DartBlur: Privacy Preservation with Detection Artifact Suppression

Baowei Jiang*, Bing Bai*, Haozhe Lin*, Yu Wang,
Yuchen Guo, Lu Fang

Tsinghua University

*authors contributed equally

Summary of highlights

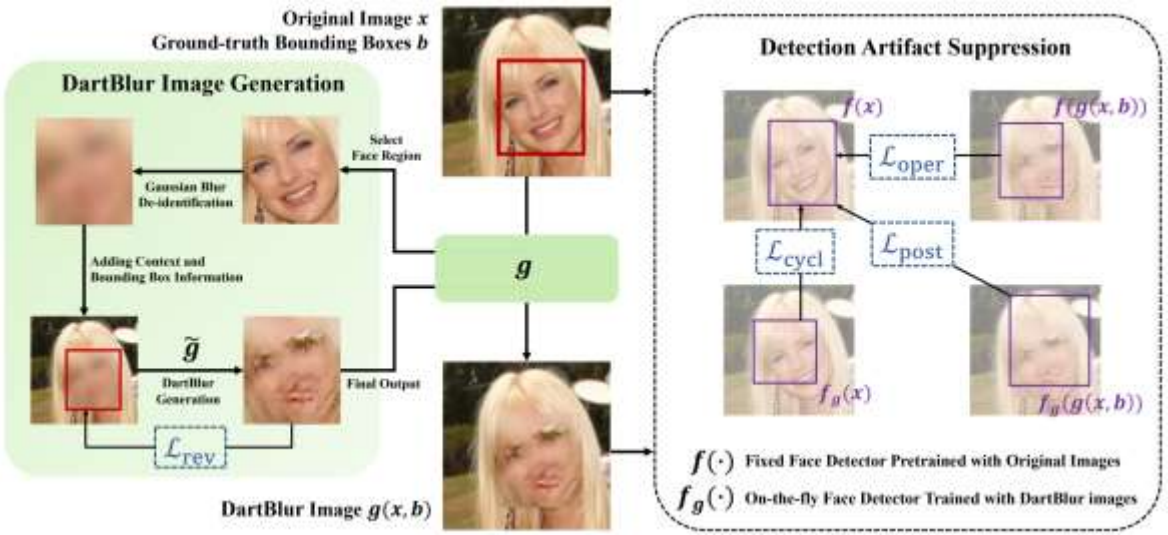
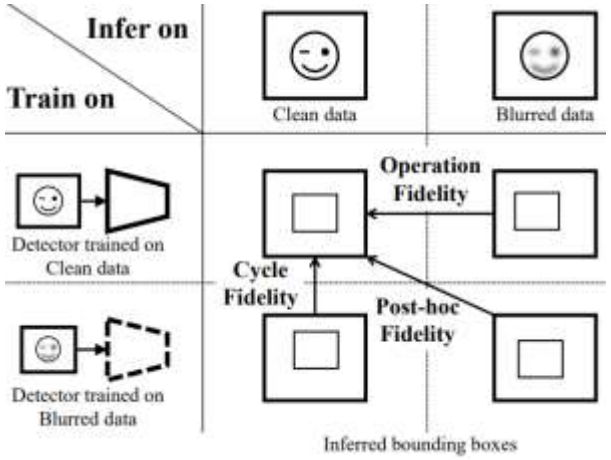


Illustration of detection artifact suppression



Algorithm 1: Training Algorithm for DartBlur

Input: Dataset $\mathcal{D}$, pretrained face detector f
Hyper-parameter: Step size α and β, threshold ϵ_{rev}
Output: Optimized parameters $\boldsymbol{\theta}_g^*$

