-making rights of data subjects. The California Consumer Privacy Act (CCPA) [5] imposes further regulations on data handling and privacy practices, reinforcing the need for stringent consent and data protection measures. Clear criteria and protocols for data anonymization should be established to ensure compliance with regulations when building large-scale visual datasets. Strict compliance with these prescribed guidelines is essential to safeguard individual privacy rights and adhere to legal mandates. However, legal constraints can challenge research and development, especially in the data-dependent field of computer vision [6]. In compliance with GDPR, as outlined in Article 4(1), personal data are defined as "*any information that can identify a person is personal data*". In other words, the GDPR does not apply if data are rendered anonymous so that the data subject is not or is no longer

The associate editor coordinating the review of this manuscript and approving it for publication was Szidonia Lefkovits.

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

identifiable (Recital 26). Consequently, practitioners have adopted traditional anonymization methods (e.g. blurring) to collect training data [7], [8].

The application of anonymization techniques inherently bears the risk of degrading the quality of the training dataset, potentially exerting an adverse impact on the efficacy of image recognition models [9]. Nonetheless, traditional anonymization methods remain the standard approach for concealing personal information within data, and their impact on model learning performance still requires more research to be conclusive. Previous studies primarily focused on analyzing the impacts of anonymization, especially related to human faces [10], [11]. To sum up, the results of their research indicate that facial anonymization can impact object classes related to identifying humans to some degree [12]. However, given the diverse nature of personal information in data and the risk of re-identification through information fusion [13], [14], it is imperative to assess the impact on a broader range of data subjects, including but not limited to human faces. In this paper, we evaluate the impact of data anonymization on model training using the MS COCO dataset [15]. Following the model training phase, we comprehensively evaluate the models using the original validation dataset. In summary, our main contributions are:

- We analyze the impact of data anonymization on deep learning-based model training for computer vision tasks.
- From a model-centric perspective, we present comprehensive insights into data anonymization. Our evaluation involves a comparison of the impact of data anonymization on both model scale and architecture. Furthermore, we analyze the inter-class influence that arises as a result of anonymization.

Our goal is to promote a better understanding of the effects of anonymization while also facilitating a balanced approach to model performance and privacy protection. The rest of the paper is organized as follows. Section II provides a summary of related work on privacy guarantees of anonymization. Section III presents the experimental setup used to assess the impact of anonymized data on deep learning models. Quantitative and qualitative results and analysis on the MS COCO dataset are presented in Section IV. Finally, Section V concludes the paper.

II. RELATED WORK

A. IMAGE ANONYMIZATION

The main goal of image anonymization is to remove privacy-sensitive information from the image. Image anonymization can involve various techniques, including traditional methods and synthetic data generation.

1) TRADITIONAL METHODS

Traditional methods for anonymization, including techniques like blurring, pixelation, distortion, mask-out methods as referenced in [16], and even k-means clustering as discussed in [17], find widespread usage in real-world situations. Given the image I and privacy-critical area mask M indicating the region to be anonymized, the anonymized images can be defined as follows:

Blurring: Gaussian blur defines the anonymized image as $I_{anonym} = I \odot (1 - M) + M \odot I_{blur}$, where $\odot$ represents elementwise multiplication, and I_{blur} is the blurred image obtained by convolving the original image I with the Gaussian filter $G_\sigma(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$, here σ represents the standard deviation. ($\sigma = 8$)

Pixelation: Given the pixelation ratio r, pixelation is achieved by first resizing the image I with width w_I and height h_I using nearest-neighbor interpolation to create a smaller block of pixels, denoted as B. This resized block B has dimensions $\frac{w_I}{r} \times \frac{h_I}{r}$. Then, the block B is resized back to the original image dimensions, $w_I \times h_I$, again using nearest-neighbor interpolation to create the pixelated image $I_{pixelated}$. Pixelation defines the anonymized image as $I_{anonym} = I \odot (1 - M) + M \odot I_{pixelated}$. (pixel_ratio $= 0.1$)

Distortion: Elastic deformation [18] defines the transformed images as $I_{trans} = I(x + \Delta x,\ y + \Delta y)$, where I_{trans} denotes the pixel value in the transformed image, $I(x, y)$ denotes the pixel value in the original image at position (x, y), and Δx and Δy are displacement fields. The displacement fields are generated as follows:

$$\Delta x = \alpha \cdot G_\sigma * U_x(h_I, w_I), \quad \Delta y = \alpha \cdot G_\sigma * U_y(h_I, w_I) \quad (1)$$

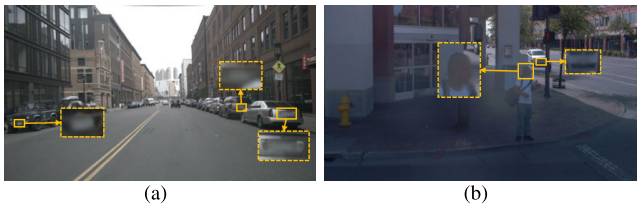
where $U(h_I, w_I)$ is a uniform random field generated over the range $[-1, 1]$ with dimensions specified by image resolution, and G_σ represents the Gaussian filter operation with standard deviation σ. The operation $G_\sigma * U(h_I, w_I)$ implies the convolution of the Gaussian filter G_σ with the uniform random field $U(h_I, w_I)$. The parameter α controls the intensity of the deformation. The affine transformation is incorporated as an initial step before applying the elastic deformation. Finally, the anonymized images are defined as $I_{anonym} = I \odot (1 - M) + I_{trans} \odot M$. ($\sigma = 8$)

Mask-out: Mask-out defines the anonymized image as $I_{anonym} = I \odot (1 - M) + M \odot 0$ [19].

2) SYNTHETIC DATA GENERATION

Some studies have observed negative consequences resulting from traditional anonymization methods as follows. Traditional anonymization may impair the intrinsic value of the original data [20]. Dvořáček and Hurtik [21] and Zhou and Beyerer [22] demonstrated that the performance, particularly object detection and instance segmentation, are adversely impacted when the urban scenes dataset is anonymized. Hukkelås et al. [23] reported declining instance segmentation performance on the anonymized autonomous driving dataset. Therefore, recent research efforts have been focused on methods aimed at substituting identifiable objects with synthetic images, effectively impeding the potential for data restoration while preserving its intrinsic value [24].

The generation of synthetic images primarily relies on Generative Adversarial Networks (GANs) [25], a class of



(a) (b)

FIGURE 1. Anonymization of faces and license plates in public autonomous driving datasets. (a) NuScenes [7]. (b) Waymo [8].

generative models, and has witnessed significant advancements in the quality of generated data. There are two primary approaches to anonymizing the data using synthetic imagery [26]: one entails a complete substitution of the identifiable information with synthetic images, while the other method concentrates on modifying specific features of the identifiable information. As an example of the method involving a complete substitution of the identifiable information with synthetic images, techniques like Face Shifting [27] are employed to replace the input face with the target face, or the generator of StyleGAN [28] is utilized to transform them into synthetic images in the desired style. A method also goes beyond the face and replaces the entire individual. The CSE-detector neural network can estimate the correspondence between a body shape and the model of the human body surface within the image. Based on this estimated body surface model and their posture, a GAN can be employed to generate virtual individuals [23]. Another approach involves altering only the target features. For example, manipulating facial features within the embedding vector space of GAN-based face generation techniques [29] or generating virtual faces to obliterate personal identification information [30]. The quality of these approaches has been improving annually, and they offer the advantage of preserving the value of data due to their resemblance to authentic images. However, GANs suffer from limitations such as a restricted number of generatable samples and the occurrence of mode collapse. Recent concerns have also been raised about the potential for restoration [31]. Conventional anonymization techniques are considered for data acquisition because they are still valid regarding processing cost and safety.

B. PUBLIC ANONYMIZED DATASETS

Autonomous driving datasets consist of image data collected in urban environments, where pedestrians and vehicles are frequently present, posing numerous anonymization challenges. NuScenes [7], for instance, comprises images from vehicles driving in Singapore and Boston, where anonymization is carried out by blurring faces and license plates by privacy regulations. Moreover, it grants data subjects the authority to request the removal of personal information from the data. The Waymo Open dataset [8], collected by Alphabet Inc., similarly employs facial and license plate blurring, affording data subjects the right to preserve privacy within the data. With the ongoing trend

TABLE 1. Evaluation result of the impact of anonymization of faces and license plates in autonomous driving datasets. For the extraction of anonymized regions, faces (categorized as human) and license plates (categorized as vehicles) are detected within the Cityscapes dataset [34]. The bounding boxes identified are then anonymized using the methods described in Section II-A1–blurring (blur), pixelation (pixel), distortion (disto), and mask-out (mask). The model under evaluation is the YOLOv8m, which is trained using datasets anonymized by four distinct methods. The relative Average Precision (AP) change is compared to the YOLOv8m model trained on the original Cityscapes dataset. Our results show that the impact of anonymizing faces and license plates has a restricted influence on model training.

