CTU
CZECH TECHNICAL
UNIVERSITY
IN PRAGUE

F3

Faculty of Electrical Engineering
Department of Cybernetics

Bachelor's Thesis

Anonymization of Faces in Images and Videos

Ditmar Hadera
Cybernetics and Robotics

May 2024
Supervisor: Ing. Jan Čech, Ph.D.

BACHELOR'S THESIS ASSIGNMENT

I. Personal and study details

Student's name:	**Hadera Ditmar**	Personal ID number:	**508514**
Faculty / Institute:	**Faculty of Electrical Engineering**		
Department / Institute:	**Department of Cybernetics**		
Study program:	**Cybernetics and Robotics**		

II. Bachelor's thesis details

Bachelor's thesis title in English:

Anonymization of faces in images and videos

Bachelor's thesis title in Czech:

Anonymizace obličejů v obrazcích a ve videu

Guidelines:

1. Anonymization (also known as de-identification) of face photographs serves to erase the identity of a person, as personal data. There are trivial methods of anonymization, where the face is blurred or replaced by a black rectangle. However, modern methods aim to change the face only so that the person is not recognizable, and at the same time all attributes (expression, gaze, age, race, etc.) are preserved and it is not obvious at first glance that the original photograph was modified, e.g., the resulting content does not contain visible seams or lighting inconsistencies.
2. Another challenge is the anonymization of faces in videos, where the novel identity replacing the original one is consistent across all frames in the video. The anonymized video does not contain artifacts of temporal inconsistency.
3. Review existing methods and use/modify existing methods, or propose your own method for face anonymization. Evaluate the method for anonymization strength (if the method really conceals the original identity), for facial expression fidelity, and for quality and consistency of the resulting images (e.g., no obvious seems should be visible). Furthermore, for videos, evaluate temporal and identity consistency across frames.

Bibliography / sources:

[1] Hakon Hukkelas, Frank Lindseth. DeepPrivacy2: Towards Realistic Full-Body Anonymization. In WACV, 2023.
[2] Marvin Klemp, Kevin Rösch, Royden Wagner, Jannik Quehl, Martin Lauer. LDFA: Latent Diffusion Face Anonymization for Self-driving Applications. In arXiv:2302.08931, 2023.
[3] Robin Rombach and Andreas Blattmann and Dominik Lorenz and Patrick Esser and Björn Ommer. High-Resolution Image Synthesis With Latent Diffusion Models. In CVPR, 2022.
[4] Lvmin Zhang and Anyi Rao and Maneesh Agrawala. Adding Conditional Control to Text-to-Image Diffusion Models. In ICCV, 2023.
[5] Jiri Moravcik. Face Anonymizer. Master's thesis, Czech Technical University in Prague, FEE, 2023. (https://dspace.cvut.cz/handle/10467/109451)

Name and workplace of bachelor's thesis supervisor:

Ing. Jan Čech, Ph.D. Visual Recognition Group FEE

Name and workplace of second bachelor's thesis supervisor or consultant:

Date of bachelor's thesis assignment: **02.02.2024** Deadline for bachelor thesis submission: **24.05.2024**

Assignment valid until: **21.09.2025**

Ing. Jan Čech, Ph.D. Supervisor's signature	prof. Dr. Ing. Jan Kybic Head of department's signature	prof. Mgr. Petr Páta, Ph.D. Dean's signature

/ Declaration

I declare that the presented work was developed independently and that I have listed all sources of information used within it in accordance with the methodical instructions for observing the ethical principles in the preparation of university theses. This includes leveraging tools like OpenAI's ChatGPT for linguistic refinement.

Prague, date 24. 5. 2024

....