Randomly initialize $\boldsymbol{\theta}_g$ and $\boldsymbol{\theta}_{f_g}$;
while *not converge* **do**
  Sample a batch of data $\boldsymbol{x}, \boldsymbol{b} \sim \mathcal{D}$;
  `// Optimize g with f_g fixed`
  Update $\boldsymbol{\theta}_g \leftarrow \boldsymbol{\theta}_g - \beta \nabla_{\boldsymbol{\theta}_g} \mathcal{L}_{\text{rev}}\left(g, \boldsymbol{x}, \boldsymbol{b}, \epsilon_{\text{rev}}\right)$;
  Update $\boldsymbol{\theta}_g \leftarrow \boldsymbol{\theta}_g - \beta \nabla_{\boldsymbol{\theta}_g} \mathcal{L}_{\text{det}}\Big(f\big(g(\boldsymbol{x}, \boldsymbol{b})\big), f(\boldsymbol{x})\Big)$;
  Update $\boldsymbol{\theta}_g \leftarrow \boldsymbol{\theta}_g - \beta \nabla_{\boldsymbol{\theta}_g} \mathcal{L}_{\text{det}}\Big(f_g\big(g(\boldsymbol{x}, \boldsymbol{b})\big), f(\boldsymbol{x})\Big)$;
  `// Optimize g considering second-order effects`
  Compute adapted parameters with gradient descent: $\boldsymbol{\theta}'_{f_g} = \boldsymbol{\theta}_{f_g} - \alpha \nabla_{\boldsymbol{\theta}_{f_g}} \mathcal{L}_{\text{det}}\Big(f_g\big(g(\boldsymbol{x}, \boldsymbol{b})\big), \boldsymbol{b}\Big)$;
  Update $\boldsymbol{\theta}_g \leftarrow \boldsymbol{\theta}_g - \beta \nabla_{\boldsymbol{\theta}_g} \mathcal{L}_{\text{det}}\big(f_g(\boldsymbol{x}; \boldsymbol{\theta}'_{f_g}), f(\boldsymbol{x})\big)$;
  `// Optimize f_g with g fixed`
  Update $\boldsymbol{\theta}_{f_g} \leftarrow \boldsymbol{\theta}_{f_g} - \beta \nabla_{\boldsymbol{\theta}_{f_g}} \mathcal{L}_{\text{det}}\Big(f_g\big(g(\boldsymbol{x}, \boldsymbol{b})\big), \boldsymbol{b}\Big)$;
end

- **Propose a new blur-based privacy preservation model DartBlur by taking into account the actual accessibility of the model, review convenience, and detection artifact suppression simultaneously.**
- **Introduce four novel training objectives that each directly addresses the desired properties and design an adversarial training strategy with a second-order optimization for model training.**
- **Effectively protect personal privacy while suppressing detection artifacts on various benchmarks.**

DartBlur: Privacy Preservation with Detection Artifact Suppression

Baowei Jiang*, Bing Bai*, Haozhe Lin*, Yu Wang,
Yuchen Guo, Lu Fang

Tsinghua University

*authors contributed equally

Motivation

- Blur-based methods are already preferred to protect private information in real-world data, e.g. ImageNet and Ego4D.

ImageNet

Ego4D

Motivation

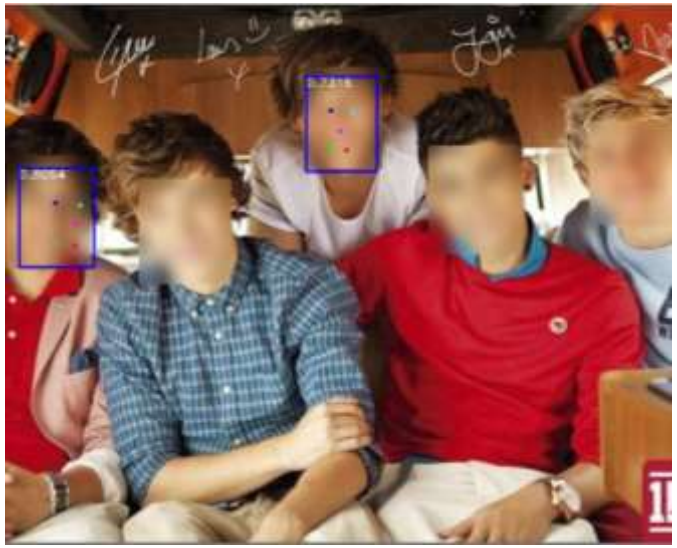
- Images with blurred faces produce artifacts for object detection task, and affect the performance of the face detector