	Object Detection			
	blur	pixel	disto	mask
Category				
Human	-0.14%	-0.21%	-0.20%	**-0.23%**
Vehicle	-0.02%	-0.05%	-0.06%	**-0.06%**
bbox AP	-0.01%	-0.03%	-0.02%	**-0.04%**

of anonymized data in open datasets, it becomes crucial to evaluate the impact of such anonymization on model training thoroughly. Our study is indispensable in deep learning technologies in developing robust and reliable systems founded on data that respects individual privacy.

C. PRIVACY-PRESERVING IN LARGE SCALE VISUAL DATASETS

In general, the areas in images that necessitate privacy protection involve human faces and vehicle license plates [32], [33]. Identifying regions containing personal information, such as faces and license plates, is a fundamental requirement for anonymization. However, faces and license plates have limited variations in aspect ratio and do not occupy a significant portion of the image. This limitation becomes particularly pronounced in autonomous driving datasets like Cityscapes [34] and BDD100k [35]. Hukkelås et al. [19] and Zhou et al. [22] reported that the anonymization of faces and license plates within training data exerts a restricted influence on model training.

In addition to the face, a gait includes unique behavioral biometric characteristics that can be utilized for personal identification [36]. Gait information includes walking and features such as leg length, height, and even aspects of appearance such as age and gender [37], [38]. With advancements in computer vision technology, it is now possible to recognize individuals based on their gait, even when their faces are not visible [39]. Silhouettes, whether utilized independently or in conjunction with other biometric characteristics, can also serve as valuable identifiers of individuals. Details about the clothing worn can function as additional information in the process of person identification [40]. Deep learning-based pose replacement algorithms have recently garnered attention for their capacity to depersonalize biometric traits and clothing information [41], [42]. The anonymization of automobiles follows a similar trajectory to that of individuals. One primary approach entails removing distinctive vehicle identification information, such as license plates [43]. Additionally, it encompasses techniques that mitigate or

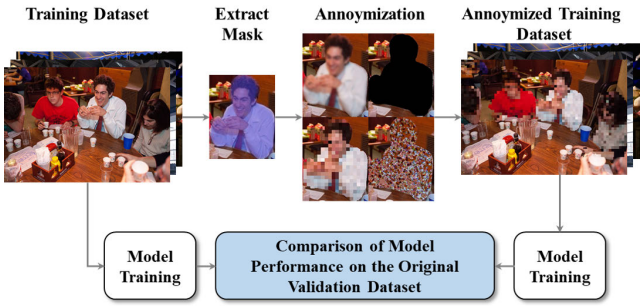


FIGURE 2. Our experimental process is designed to assess the impact of anonymized data on model training. Initially, we apply anonymization to the widely-used COCO training dataset for computer vision. Then, we proceed to train separate models, specifically YOLOv8 and RT-DETR, using both the anonymized and the original data. Finally, we evaluate the performance of these models on the original validation set, setting up a baseline with the models trained on the original data for a comparative analysis of the anonymization impact.

obfuscate the distinctive exterior features of the vehicle, making it challenging to identify the make or brand of the vehicle [44]. The methods used for information removal can be categorized, much like in human anonymization, into altering all identifying information or modifying select attributes that serve as unique identifiers.

The increasing volume of data and the emergence of new visual cues continue to heighten the risks associated with personal identification. In the case of humans, not only face information but also data like gait and silhouettes can be utilized for identity recognition. Several studies investigate the impact of full-body anonymization to address privacy concerns [23], [24]. Similarly, with vehicles, apart from license plates, aspects such as vehicle color, type, brand, or accessories can facilitate personal identification. In vehicles, personal identification is risky not only through license plates but also due to distinctive features (e.g., supercars, unique colors) and data fusion with edge infrastructure [45]. Consequently, to mitigate the privacy risks that may arise from data fusion, anonymization should extend beyond faces and license plates to encompass full-body and vehicle anonymization, as this approach proves to be more effective.

III. METHODOLOGY

A. PROBLEM FORMULATION

Fig. 2 shows our experimental process. Let the original training dataset denote by D_{orig} and the anonymized dataset by D_{anonym}. Each dataset consists of pairs of images and their associated labels: $D = \{(x_i, y_i)\}_{i=1}^{N}$, where x_i is the i-th image, y_i is the corresponding label or annotation, and N is the number of samples in the dataset.

The anonymization process can be described as a function $f(\cdot)$ that transforms the original dataset D_{orig} into the anonymized dataset D_{anonym}:

$$D_{\text{anonym}} = f(D_{\text{orig}}) \tag{2}$$

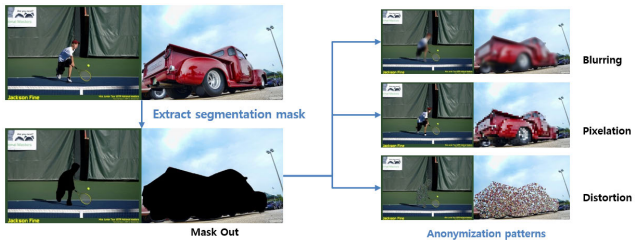


FIGURE 3. Anonymized data generation procedure: (i) an area to be anonymized is extracted with a pixel-level split mask; (ii) anonymizes the area corresponding to the split mask.

The machine learning models are trained using a training function $\mathcal{T}(\cdot)$, which takes a dataset and outputs a model M:

$$M_{\text{orig}} = \mathcal{T}(D_{\text{orig}}) \tag{3}$$

$$M_{\text{anonym}} = \mathcal{T}(D_{\text{anonym}}) \tag{4}$$

where M_{orig} and M_{anonym} are models trained on the original and anonymized datasets, respectively. The performance of the models is evaluated using a performance metric P on a validation dataset D_{val}:

$$P_{\text{orig}} = P(M_{\text{orig}},\ D_{\text{val}}) \tag{5}$$

$$P_{\text{anonym}} = P(M_{\text{anonym}},\ D_{\text{val}}) \tag{6}$$

where P_{orig} and P_{anonym} are the performance metrics of the original model M_{orig} and the anonymized model M_{anonym}, respectively, evaluated on the validation dataset D_{val}. Our goal is to assess the impact of anonymization on model performance, quantified as the difference:

$$\Delta P = P_{\text{anonym}} - P_{\text{orig}} \tag{7}$$

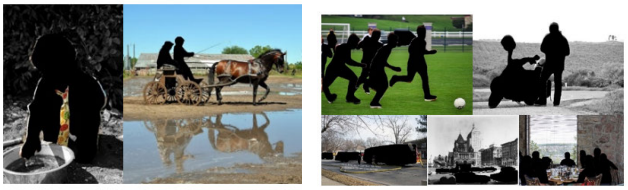
If $\Delta P < 0$, it suggests that the anonymization process has a negative impact on model performance. As ΔP approaches zero, it is considered that anonymization has little impact on model performance and $\Delta P > 0$ indicates that the anonymized model M_{anonym} performs better than the original model M_{orig}, which is a less likely occurrence.

B. EXPERIMENT SETUP

1) ANONYMIZED DATA GENERATION

In this work, we use conventional anonymization techniques such as blurring, pixelation, distortion, and mask-out to create anonymized data to analyze how this process affects model training. Fig. 3 shows our anonymized data generation pipeline. As mentioned in Section II-C, with the increasing amount of information that can be used to identify individuals, we are expanding the scope of our anonymization efforts to include full-body and vehicles, in addition to faces and license plates. To do this, we used a pipeline with instance segmentation and anonymization processing components.

We applied this pipeline to the COCO 2017 training set, which contains around 118,000 images. We used pre-trained weights of the advanced RTMDet-Ins model on COCO for pixel-level mask extraction, including full-body and vehicles.



(a) (b)

FIGURE 4. Pre-defined instance segmentation annotations for the person/car/motorcycle/bus/truck classes. (a) Aiming to minimize data loss outside the annotated classes. (b) Achieve a pixel level anonymization.

In Section II-A1, anonymize the mask extracted as mentioned using the method described.

