brighter AI

Whitepaper

The Impact of Deep Natural Anonymization on the Training of Machine Learning Models

Abstract

AI innovation and machine learning need high quality image and video data. **As data protection regulations are tightening, companies anonymize data to comply with the regulations.** However, traditional anonymization techniques such as pixelation and black bars cannot preserve the accuracy and integrity of the original data.

brighter AI's Deep Natural Anonymization (DNAT) is a generative AI based solution that preserves high data quality for analytics and machine learning.

This whitepaper aims to test machine learning model's accuracy when it is trained on image data anonymized by DNAT.

To measure the model's accuracy, we compared the Mask R-CNN model's performance in the segmentation task between two cases. One case is when the model is trained with unmodified data, the other case is when it is trained with anonymized data. The result is measured with Average Precision (AP) and Mean Average Precision (mAP).

After analyzing the results, we conclude the machine learning model trained on anonymized data **has the same accuracy as when it is trained on unmodified data.**

I. INTRODUCTION

Data protection regulations are tightening, especially those regarding publicly-recorded images and videos. Though the regulations protect privacy as a fundamental human right, they block exciting AI and machine learning use cases. Anonymization is an effective method to comply with privacy regulations. However, the quality of data anonymized by traditional solutions **is not high enough for analytics and machine learning.**

Therefore, brighter AI developed Deep Natural Anonymization (DNAT) which is designed to protect personally identifiable information (PII) in image and video data. This whitepaper aims to evaluate whether a machine learning model's accuracy is impacted if it is trained on images anonymized by DNAT compared to when it is trained on unmodified data.

We trained the model with both unmodified and anonymized data while keeping the same hyperparameters. In this way, we ensured that the two datasets were the only reason to cause differences in accuracy.

II. DATASET

The dataset we used is Cityscapes[1] with fine labels. Cityscapes is a standardized publicly available dataset that contains images of street scenes recorded from various locations, in different weather conditions, and covers different dates and times. Figure 1 shows a sample image from the Cityscapes dataset. As DNAT is designed for human faces and license plates, we focused on object classes such as bus, truck, car, person, and rider. We used brighter's AI DNAT to create a training dataset with anonymized images, where the PII in every image is anonymized.

Figure 2 shows a sample image where the PII was anonymized by DNAT. In this example, DNAT anonymized the face by generating a synthetic overlay, it anonymized the license plate by changing its alphanumeric combination.