ImageNet

Ego4D

‘Blurred’ RetinaFace on clean faces

‘Clean’ RetinaFace on blurred faces



Motivation

- Based on visualization and evaluation, both Gaussian blur and Pixelation produce artifacts on the face detector

easy medium hard

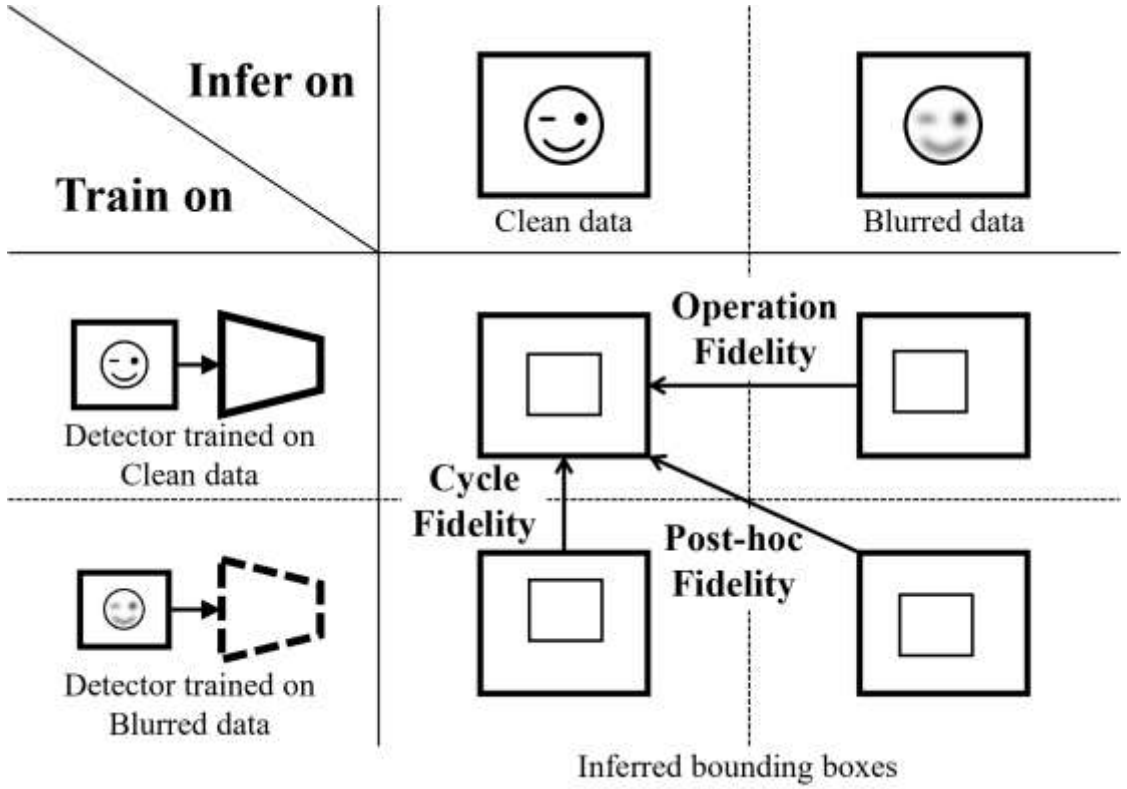
RetinaFace training and evaluation

Train \ val dataset	Origin	Gaussian Blur	Pixelation
Origin	0.921005 0.901564 0.745564	0.784570 0.790573 0.651209	0.764316 0.708425 0.544112
Gaussian Blur	0.000000 0.000037 0.011646	0.996980 0.993339 0.897396	
Pixelation	0.000000 0.000076 0.000121		0.996482 0.995469 0.952062