2) DEEP NEURAL NETWORKS FOR EVALUATIONS

We use Convolutional Neural Network (CNN) [47] and Vision Transformer (ViT) [48] based models designed for diverse computer vision tasks, including object detection, instance segmentation, and keypoint detection. We assess the influence of data anonymization on model training with YOLO, employing the CNN architecture [49]. YOLO supports classification, object detection, instance segmentation, and keypoint detection (human pose estimation) tasks. The latest addition to the YOLO series, YOLOv8 [50], was unveiled by Ultralytics in January 2023. Through comparisons with its predecessors, YOLOv8 has already proven its superiority as the cutting-edge state-of-the-art in YOLO. Fig. 5 shows that when compared to conventional YOLO models, YOLOv8, trained at a resolution of 640, demonstrates superior performance with a similar number of parameters. Figure 5 indicates that YOLOv8 prioritizes high-performance inference over real-time capabilities compared to its conventional YOLO counterparts. Therefore, we conduct experiments utilizing YOLOv8, which ensures robust generalization performance on anonymized data and offers a range of model sizes required for assessing the impact of the number of parameters on anonymization.

DETR [51] is a deep learning model designed for object detection, utilizing the ViT architecture. This architecture leverages self-attention mechanisms, considering attention heads and positional information to detect objects precisely. DETR generates specific location tokens to predict the positions of individual objects and decodes the positional information to produce outputs. YOLO primarily depends on the relative positions within grid cells for acquiring object location information without significantly relying on the positional data of grid cells. Consequently, YOLO and DETR exhibit distinct architectural differences. To investigate the impact of data anonymization based on model architecture, we select RT-DETR [52], an extension of DETR, as a comparative benchmark against YOLO.

Pre-trained models are harnessed following transfer learning paradigms, where initial training on large-scale datasets facilitates the extraction of diverse image features [53], [54]. These extracted features provide a versatile foundation that,

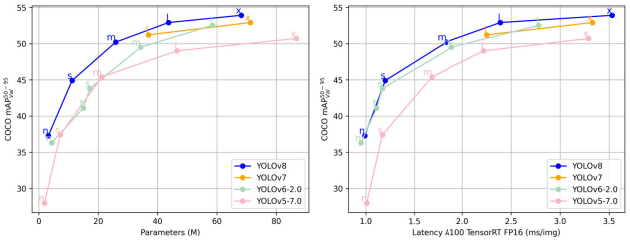


FIGURE 5. Performance comparison of each YOLO version trained on the MS COCO dataset. Compared to other versions, YOLOv8 shows superior performance despite having comparable parameters. This suggests that YOLOv8 emphasizes high-performance inference more than real-time capabilities than its conventional YOLO counterparts [50].

TABLE 2. Our hyperparameter setting for training the YOLOv8 and RT-DETR models on the anonymized and original data.

config	YOLOv8	RT-DETR
backbone	ResNet50	HGNetv2
epochs	100	
batch size	16	
image size	640	
optimizer	SGD	
learning rate	0.01	
momentum	0.937	
weight decay	0.0005	
warm-up	True	

due to their generalized nature, can be effectively modified to suit a range of novel tasks. Fine-tuning these models with user-specific data enables customization to meet the precise needs of the target domain. However, using such pre-trained models may retain features from the previous dataset in the model [55]. As a result, in our training process, we abstain from using pre-trained models and instead opt to train the model from scratch. The Table. 2 lists our hyperparameters used in this work. All experiments were conducted using PyTorch 1.9.0 with CUDA 11.1 and executed on NVIDIA RTX 3090 GPUs.

C. EVALUATION METRICS

1) AVERAGE PRECISION (AP)

In this paper, we use the mean Average Precision (mAP) matrix for analyzing the performance of each tasks [56], [57], [58]. The mAP is a widely recognized performance metric in computer vision, notable for incorporating Precision-Recall (PR) curves at different thresholds. PR curves are constructed by predictions and plotting two key metrics: precision and recall. Let Precision (P) and Recall (R) as follows: $P = \frac{TP}{TP+FP}$, $R = \frac{TP}{TP+FN}$, where TP, FP, TN, and FN are the number of true positives, false positives, true negatives, and false negatives respectively. The AP for a class is the area under the Precision-Recall (PR) curve, and is calculated by integrating precision as a function of recall:

$$AP = \int_0^1 P(R)\, dR \tag{8}$$

The mAP across all classes is then calculated as:

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{9}$$

where AP_i is the AP for the i-th class, and N is the total number of classes. In this paper, we calculate the mAP using IoU thresholds ranging from 0.5 to 0.95, with a step size of 0.05. Additionally, mAP_{50} refers to the mAP caculated at an IoU threshold 0.5.

- The bbox AP (Bounding Box Average Precision) measures the accuracy of the bounding boxes used in object detection tasks, reflecting the precision and recall of these boxes against ground truth.
- The mask AP (Mask Average Precision) evaluates the accuracy of predicted instance masks in instance segmentation tasks, assessing how precisely the model segments each object at the pixel level.
- The keypoint AP (Keypoint Average Precision) is used in human pose estimation, measuring the accuracy of the predicted keypoints (like elbows, knees, etc.) on the human body compared to the ground truth keypoints.

2) GradCAM

In computer vision, the class activation map (CAM) [59] is widely utilized to visualize the relationship between essential areas within an image and the classification results of a CNN. By using the CAM algorithm, one can better understand the decision-making process of the CNN, ultimately enhancing confidence in the classification results. This algorithm incorporates a Global Average Pooling (GAP) layer and visually demonstrates the influence of specific regions within an image on the classification outcome. The CAM can be expressed as follows:

$$\mathcal{L}_{\text{cam}}(c) = \sum_{k=1}(\omega_k^c \mathcal{F}_k) \tag{10}$$

where $\mathcal{F}_k$ is the k-th feature map of the final convolution layer, and the weights ω_k^c denotes the weights for class c on the with individual feature map $\mathcal{F}_k$. CAM is generated by computing a weighted sum of each feature map based on the output scores of a specific class.

The gradient class activation map (Grad-CAM) [60] is a direct extension of the original CAM implementation, eliminating the need to train a new set of weights. Grad-CAM utilizes gradient information from the final convolution layer of the CNN to understand the importance of each neuron for class prediction. If we denote the score for a specific class c with respect to the model output as y^c, then the gradient value of y^c for class c in the feature map $\mathcal{F}_k$ can be expressed as follows:

$$g(c, \mathcal{F}_k) = \frac{\partial y^c}{\partial \mathcal{F}_k} \tag{11}$$

This indicates how much the output score y^c has changed concerning the feature map $\mathcal{F}_k$, allowing us to assess the influence of the feature map on the specific class output. Subsequently, we calculate the importance weight ω_k^c for class c by averaging the gradients across each channel. Formally,

$$\omega_k^c = \frac{1}{Z}\sum_i\sum_j \frac{\partial y^c}{\partial \mathcal{F}_{k,(i,j)}} \tag{12}$$

where $\mathcal{F}_{k,(i,j)}$ is the value at the position (i, j) within the k-th feature map, and Z denotes the sum across all feature maps. Ultimately, Grad-CAM is eventually computed for all feature maps by taking gradients, performing weighted summation, and applying the ReLU operator is calculated by:

$$\mathcal{L}_{\text{Grad-CAM}} = \text{ReLU}(\sum_k \omega_k^c \mathcal{F}_k) \tag{13}$$

We use Grad-CAM to evaluate which pixels contribute to classifying the class person and vehicles in our experiment.

IV. RESULTS AND ANALYSIS

We evaluate the impact of the anonymized COCO training set on the training of YOLO and RT-DETR. As mentioned in Section III-B1, the anonymized dataset is generated by blurring, pixelating, distorting, and mask-out the person and vehicle classes included in the COCO training set. As mentioned in Section III-C, models trained with each anonymized data are evaluated using AP, with the model trained on the original COCO training set as the baseline. The impact of anonymization on model performance is measured by the change in performance of the model trained with anonymized data relative to the baseline, as defined in Section III-A.

Section IV-A investigates the relationship between the number of parameters in deep learning models and the anonymization of training data. We employ YOLOv8, which offers a range of model architectures at varying scales, to evaluate performance changes due to data anonymization in object detection, instance segmentation, and keypoint detection. Section IV-B analyzes the relationship between deep learning model architectures and the anonymization of training data. Here, we compare and assess the CNN-based YOLOv8 and the ViT-based RT-DETR at similar model scales. In Section IV-C, we evaluate the impact of anonymization on model performance for anonymized classes, such as a person, car, bus, motorcycle, and truck, while Section IV-D assesses negative impact for non-anonymized classes. In Section IV-C and Section IV-D, we use the YOLOv8 model architecture and compare the results in object detection and instance segmentation tasks.