Abstrakt / Abstract

Tato práce představuje novou metodu pro anonymizaci obličejů na obrázcích a ve videích, která mění identitu obličeje při zachování klíčových obličejových atributů, jako jsou věk, pohlaví, rasa, pozice a výraz. Náš přístup je založen na inpaintingu pomocí moderního difuzního modelu, konkrétně Stable Diffusion od Stability AI. Provedli jsme rozsáhlé experimentální testování a kvantitativně jsme metodu vyhodnotili pomocí několika navrhovaných statistik, které měří stupeň deidentifikace, zachování obličejových atributů a úroveň perceptuálních artefaktů. Představujeme velké množství kvalitativních výsledků. Perceptuální realismus anonymizovaných obličejových obrázků je také měřen pomocí malé uživatelské studie. Naši metodu jsme testovali proti populární nedávné metodě DeepPrivacy 2 (Hukkelas, 2023) s nadějnými výsledky. V mnoha aspektech naše metoda dosahuje srovnatelných výsledků a překonává DeepPrivacy 2 v zachování výrazů.

Klíčová slova: anonymizace obličeje, deidentifikace, inpainting, generativní model, Stable Diffusion

Překlad titulu: Anonymizace obličejů v obrázcích a ve videu

The thesis proposes a novel method for anonymizing faces in images and videos that alters the identity of the face while preserving key facial attributes, such as age, gender, race, pose, and expression. Our approach is based on inpainting using a recent diffusion model, specifically the Stable Diffusion by Stability AI. We have conducted extensive experimental testing and quantitatively evaluated the method using several proposed statistics that measure the degree of de-identification, preservation of facial attributes, and the level of perceptual artifacts. We present a large number of qualitative results. The perceptual realism of the anonymized face images is also measured using a small-scale user study. Our method was tested against the popular recent Deep Privacy v2 (Hukkelas, 2023) method with promising results. In many aspects, our method achieves comparable results and outperforms the Deep Privacy v2 in preserving expressions.

Keywords: facial anonymizatinon, de-identification, inpainting, generative model, Stable Diffusion

Contents /

Tables / Figures

Chapter 1
Introduction

Facial anonymization or de-identification is the process obscuring or altering faces of individuals in images or videos in a way that they cannot be easily recognized or identified.

Sharing datasets including raw video footage is critical to make progress in psychological science [1]. This is also - perhaps even more - important when studying development, which needs to be sampled frequently [2]. However, video footage of infants and children is sensitive data - more sensitive than recordings of adults. Therefore, changing the visual appearance of infants in videos to hide their identity while preserving performance in downstream tasks (pose estimation, gaze estimation) is key for progress in developmental science.

Despite the availability of various anonymization methods, many do not meet the specific needs for maintaining the usability of data in downstream tasks. Current techniques often struggle to balance effective de-identification with the preservation of essential facial features.

To address this issue, we propose a new method for facial anonymization applicable to both images and videos. Our approach utilizes a diffusion-based image generator, specifically Stable Diffusion. We start by creating a binary mask for each face in an image and then iteratively apply inpainting to obscure all the faces in the image. To preserve facial features, we condition Stable Diffusion with multiple parameters.

Our contributions include the development of this novel anonymization method, the creation of a benchmarking suite for its evaluation, and a comparative analysis with an existing anonymization method with similar objectives.

Chapter 2
Related Work

There exists a plethora of anonymization techniques. An overview of them can be seen in [3]. They all differ in mainly 2 aspects. First is how well they anonymize a face. Meaning how hard is it to figure out who the person in the image was before the anonymization method has been applied. Second is how much the anonymized face still look like a human face.

2.1 Face Obscuring Methods

These methods are based on obfuscating a face or its part. They include blurring, pixelization or masking. A few images of faces anonymized using these techniques can be seen in Fig. 2.1.