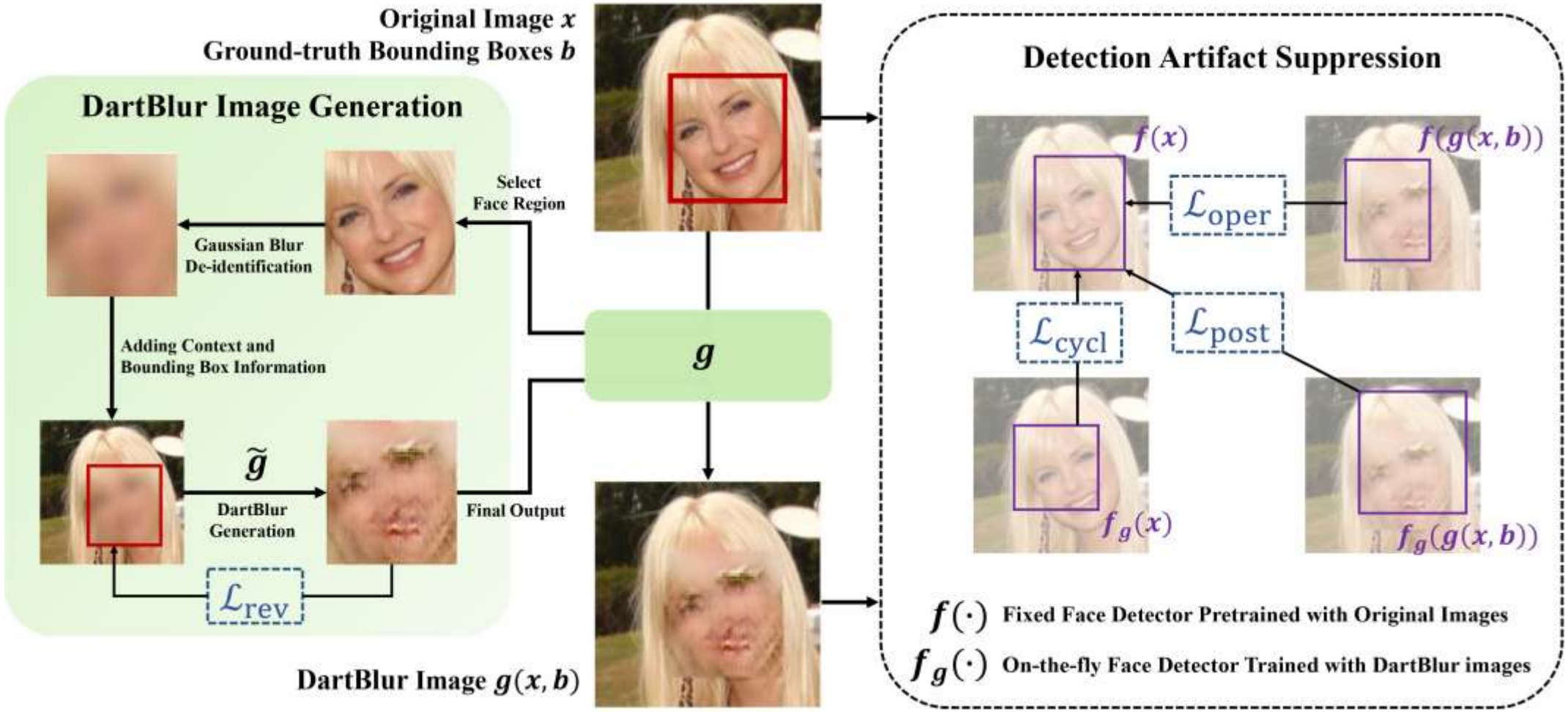
Overview

We design a novel blur-based privacy protection paradigm with the following goals:

- Accessibility.
- Review convenience.
- Detection artifact suppression.



Pipeline


Original Image x
Ground-truth Bounding Boxes b
DartBlur Image Generation
Select Face Region
Gaussian Blur De-identification
Adding Context and Bounding Box Information
$\tilde{g}$
DartBlur Generation
Final Output
$\mathcal{L}_{rev}$
g
DartBlur Image $g(x, b)$
Detection Artifact Suppression
$f(x)$
$f(g(x, b))$
$\mathcal{L}_{oper}$
$\mathcal{L}_{cycl}$
$\mathcal{L}_{post}$
$f_g(x)$
$f_g(g(x, b))$
$f(\cdot)$ Fixed Face Detector Pretrained with Original Images
$f_g(\cdot)$ On-the-fly Face Detector Trained with DartBlur images