A. PERFORMANCE COMPARISON BASED ON THE NUMBER OF MODEL PARAMETERS

In deep learning models, the number of parameters is a critical determinant of model complexity and capacity [61]. These parameters significantly impact not only the performance but also its generalization ability. Selecting an appropriate number of parameters when designing a model emerges as a

TABLE 3. Comparative analysis of model scale-specific mAP results for YOLOv8 models on the COCO validation set. YOLOv8 model is trained for object detection, instance segmentation, and keypoint detection tasks using both original and anonymized datasets. These models are trained utilizing the YOLOv8 architecture across five distinct scales. The baseline is the model trained on the original dataset, whereas models trained on anonymized data are denoted as follows: blurring (Blur), pixelation (Pixel), distortion (Disto), and mask-out (Mask). **Bold**: Largest decline in row. Underline: 2nd largest decline in row.

Object Detection										
Params(M)	bbox mAP					bbox mAP_{50}				
	Baseline	Blur	Pixel	Disto	Mask	Baseline	Blur	Pixel	Disto	Mask
3.2	33.1	-3.9	**-4.5**	**-4.5**	-4.4	41.7	-6.0	**-6.7**	-6.6	**-6.7**
11.2	41.4	-4.0	-4.1	**-4.9**	-4.1	57.4	-5.4	-5.6	**-6.7**	-5.5
25.9	46.4	-3.7	-3.5	**-4.6**	-4.4	62.7	-4.7	-4.6	**-5.8**	-5.4
43.7	48.3	-3.8	-3.6	**-3.9**	-3.3	64.6	-4.4	-4.3	**-4.9**	-4.3
68.2	48.4	-2.6	-2.4	**-3.1**	-2.5	65.0	-3.4	-3.4	**-4.0**	-3.4
Instance Segmentation										
Params(M)	mask mAP					mask mAP_{50}				
	Baseline	Blur	Pixel	Disto	Mask	Baseline	Blur	Pixel	Disto	Mask
3.4	27.6	-4.1	-3.8	**-4.4**	-4.1	44.5	-6.5	-6.1	**-6.9**	**-6.9**
11.8	34.4	-3.4	-3.5	**-4.4**	-3.8	54.6	-5.4	-5.7	**-6.7**	-6.0
27.3	38.6	-3.4	-3.0	**-3.7**	-3.1	60.0	-5.0	-5.0	**-5.7**	-4.9
46.0	41.1	-3.6	-3.5	**-3.7**	**-3.7**	63.2	-5.3	-5.0	-5.4	**-5.5**
71.8	41.5	**-4.1**	-3.2	-3.5	-2.8	63.7	**-6.0**	-4.4	-4.8	-4.1
Keypoint Detection										
Params(M)	keypoint mAP					keypoint mAP_{50}				
	Baseline	Blur	Pixel	Disto	Mask	Baseline	Blur	Pixel	Disto	Mask
3.3	42.8	-16.2	-16.7	-26.6	**-27.0**	41.7	-17.3	-17.6	-30.5	**-31.9**
11.6	51.7	-15.3	-15.3	**-26.4**	-23.1	57.4	-13.6	-13.4	**-24.9**	-21.0
26.4	57.7	-13.8	-14.3	**-25.6**	-23.5	62.7	-10.8	-10.7	**-19.9**	-17.9
44.4	60.9	-12.1	-12.8	**-24.2**	-22.1	64.6	-8.6	-8.5	**-16.8**	-14.8
69.4	62.8	-12.2	-12.7	**-24.8**	-22.1	65.0	-7.8	-7.7	**-16.3**	-13.2
99.1	62.8	-13.6	-13.5	**-24.6**	-23.2	65.0	-8.7	-7.9	**-16.0**	-15.3

crucial consideration, aiming to optimize performance while mitigating issues such as overfitting [62].

In this section, we investigate the influence of varying the number of parameters on the impact of anonymized data on model performance. This analysis illuminates the interplay among model capacity, parameter quantity, and the consequences of data anonymization, offering valuable insights for engineering applications.

1) QUANTITATIVE RESULTS

Table 3 shows the performance variation of models trained with anonymized data with respect to different numbers of model parameters. We observe a noticeable performance decline in the COCO validation set due to the data anonymization methods. In object detection and instance segmentation tasks, image distortion has the most detrimental impact on model performance. The impact of blurring, pixelation, and mask-out on model training are relatively similar. In the case of keypoint detection, only the person class is anonymized. The degradation in performance is more conspicuous due to image distortion and mask-out than blurring and pixelation. Keypoint detection is most susceptible to distortion, with the following highest vulnerability being mask-out. Distortion and mask-out processing primarily alter or remove visual information, resulting in the loss of color information associated with the person class. Keypoint detection relies on color, texture, contrast, and structural information to identify human body parts and joints [63], [64], [65]. Therefore, the loss of such information can have a detrimental impact on estimating the human pose.

Our experimental results show that models with more parameters exhibit increased resilience to the impact of data anonymization. This phenomenon may be associated with several factors. First, models with more parameters enhance their capacity for feature learning, allowing them to adapt more effectively to variations introduced by anonymization, such as image distortion and mask-out [66]. Secondly, larger models tend to reduce the risk of overfitting, which is particularly advantageous when handling anonymized data, as they excel at distinguishing informative patterns from noise [67]. Furthermore, models with higher expressiveness can better approximate hidden relationships within anonymized data to compensate for information loss due to anonymization. Lastly, larger models effectively utilize regularization techniques to prevent overfitting and enhance robustness against data variations introduced by anonymization [68]. These results emphasize the significant role of model scale in influencing performance when dealing with data anonymization, providing valuable insights for practical applications in domains that require sensitive data protection.

2) QUALITATIVE RESULTS

We compare several object detection from the models trained with the anonymized COCO dataset in Fig. 6. The results show that the proportion of false positive (FP) samples increases as the number of parameters decreases.

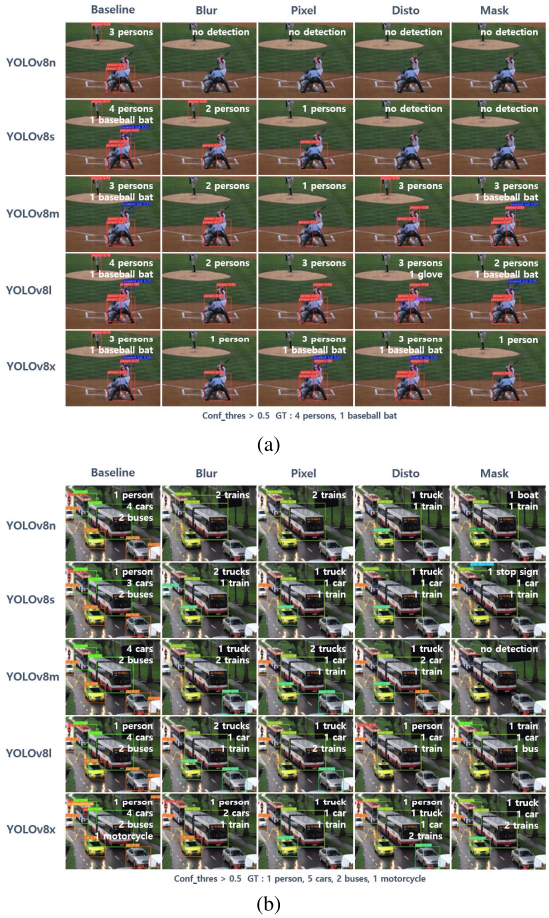


FIGURE 6. Inference results of YOLOv8 (detection task) with the ResNet50 backbone are as follows: (a) Person class. (b) Vehicle classes. The baseline is the model trained on the original dataset, whereas models trained on anonymized data are denoted as follows: blurring (Blur), pixelation (Pixel), distortion (Disto), and mask-out (Mask).

We observed a degradation in detection performance, specifically for the bus and truck classes within the vehicle category. The magnitude of performance degradation is inversely proportional to the model scale, consistent with quantitative results. Larger model scales result in the detection of objects that closely resemble the ground truth bounding box labels. In image recognition, the most significant threat arises from anonymization techniques involving distortion and mask-out.