FIG. 1. Sample image from the Cityscapes dataset with blurred PII

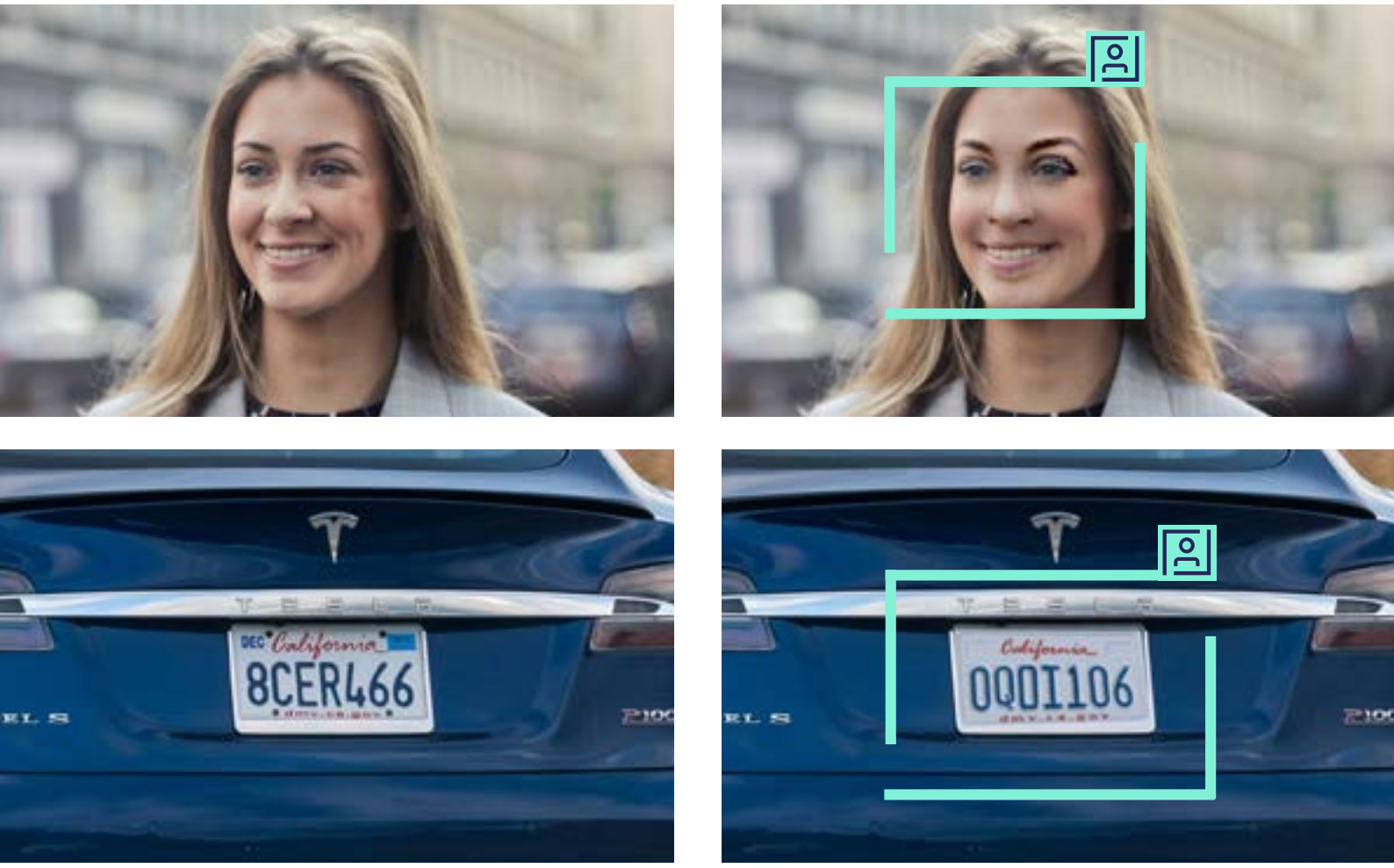
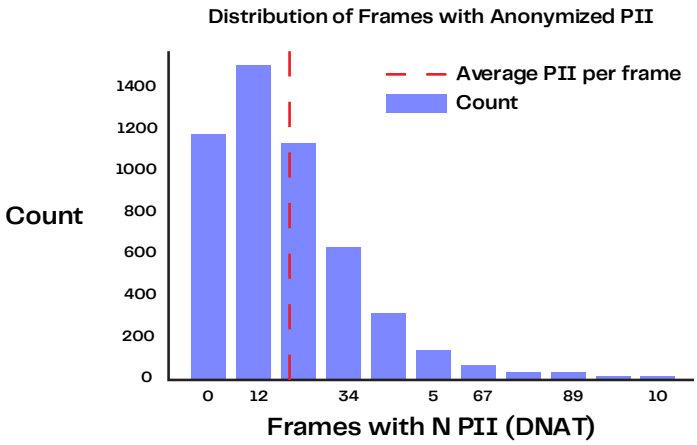


FIG. 2. Image data with face and license plate before and after DNAT (the original image is not included in the dataset).

To understand how the dataset is characterized in terms of PII, we analyzed the distribution of the mean occurrence of PII per frame in the set. Figure 3 shows this distribution, and the set is characterized by having two anonymized PII per frame on average.

FIG. 3. Number of frames of the training data with N anonymized PII. The training dataset is characterized by having two anonymized PII per frame on average.



III. EXPERIMENT

The experiment compared the model's performance in the segmentation task between two cases. One case is when the model is trained with unmodified data, the other is trained with anonymized data. Note that the anonymized data corresponds to an anonymized version of the training split shown in Table I.