Optimization

- Input image $\mathbf{x}$ and bbox $\mathbf{b}$, Unet $\mathbf{g}$
- Fixed $\mathbf{f_g}$
  - Solve $\min_{\mathbf{g}} \|\mathbf{g}(\mathbf{x}) - \mathbf{x}\|$
  - Solve $\min_{\mathbf{g}} \|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{g}(\mathbf{x}))\|$

    $\min_{\mathbf{g}} \|\mathbf{f_g}(\mathbf{x}) - \mathbf{f}(\mathbf{g}(\mathbf{x}))\|$
  - Solve $\min_{\mathbf{g}} \|\mathbf{f'_g}(\mathbf{x}) - \mathbf{f}(\mathbf{x})\|, \mathbf{f_g} \xrightarrow{\mathbf{b}} \mathbf{f'_g}$
- Fixed $\mathbf{g}$
  - Train $\mathbf{f_g}$ on $\mathbf{g}(\mathbf{x}, \mathbf{b})$ dataset

Algorithm 1: Training Algorithm for DartBlur

Input: Dataset $\mathcal{D}$, pretrained face detector f
Hyper-parameter: Step size α and β, threshold ϵ_{rev}
Output: Optimized parameters $\boldsymbol{\theta}_g^*$

Randomly initialize $\boldsymbol{\theta}_g$ and $\boldsymbol{\theta}_{f_g}$;
while *not converge* **do**
  Sample a batch of data $\boldsymbol{x}, \boldsymbol{b} \sim \mathcal{D}$;
  // Optimize g with f_g fixed
  Update $\boldsymbol{\theta}_g \leftarrow \boldsymbol{\theta}_g - \beta \nabla_{\boldsymbol{\theta}_g} \mathcal{L}_{\text{rev}}\left(g, \boldsymbol{x}, \boldsymbol{b}, \epsilon_{\text{rev}}\right)$;
  Update $\boldsymbol{\theta}_g \leftarrow \boldsymbol{\theta}_g - \beta \nabla_{\boldsymbol{\theta}_g} \mathcal{L}_{\text{det}}\left(f\left(g(\boldsymbol{x}, \boldsymbol{b})\right), f(\boldsymbol{x})\right)$;
  Update $\boldsymbol{\theta}_g \leftarrow \boldsymbol{\theta}_g - \beta \nabla_{\boldsymbol{\theta}_g} \mathcal{L}_{\text{det}}\left(f_g\left(g(\boldsymbol{x}, \boldsymbol{b})\right), f(\boldsymbol{x})\right)$;
  // Optimize g considering second-order effects
  Compute adapted parameters with gradient descent:
  $\boldsymbol{\theta}'_{f_g} = \boldsymbol{\theta}_{f_g} - \alpha \nabla_{\boldsymbol{\theta}_{f_g}} \mathcal{L}_{\text{det}}\left(f_g\left(g(\boldsymbol{x}, \boldsymbol{b})\right), \boldsymbol{b}\right)$;
  Update $\boldsymbol{\theta}_g \leftarrow \boldsymbol{\theta}_g - \beta \nabla_{\boldsymbol{\theta}_g} \mathcal{L}_{\text{det}}\left(f_g(\boldsymbol{x}; \boldsymbol{\theta}'_{f_g}), f(\boldsymbol{x})\right)$;
  // Optimize f_g with g fixed
  Update $\boldsymbol{\theta}_{f_g} \leftarrow \boldsymbol{\theta}_{f_g} - \beta \nabla_{\boldsymbol{\theta}_{f_g}} \mathcal{L}_{\text{det}}\left(f_g\left(g(\boldsymbol{x}, \boldsymbol{b})\right), \boldsymbol{b}\right)$;
end

Evaluation

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

- DartBlur trained with WIDER FACE successfully suppressed detection artifacts, and was generalizable across different datasets, and could be well generalized across different datasets and architectures for detection.