3) VISUALIZATION RESULTS

We employ GradCAM to assess how the pixels contributing to the classification of person and vehicle classes vary about the number of model parameters. In Fig. 7, we observe the gradients of the last convolutional layer in the decoder head as part of our experiments. In models trained on the original images, activations predominantly focus on the overall shape and contour of the target object, suggesting a precise recognition of salient features of the object. For images processed with blur, the activation patterns largely resemble the original; however, a reduction in activation intensity is observed in certain granular regions, presumably

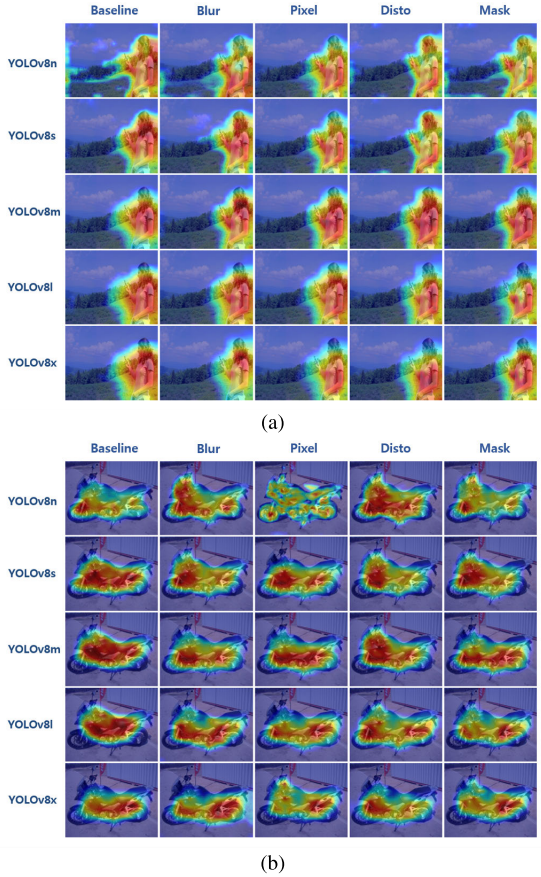


FIGURE 7. Evaluation of the impact of person class on non-anonymized class. Each point represents a class within the COCO dataset excluding the person class, differentiated by color according to the method of data anonymization. The graph includes a scatter plot with a regression line illustrating the relationship between the frequency of images co-occurring with the person class on the x-axis and the performance percentile relative to the baseline for each class on the y-axis. The solid line through the data points is the average trend. The bands represent confidence intervals, showing the range within which the regression line is expected to be found with a certain probability. The models for (a) object detection and (b) instance segmentation have been trained on anonymized COCO data utilizing the YOLOv8n, with evaluations conducted on the COCO validation set.

due to the loss of details caused by the blurring process. Models trained on data anonymized using mosaic, distortion, and mask techniques demonstrate distinct activation patterns compared to those trained on the original images. This deviation indicates that the model might perceive features differently or interpret objects in an altered manner. Such characteristics underscore the imperative for a comprehensive understanding of how anonymization techniques impact the interpretability and performance of deep learning models.

B. ANALYSIS OF DATA ANONYMIZATION IMPACT ACROSS MODEL ARCHITECTURES

In this section, we evaluate the impact of anonymization based on the architecture, representing the structure of deep learning models. Given the diversity of model architectures in deep learning methodologies, understanding the responsiveness of each architecture to anonymized data aids in

TABLE 4. Comparative evaluation of the impact of data anonymization on two neural network architectures: YOLOv8x (CNN-based) and RT-DETR-L (ViT-based). The evaluation utilizes the COCO validation set. Despite both architectures having a similar number of parameters (67-68M), YOLOv8x architecture demonstrates superior robustness in maintaining performance metrics, as opposed to the RT-DETR-L. The baseline is the model trained on the original dataset, whereas models trained on anonymized data are denoted as follows: blurring (Blur), pixelation (Pixel), distortion (Disto), and mask-out (Mask). **Bold**: Largest decline in row. Underline: 2nd largest decline in row.

architec.	bbox mAP					bbox mAP_{50}				
	Baseline	Blur	Pixel	Disto	Mask	Baseline	Blur	Pixel	Disto	Mask
YOLOv8x	48.4	-2.6	-2.4	**-3.1**	-2.5	65.0	-3.4	-3.4	**-4.0**	-3.4
RT-DETR-L (ViT)	48.5	**-7.5**	-4.6	-5.8	-5.4	61.7	**-7.7**	-5.6	-7.0	-6.3

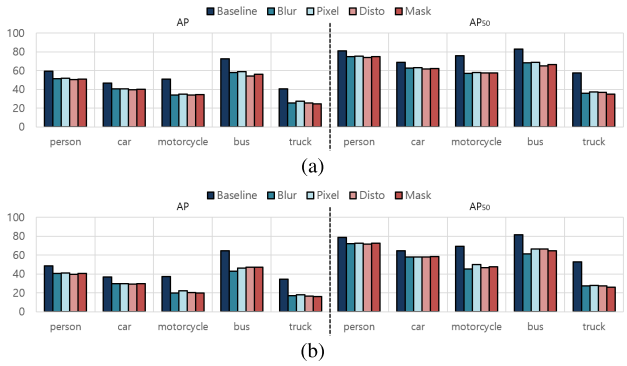


FIGURE 8. Results of the performance degradation of anonymized classes using the YOLOv8m trained on original and anonymized data, evaluated on the COCO validation set: (a) Object detection. (b) Instance segmentation. The baseline is the model trained on the original dataset, whereas models trained on anonymized data are denoted as follows: blurring (Blur), pixelation (Pixel), distortion (Disto), and mask-out (Mask).

determining which models demonstrate greater resilience or susceptibility in the face of data anonymization. This provides insights into selecting the most suitable model for real-world applications. For these reasons, we evaluate the interaction between data anonymization and model architectures from the experimental results. We compare YOLO based on CNN and DETR based on the ViT, trained with a similar model scale, to assess the impact of anonymization training data depending on the model architecture.

According to previous studies, ViT-based models are generally recognized to exhibit greater resilience than CNN-based models against perturbation attacks targeting the entire image [69]. However, our results show that despite the same model scale, the real-time detection transformer (RT-DETR) exhibited a more significant performance degradation than YOLO. Based on the data presented in Table. 4, the observed results can be attributed to several underlying factors. RT-DETR relies on its inherent self-attention mechanism for learning object relationships, which may pose challenges when certain classes are anonymized in the training data, rendering objects less distinguishable [70]. Consequently, RT-DETR may struggle to capture the appropriate inter-object relationships under such circumstances. RT-DETR and YOLO employ distinct object detection architectures, with RT-DETR relying heavily on self-attention for predicting object positions, while YOLO adopts a grid-based approach [51], [71]. This architectural contrast accentuates the reliance of RT-DETR on learning object relationships, particularly when faced with anonymized classes in the training data. Furthermore, ViT architecture processes visual data differently than CNN architecture, employing input tokenization [48]. This distinction could lead to varying representations of anonymized data, potentially contributing to RT-DETR exhibiting a more significant performance decline.

In conclusion, the performance degradation in RT-DETR stems from its inherent mechanism for learning object relationships and its reliance on the prediction of object positions, both of which can be affected by data anonymization. Additionally, architectural differences and input representation methods between models may also play a role in influencing the observed performance decline.

C. EVALUATION OF THE IMPACT ON ANONYMIZED CLASSES

In this section, we compare the performance of models trained with anonymized data with respect to the anonymized classes. We select person, car, motorcycle, truck, and bus classes in the COCO dataset as target classes for anonymization. Fig. 8 presents experimental results on the relative performance of anonymized target classes using traditional anonymization methods. We observe performance degradation in object detection and instance segmentation due to anonymization processing. In both tasks, the recognition performance significantly drops for the motorcycle, bus, and truck classes. However, the performance change for the person class is relatively minor. We believe this is due to the class imbalance present in the COCO dataset [72]. In the COCO training dataset, the person class accounts for 58.7% of the total, while the vehicle category classes represent only 4.5%. Such class imbalance contributes to differences in generalization performance.

The performance degradation was similar across anonymization methods. Nevertheless, distortion and mask-out, despite completely obscuring image regions, had a more pronounced impact on model performance compared to other anonymization methods. We attribute this phenomenon to the substantial information loss in areas subjected to distortion and mask-out. While blurring and pixelation tend to preserve some information while safeguarding personal data, these methods involve deleting specific image regions, restricting the model's ability to learn valuable visual

TABLE 5. Comparative analysis of AP for object detection and instance segmentation on non-anonymized classes. The COCO validation set is utilized for evaluation. The model under evaluation, YOLOv8m, is trained using both original and anonymized datasets. The baseline is the model trained on the original dataset, whereas models trained on anonymized data are denoted as follows: blurring (Blur), pixelation (Pixel), distortion (Disto), and mask-out (Mask). **Bold**: Largest decline in row. Underline: 2nd largest decline in row.