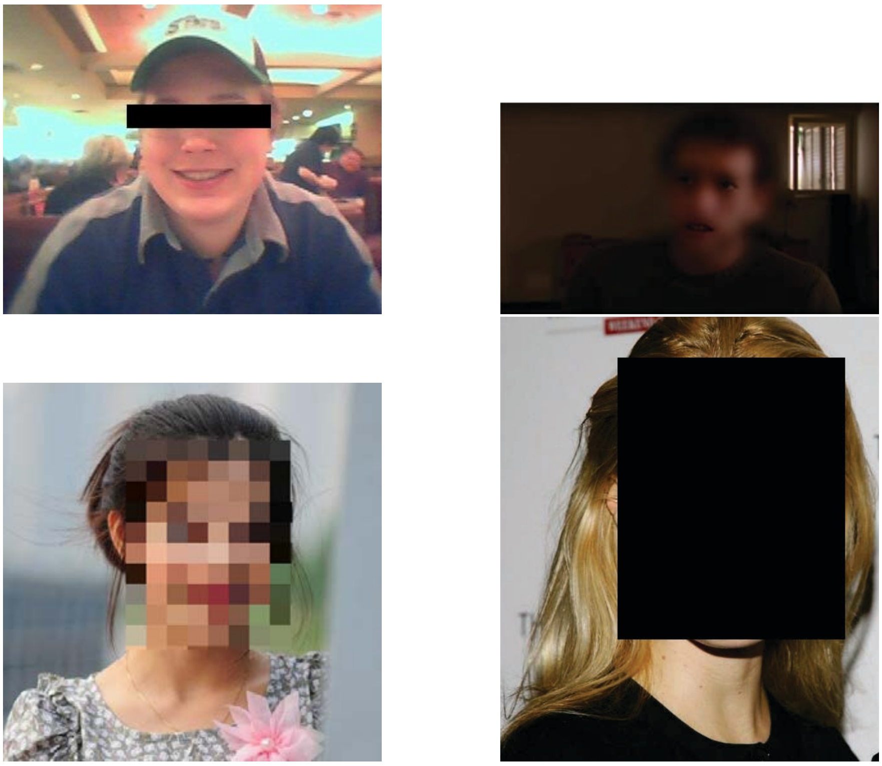


Figure 2.1. Comparison of different face obscuring anonymization methods.

The advantage of these methods is their simplicity. It allows them to be fast and not resource intensive. On the other hand their disadvantage is that their de-identification effectiveness is directly proportional to how much the anonymized face still looks like a real face. Using images generated with these methods on downstream tasks is therefore impossible most of the time.

2.2 Face Modifying Methods

Unlike the face obscuring methods, these methods try to modify the faces they are anonymizing. This allows them to anonymize their identities while still keeping all the aspects of a human. This reality makes them usable for anonymizing data and not loosing the possibility of using it for downstream tasks. Unfortunately it also makes them more complex, which result in either higher resource intensivness and longer anonymization times.

The existing methods can be distinguished based on whether they utilize deep learning techniques or not.

The non-deeplearning methods mostly utilize face swapping techniques. The face that is being anonymized is compared to many faces from a large dataset. Then one or multiple of the faces that are the closest by some metric are picked. They are the image is then anonymized through interpolatin. Examples of these methods are the k-Same algorithm [4] and AnonySwap [5].

The methods utilizing deep learning approach the anonymization in multiple different ways. Examples of these methods include StyleGAN3 [6], or DeepPrivacy 2 [7].

Chapter 3
Method

We introduce our method for anonymizing images. It consists of four steps:

1. Detecting all faces in an image.
2. Finding facial landmarks of each detected face.
3. Generating a binary mask for the detected face.
4. Inpainting the masked-out face in the original image to anonymize it.

First, all faces are detected in the original image. The subsequent three steps are performed repeatedly on each detected face until all faces are anonymized. These steps can be applied to any image, regardless of the number of faces it contains.