Dataset	Fidelity	Block	Blur	Pixel.	DartBlur
WIDER FACE	Oper. Fid.	19.20	83.10	46.35	**96.76**
	Post-hoc Fid.	77.77	84.04	80.08	**84.70**
	Cycle Fid.	0.10	1.97	7.64	**47.22**
FDDB	Oper. Fid.	4.63	89.00	4.82	**98.06**
	Post-hoc Fid.	85.49	93.41	88.34	**94.64**
	Cycle Fid.	0.00	0.05	0.02	**41.93**
CrowdHuman	Oper. Fid.	16.69	75.74	45.26	**81.64**
	Post-hoc Fid.	58.23	57.51	59.60	**62.52**
	Cycle Fid.	0.83	36.72	28.75	**45.11**

Evaluation results of detection artifact suppression and cross-dataset transferability

Architecture	Fidelity	Block	Blur	Pixel.	DartBlur
PyramidBox	Oper. Fid.	21.34	84.60	30.55	**95.18**
	Post-hoc Fid.	75.04	70.59	65.18	**75.16**
	Cycle Fid.	0.01	1.70	10.98	**24.68**
YOLOv5	Oper. Fid.	35.93	84.84	35.22	**96.17**
	Post-hoc Fid.	85.68	87.44	86.01	**91.72**
	Cycle Fid.	0.21	10.00	0.36	**37.15**

Cross-architecture transferability evaluation on WIDER FACE

Privacy recovering

An indomain model is trained to reconstruct the original images from DartBlur images. As illustrated in the following equation. the function $\bar{g}$ acts only on the faces region,

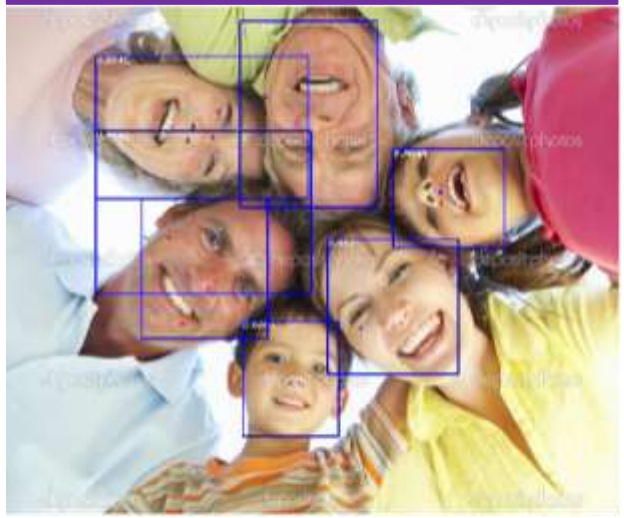
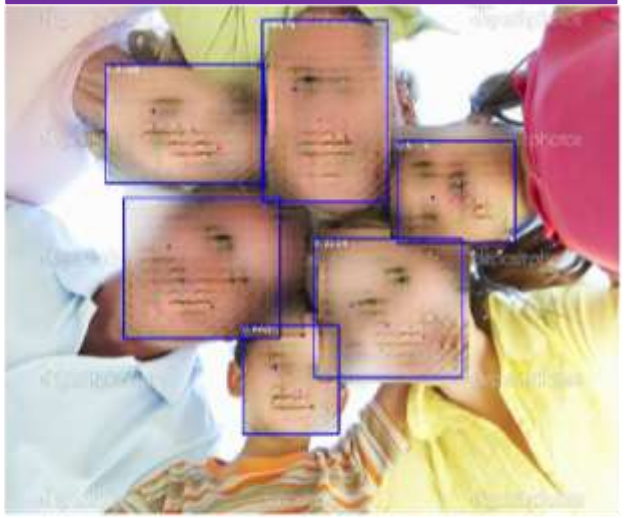
$$\bar{g}(g(x,b),b) = x \odot (1-b) + \bar{g}(g(x,b)) \odot b,$$

where $\bar{g}$ maintain the same structure as $\tilde{g}$. The figure shows 3 cases and demonstrates that erased privacy information cannot be trivially restored.

Dartblur-Faces

Detection results comparison

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Original	*Gaussian Blur*	*DartBlur*
		
		

Thanks!