Group	Class	Object Detection bbox AP					bbox AP_{50}				
		Baseline	Blur	Pixel.	Disto	Mask	Baseline	Blur	Pixel	Disto	Mask
Riding	bicycle	35.5	-5.6	-6.7	-7.0	**-7.9**	58.4	-10.1	-12.3	-12.6	**-13.2**
	motorcycle	76.1	**-19.2**	-16.3	-18.5	-18.6	51.1	-16.9	-16.3	**-17.0**	-16.7
Equipment	skis	51.5	**-11.3**	-9.9	-10.7	-9.1	28.9	-5.4	-5.2	**-5.7**	-4.1
	snowboard	51.4	-15.8	-10.7	-15.3	**-17.4**	38.7	-12.0	-7.9	-12.2	**-13.4**
	baseball bat	58.9	-19.5	-19.7	-21.0	**-22.1**	38.1	-18.2	-17.5	-17.8	**-18.8**
	baseball glove	65.5	-23.3	-25.6	-25.9	**-27.6**	42.0	-15.5	-17.2	-16.8	**-18.6**
	skateboard	65.5	-23.3	-25.6	-25.9	**-27.6**	42.0	-15.5	-17.2	-16.8	**-18.6**
	surfboard	81.0	-18.0	-15.0	-17.0	**-18.3**	58.5	**-15.2**	-12.2	-13.0	-14.3
	tennis racket	61.7	**-6.2**	-5.5	-4.0	-5.0	41.6	-4.4	-4.7	-4.3	**-5.1**
Accessory	backpack	29.9	-10.9	-10.3	**-12.1**	-10.7	17.2	**-6.8**	-6.4	-6.7	-6.6
	umbrella	65.3	-3.2	**-3.9**	-3.1	-3.6	45.5	-3.4	-2.3	-3.0	**-3.9**
	handbag	31.5	-12.1	-13.3	**-14.2**	-13.9	18.5	-8.1	-8.3	-8.6	**-8.9**
	tie	55.0	-17.1	-17.0	**-21.1**	-20.4	36.4	-13.3	-12.1	-15.8	**-16.8**
	suitcase	64.1	-4.0	**-5.8**	-4.4	-1.8	44.6	-3.1	**-5.0**	-3.4	-1.9
Group	**Class**	**Instance Segmentation mask AP**					**mask AP_{50}**				
		Baseline	**Blur**	**Pixel.**	**Disto**	**Mask**	**Baseline**	**Blur**	**Pixel**	**Disto**	**Mask**
Riding	bicycle	48.6	-11.0	-10.0	-10.8	**-11.6**	20.8	-4.7	-4.5	-4.4	**-5.5**
	motorcycle	69.2	**-23.8**	-19.4	-22.6	-21.7	37.5	**-17.7**	-15.2	-17.0	-17.4
Equipment	skis	36.4	-9.2	**-11.2**	-10.6	-10.1	9.5	**-2.7**	-2.6	-2.2	**-2.7**
	snowboard	52.6	-18.5	-19.8	**-23.4**	-14.4	29.3	-9.9	-10.3	**-12.8**	-6.9
	baseball bat	52.3	-11.4	**-16.7**	-11.2	-15.8	26.5	-8.7	**-10.9**	-7.2	-9.9
	baseball glove	64.6	-24.7	-25.1	-24.0	**-26.1**	37.9	-13.5	-13.2	-13.7	**-15.0**
	skateboard	77.1	**-22.2**	-19.9	-19.8	-19.4	41.5	-13.6	-13.2	**-14.1**	-13.6
	surfboard	64.2	**-9.9**	-6.5	-8.9	-9.8	34.7	**-4.9**	-3.2	-4.8	-4.4
	tennis racket	81.8	-14.4	-12.2	**-15.1**	-14.9	53.7	-8.9	-9.0	-10.5	**-11.5**
Accessory	backpack	32.1	**-14.0**	-13.5	-13.7	-13.6	16.5	**-7.4**	-6.7	-6.7	-5.9
	umbrella	67.9	-2.9	-3.0	**-3.5**	-2.6	46.8	-2.7	-1.4	**-3.2**	-2.7
	handbag	33.0	-14.5	-15.7	**-16.9**	-13.2	16.6	-7.9	-8.1	**-8.2**	-7.3
	tie	55.4	-15.9	-12.8	**-17.0**	-15.2	33.5	-11.6	-9.8	**-12.5**	-11.7
	suitcase	61.7	-2.0	-5.2	**-5.3**	-1.7	41.8	-2.1	-4.3	**-5.1**	-2.2

features. Consequently, the model needs more substitute information or cues for the removed regions, leading it to disregard these areas during training and focus on other parts of the image. Our results highlight the need to carefully consider how anonymization methods impact information loss when selecting them to optimize model performance. Thus, it is imperative to exercise discretion when selecting anonymization methods to strike a delicate equilibrium between personal data safeguarding and specific class performance requirements.

D. EVALUATION OF THE IMPACT ON NON-ANONYMIZED CLASSES

We evaluate the impact on non-anonymized classes, which hold equal significance in practical applications. For instance, within the context of autonomous vehicles, not only the precise recognition of vehicles and pedestrians is crucial, but also the identification of roads, signals, signs, and various dynamic objects is vital for ensuring safety [73]. Therefore, we analyze non-anonymized classes to garner more profound insights into the interactions between anonymized and non-anonymized classes.

Our experiment results present a detrimental effect of anonymized classes on the detection performance of non-anonymized classes. Table. 5 shows the performance of riding, equipment, and accessory classes. Significant performance degradation is observed in all classes within the three groups. Object detection shows that classes belonging to the riding and equipment categories manifest susceptibility to mask-out. Conversely, while influenced somewhat, classes within the accessory group do not demonstrate pronounced vulnerabilities to particular anonymization methodologies. This pattern of observations is consistently reflected in instance segmentation as well. Across both tasks, the classes related to equipment or wear exhibited more substantial performance drops. We believe this is due to the significant overlap between these classes and the anonymized classes, leading to performance challenges. Our study underscores the importance of a thoughtful approach to anonymization, emphasizing the need to balance individual class performance considerations for achieving the equilibrium between personal data protection and model performance. For example, in industrial safety, detecting whether a worker is appropriately wearing safety equipment, such as helmets,

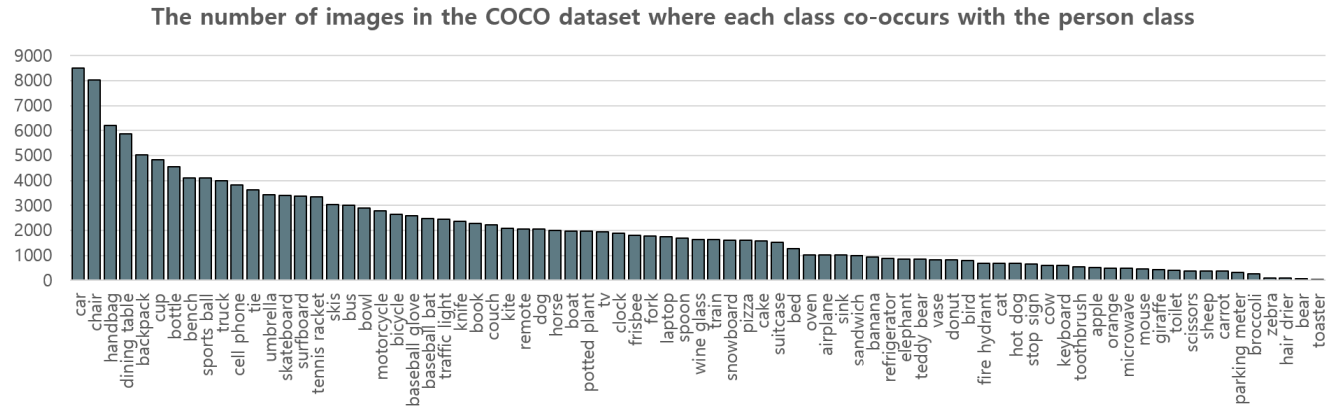
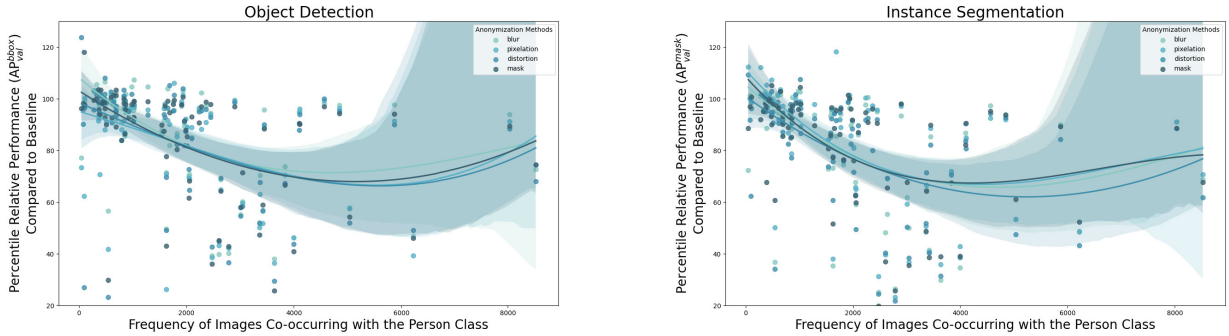


FIGURE 9. Frequency of Co-occurrence of Person Class with Other Classes. x-axis: COCO classes, y-axis: number of images with co-occuring person class.