TABLE I. Cityscapes Dataset Distribution.

Training	2975
Validation	500
Test	1525
Total Images	5000

We chose Mask R-CNN as the model for the experiment because it is a well-known community standard. Its implementation is publicly available via the Detectron2 [2] repository hosted by Facebook AI Research. We used the aforementioned repository to conduct our experiment.

Detectron2 comes with pre-configured configurations. Our training and testing used the configuration named "cityscapes/mask rcnn R 50 FPN.yaml". This configuration utilizes a pre-trained ResNet50 backbone (pre-trained in ImageNet dataset) and uses eight GPUs for training. Based on the available resources, we changed this default configuration to four GPUs and adjusted the learning rate accordingly to 0.005.

We used the Average Precision (AP) [3] metric to evaluate the model's performance in the segmentation task. To incorporate the model's performance in different segmentation classes, we calculated the mean AP (mAP) [3] across these classes, while the open-source repository of Detectron2 [2] already provides the AP as an evaluation metric.

In order to ensure the consistency of the experiment and to minimize its variance, we trained the model five times and aggregated these results to calculate the AP.

Table II shows the testing results for training the Mask R-CNN machine learning model on the public ImageNet dataset, using the ResNet-50 backbone that is pre-trained on the public ImageNet dataset. The table shows the AP of the two cases: when the model is trained on unmodified data (left) and when it is trained on anonymized data (right).

The results do not show an obvious difference. The mAP of all eight object classes is 0.315, regardless of the class. AP of individual classes does not show significant differences between the two cases either.

	AP	
Class	**Unmodified Data**	**Anonymized Data**
Person	0.309 ± 0.002	0.310 ± 0.002
Rider	0.246 ± 0.004	0.243 ± 0.005
Car	0.501 ± 0.003	0.505 ± 0.002
Truck	0.278 ± 0.006	0.289 ± 0.015
Bus	0.503 ± 0.015	0.501 ± 0.013
Train	0.329 ± 0.018	0.314 ± 0.021
Motorcycle	0.168 ± 0.005	0.165 ± 0.010
Bicycle	0.187 ± 0.002	0.189 ± 0.005
All classes (mAP)	0.315 ± 0.004	0.315 ± 0.004

TABLE II. Results of the model trained with original and anonymized data. The first eight rows show the mean of the average precision and its standard deviation, calculated over five training runs. The last row shows the mean average precision, where the mean is taken across all eight classes.

V. CONCLUSION

This whitepaper proved that image data trained by brighter AI's DNAT does not impact the accuracy of a machine learning model. We ran the training and test on a state-of-the-art machine learning model Mash R-CNN on the public Cityscapes dataset. The accuracy of the results is measured by AP and mAP (see Table II). Through our experiments, we conclude that machine learning models trained on anonymized data have the same accuracy level as those trained on unmodified data.

Data privacy is an increasingly important social topic. For businesses and organizations who wish to train machine learning models with publicly collected data but are bound by data protection legislations [4] (e.g., the GDPR), this whitepaper provides another option: use data anonymized by brighter AI's DNAT to train machine learning models. **DNAT protects individuals' privacy, keeps high accuracy for machine learning model training and enables you to run innovative projects.**

[1] The Cityscapes Dataset, https://www.cityscapes-dataset.com/

[2] Francisco Massa and Ross Girshick. maskrcnn-benchmark: Fast, modular reference implementation of Instance Segmentation and Object Detection algorithms in PyTorch. 2018, https://github.com/facebookresearch/detectron2

[3] Evaluation Metrics: Mean Average Precision, https://en.wikipedia.org/wiki/Evaluation_measures_(information_retrieval)#Mean%20average%20precision

[4] brighter AI & German AI Association. Privacy regulations on video data collection worldwide. 2020, https://brighter.ai/resources/global-ai-projects-local-privacy-laws-embrace-privacy-by-design-worldwide/

Brighter AI Technologies GmbH, Registry Court: Berlin Charlottenburg hrb 188087, Managing directors: Marian Gläser and Tobias Stelzer; hello@brighter.ai