(a) (b)

FIGURE 10. Evaluation of the impact of person class on non-anonymized class. Each point represents a class within the COCO dataset excluding the person class, differentiated by color according to the method of data anonymization. The graph includes a scatter plot with a regression line illustrating the relationship between the frequency of images co-occurring with the person class on the x-axis and the performance percentile relative to the baseline for each class on the y-axis. The solid line through the data points is the average trend. The bands represent confidence intervals, showing the range within which the regression line is expected to be found with a certain probability. The models for (a) object detection and (b) instance segmentation have been trained on anonymized COCO data utilizing the YOLOv8n, with evaluations conducted on the COCO validation set.

safety belts, and protective glasses, is crucial. In this context, face anonymization may obscure specific parts of the equipment or attire. Therefore, selecting an appropriate anonymization method is imperative while considering the importance of the utilized classes for the specific task.

We observed that the anonymized person class influences the recognition performance of non-anonymized classes. Particularly within the COCO dataset, we have observed that classes frequently co-occurring with the person class are more significantly impacted. Fig. 9 shows the frequency of images co-occurs with the person class by class.

Fig. 10 shows the influence of the co-occurrence frequency of the person class on the object detection performance of non-anonymized classes. Generally, it is observed that the detection accuracy of non-anonymized classes declines as the co-occurrence frequency with the person class increases. Our experimental result suggests that even without direct anonymization of objects in the training data, non-anonymized classes can be adversely affected by the presence of anonymized classes. Anonymized classes may obscure or act as noise, affecting the visibility and recognition of non-anonymized classes within images. This side effect could lead to difficulty detecting these objects due to potential occlusion or misidentification.

V. CONCLUSION

In this paper, we thoroughly analyze the effects of different anonymization techniques on the performance of deep learning-based models in computer vision tasks, with a particular emphasis on presenting a model-centric perspective such as the type of deep learning model, and the number of parameters. For this, we investigate the impact of anonymized data for model training, focusing on large-scale object datasets.

Our quantitative experimental results show that anonymizing the human and vehicle categories influences the performance of object recognition models. We report that the degree of influence varies by class depending on the proportion of

samples included in the data and the anonymization technique applied. We observed that as the number of model parameters decreases, it leads to more significant performance degradation due to data anonymization. However, the impact of data anonymization was negligible for classes unrelated to the person class or other classes with infrequent frequency of frequencies within the data. Also we observed that the most influential anonymization method varies depending on the computer vision task, model architecture, and classes. Based on the observed results, the following conclusions are drawn:

- We have identified the need for a more robust level of data anonymization beyond faces and license plates. We conduct pixel-level anonymization, and assess the impact of anonymization on model training.
- The larger the scale of the deep learning model, the more robust it becomes in handling data anonymization.
- The ViT incorporating self-attention is relatively less robust to data anonymization compared to CNNs.
- The influence of anonymization was noticeable even in classes closely related to anonymized objects. Specifically, it is observed that recognition performance decreases as the frequency of concurrent occurrences with anonymized classes increases within the image.

REFERENCES

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[2] A. Bailly, C. Blanc, É. Francis, T. Guillotin, F. Jamal, B. Wakim, and P. Roy, "Effects of dataset size and interactions on the prediction performance of logistic regression and deep learning models," *Comput. Methods Programs Biomed.*, vol. 213, Jan. 2022, Art. no. 106504.

[3] A. Althnian, D. AlSaeed, H. Al-Baity, A. Samha, A. B. Dris, N. Alzakari, A. A. Elwafa, and H. Kurdi, "Impact of dataset size on classification performance: An empirical evaluation in the medical domain," *Appl. Sci.*, vol. 11, no. 2, p. 796, Jan. 2021.

[4] European Union. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27, Apr. 2016 on the Protection of Natural Persons With Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing, Directive 95/46/EC (General Data Protection Regulation)*. [Online]. Available: http://data.europa.eu/eli/reg/2016/679/oj

[5] California Legislature. (2020). *California Consumer Privacy Act of 2018 (as Amended by the California Privacy Rights Act of 2020)*. [Online]. Available: https://www.oag.ca.gov/privacy/ccpa

[6] R. A. Jarvis, "A perspective on range finding techniques for computer vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-5, no. 2, pp. 122–139, Mar. 1983.

[7] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "NuScenes: A multimodal dataset for autonomous driving," 2019, *arXiv:1903.11027*.

[8] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, and V. Vasudevan, "Scalability in perception for autonomous driving: Waymo open dataset," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2443–2451.

[9] L. Schnabel, S. Matzka, M. Stellmacher, M. Pätzold, and E. Matthes, "Impact of anonymization on vehicle detector performance," in *Proc. 2nd Int. Conf. Artif. Intell. Industries (AI4I)*, Sep. 2019, pp. 30–34.

[10] S. R. Klomp, M. Van Rijn, R. G. J. Wijnhoven, C. G. M. Snoek, and P. H. N. De With, "Safe fakes: Evaluating face anonymizers for face detectors," in *Proc. 16th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Dec. 2021, pp. 1–8.

[11] K. Yang, J. H. Yau, L. Fei-Fei, J. Deng, and O. Russakovsky, "A study of face obfuscation in ImageNet," in *Proc. Conf. Mach. Learn. (ICML)*, 2022, pp. 25313–25330.

[12] M. Tomei, L. Baraldi, S. Bronzin, and R. Cucchiara, "Estimating (and fixing) the effect of face obfuscation in video recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2021, pp. 3257–3263.

[13] A. Cecaj, M. Mamei, and F. Zambonelli, "Re-identification and information fusion between anonymized CDR and social network data," *J. Ambient Intell. Humanized Comput.*, vol. 7, no. 1, pp. 83–96, Feb. 2016.

[14] B. Lubarsky. (2020). *Re-Identification of 'Anonymized' Data. Georgetown Law Technology Review*. [Online]. Available: https://georgetownlawtechreview.org/re-identification-of-anonymized-data/GLTR-04-2017/

[15] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollr, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 740–755.

[16] J. He, B. Liu, D. Kong, X. Bao, N. Wang, H. Jin, and G. Kesidis, "PUPPIES: Transformation-supported personalized privacy preserving partial image sharing," in *Proc. 46th Annu. IEEE/IFIP Int. Conf. Dependable Syst. Netw. (DSN)*, Jun. 2016, pp. 359–370.

[17] E. M. Newton, L. Sweeney, and B. Malin, "Preserving privacy by de-identifying face images," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 2, pp. 232–243, Feb. 2005.

[18] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proc. 7th Int. Conf. Document Anal. Recognit. (ICDAR)*, vol. 3, 2003, pp. 1–6.

[19] H. Hukkelas and F. Lindseth, "Does image anonymization impact computer vision training?" in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2023, pp. 140–150.

[20] H. d. O. Silva, T. Basso, and R. L. d. O. Moraes, "Privacy and data mining: Evaluating the impact of data anonymization on classification algorithms," in *Proc. 13th Eur. Dependable Comput. Conf. (EDCC)*, Sep. 2017, pp. 111–116.

[21] P. Dvo ek and P. Hurtik, "What is the cost of privacy?" in *Proc. CCIS*, vol. 1602, 2022, pp. 696–706.

[22] J. Zhou and J. Beyerer, "Impacts of data anonymization on semantic segmentation," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2022, pp. 997–1004.

[23] H. Hukkelås, M. Smebye, R. Mester, and F. Lindseth, "Realistic full-body anonymization with surface-guided GANs," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 1430–1440.

[24] H. Hukkelås and F. Lindseth, "DeepPrivacy2: Towards realistic full-body anonymization," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 1329–1338.

[25] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[26] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Deepfakes and beyond: A survey of face manipulation and fake detection," *Inf. Fusion*, vol. 64, pp. 131–148, Dec. 2020.

[27] L. Li, J. Bao, H. Yang, D. Chen, and F. Wen, "FaceShifter: Towards high fidelity and occlusion aware face swapping," 2019, *arXiv:1912.13457*.

[28] E. Richardson, Y. Alaluf, O. Patashnik, Y. Nitzan, Y. Azar, S. Shapiro, and D. Cohen-Or, "Encoding in style: A styleGAN encoder for image-to-image translation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2287–2296.

[29] Y. Shen, J. Gu, X. Tang, and B. Zhou, "Interpreting the latent space of GANs for semantic face editing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9240–9249.

[30] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality stability and variation," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–12.

[31] H. Proença, "The UU-Net: Reversible face de-identification for visual surveillance video footage," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 2, pp. 496–509, Feb. 2022.

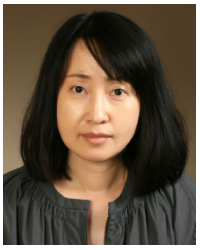
[32] J. Yu, H. Xue, B. Liu, Y. Wang, S. Zhu, and M. Ding, "GAN-based differential private image privacy protection framework for the Internet of Multimedia Things," *Sensors*, vol. 21, no. 1, p. 58, Dec. 2020.

[33] L. Du and H. Ling, "Preservative license plate de-identification for privacy protection," in *Proc. Int. Conf. Document Anal. Recognit.*, Sep. 2011, pp. 468–472.

[34] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3213–3223.
[35] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "BDD100K: A diverse driving dataset for heterogeneous multitask learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2633–2642.
[36] M. S. Nixon, J. N. Carter, J. M. Nash, P. S. Huang, D. Cunado, and S. V. Stevenage, "Automatic gait recognition," *Biometrics*, vol. 7, no. 2, p. 3, 1999.
[37] J.-H. Yoo, D. Huang, and M. K. Nixon, "Gender classification in human gait using support vector machine," in *Proc. Int. Conf. Adv. Concepts Intell. Vis. Syst.* Antwerp, Belgium: Springer, 2005, pp. 138–145.
[38] L. Lee and W. E. L. Grimson, "Gait analysis for recognition and classification," in *Proc. 5th IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2002, pp. 155–162.
[39] M. Ivasic-Kos, A. Iosifidis, A. Tefas, and I. Pitas, "Person de-identification in activity videos," in *Proc. 37th Int. Conv. Inf. Commun. Technol., Electron. Microelectron.*, May 2014, pp. 1294–1299.
[40] R. T. Collins, R. Gross, and J. Shi, "Silhouette-based human identification from body shape and gait," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2002, pp. 366–371.
[41] Y. Ren, G. Li, S. Liu, and T. H. Li, "Deep spatial transformation for pose-guided person image generation and animation," *IEEE Trans. Image Process.*, vol. 29, pp. 8622–8635, 2020.
[42] A. Cui, D. McKee, and S. Lazebnik, "Dressing in order: Recurrent person image generation for pose transfer, virtual try-on and outfit editing," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 14618–14627.
[43] R. Peter, A.-K. Grosselfinger, D. Münch, and M. Arens, "Automated license plate detection for image anonymization," *Proc. SPIE*, vol. 11166, pp. 218–231, Oct. 2019.
[44] L. Trinh, P. Pham, H. Trinh, N. Bach, D. Nguyen, G. Nguyen, and H. Nguyen, "PP4AV: A benchmarking dataset for privacy-preserving autonomous driving," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 1206–1215.
[45] X. Liu, W. Liu, T. Mei, and M. Huadong, "A deep learning-based approach to progressive vehicle re-identification for urban surveillance," in *Proc. ECCV*, vol. 1, 2016, pp. 869–884.
[46] C. Lyu, W. Zhang, H. Huang, Y. Zhou, Y. Wang, Y. Liu, S. Zhang, and K. Chen, "RTMDet: An empirical study of designing real-time object detectors," 2022, *arXiv:2212.07784*.
[47] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," 2015, *arXiv:1511.08458*.
[48] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
[49] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.
[50] Ultralytics. (2023). *YOLOv8* Accessed: Nov. 1, 2023. [Online]. Available: https://github.com/ultralytics/ultralytics
[51] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 213–229.
[52] W. Lv, Y. Zhao, S. Xu, J. Wei, G. Wang, C. Cui, Y. Du, Q. Dang, and Y. Liu, "DETRs beat YOLOs on real-time object detection," 2023, *arXiv:2304.08069*.
[53] S. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, pp. 1345–1359, Nov. 2010.
[54] X. Han, Z. Zhang, N. Ding, Y. Gu, X. Liu, and Y. Huo, "Pre-trained models: Past present and future," *AI Open*, vol. 2, pp. 225–250, Jan. 2021.
[55] G. Vrbancic and V. Podgorelec, "Transfer learning with adaptive fine-tuning," *IEEE Access*, vol. 8, pp. 196197–196211, 2020.
[56] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 580–587.
[57] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Jun. 2010.
[58] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *Proc. CVPR*, Jun. 2011, pp. 1297–1304.
[59] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2921–2929.
[60] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.
[61] X. Hu, L. Chu, J. Pei, W. Liu, and J. Bian, "Model complexity of deep learning: A survey," *Knowl. Inf. Syst.*, vol. 63, no. 10, pp. 2585–2619, Oct. 2021.
[62] C. Zhang, O. Vinyals, R. Munos, and S. Bengio, "A study on overfitting in deep reinforcement learning," 2018, *arXiv:1804.06893*.
[63] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 172–186, Jan. 2021.
[64] N. Ahmad, J. Khan, J. Y. Kim, and Y. Lee, "Joint human pose estimation and instance segmentation with PosePlusSeg," in *Proc. AAAI Conf. Artif. Intell.*, Jun. 2022, vol. 36, no. 1, pp. 69–76.
[65] W. Liu, Q. Bao, Y. Sun, and T. Mei, "Recent advances of monocular 2D and 3D human pose estimation: A deep learning perspective," *ACM Comput. Surv.*, vol. 55, no. 4, pp. 1–41, Apr. 2023.
[66] Y. L. Sung, V. Nair, and C. A. Raffel, "Training neural networks with fixed sparse masks," in *Proc. Neural Inf. Process. Syst. (NIPS)*, 2021, pp. 24193–24205.
[67] S. Tufail, H. Riggs, M. Tariq, and A. I. Sarwat, "Advancements and challenges in machine learning: A comprehensive review of models, libraries, applications, and algorithms," *Electronics*, vol. 12, no. 8, p. 1789, Apr. 2023.
[68] J. Brownlee, "How to avoid overfitting in deep learning neural networks," *Mach. Learn. Mastery*, vol. 17, p. 12, Dec. 2018.
[69] Y. Bai, J. Mei, A. L. Yuille, and C. Xie, "Are transformers more robust than CNNs?" in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2021, pp. 26831–26843.
[70] M.-H. Guo, T.-X. Xu, J.-J. Liu, Z.-N. Liu, P.-T. Jiang, T.-J. Mu, S.-H. Zhang, R. R. Martin, M.-M. Cheng, and S.-M. Hu, "Attention mechanisms in computer vision: A survey," 2021, *arXiv:2111.07624*.
[71] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[72] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2999–3007.
[73] A. Balasubramaniam and S. Pasricha, "Object detection in autonomous vehicles: Status and open challenges," 2022, *arXiv:2201.07706*.



JUN HA LEE received the M.S. degree in computer science from Chung-Ang University, Seoul, South Korea, in 2020. In 2020, he joined the AI and Robotics Research and Development Division, Korea Institute of Industrial Technology, Ansan-si, South Korea, where he is currently a Researcher. His research interests include computer vision and machine learning.



SU JEONG YOU received the B.S. and M.S. degrees in control and instrumentation engineering and the Ph.D. degree in electrical engineering and computer sciences from Seoul National University, Seoul, South Korea, in 1993, 1995, and 2013, respectively. She was a Research Engineer with LG, Seoul, for four years, with a focus on projects with machine learning and artificial neural networks. She is currently a Principal Researcher with the AI and Robot Research and Development Department, Korea Institute of Industrial Technology, Ansan-si, South Korea. Her research interests include computer vision and machine learning.