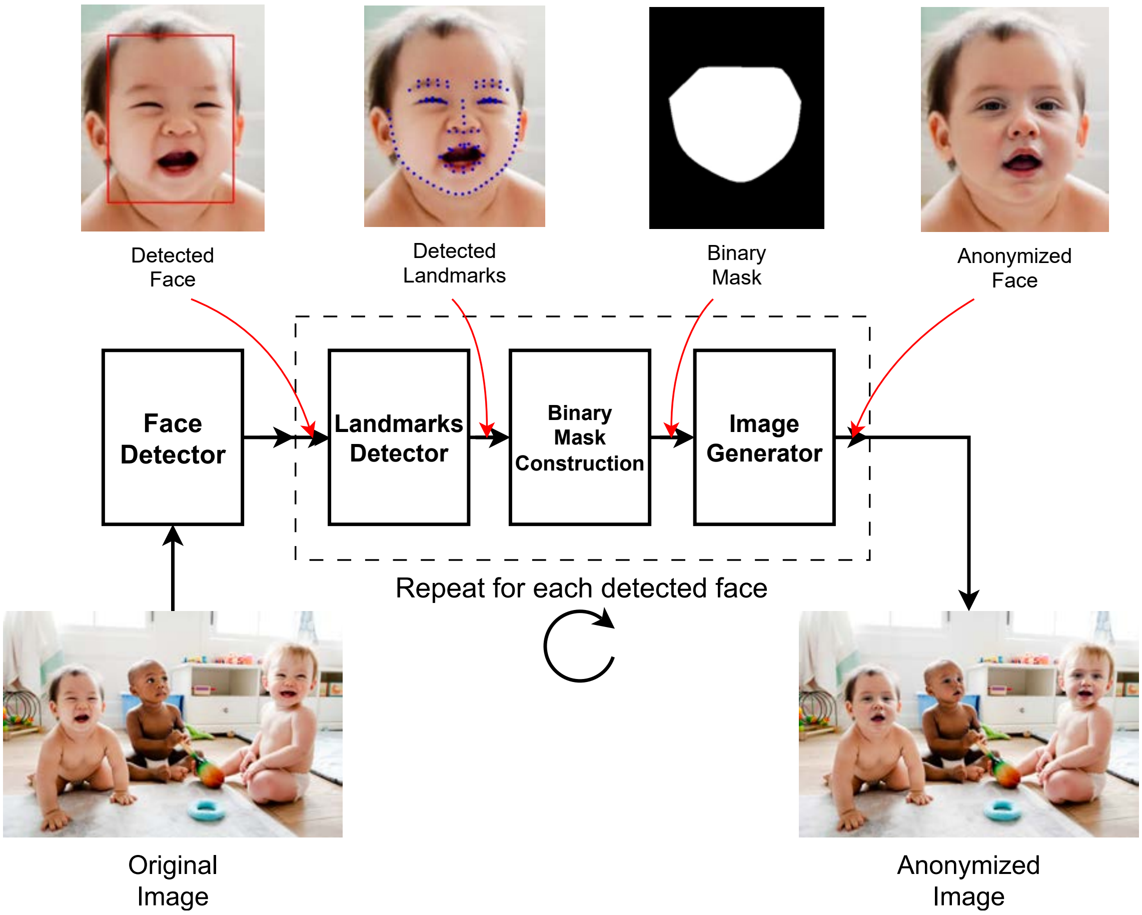


Figure 3.1. Diagram of the proposed anonymization method.

The method is designed to be modular, allowing for the substitution of different libraries for each step without disrupting the overall pipeline.

3.1 Face Detection

The initial step of our method involves detecting all faces in the current image. The positions of these faces are represented by rectangular bounding boxes, which are then passed to the next step of our method. An example of an image with a detected face and its bounding box is shown in Fig. 3.2.

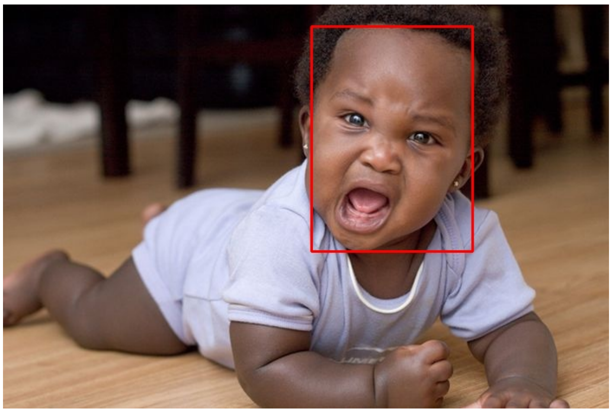


Figure 3.2. Detected face surrounded by its bounding box.

For the face detection, we use YOLOv8 [8] with a model for detecting faces [9]. When detecting faces the minimum detection confidence is set to 50%.

3.2 Facial Landmark Detection

In the second step, the bounding boxes detected in the previous step are used to find the positions of corresponding facial landmarks. These landmarks correspond to important facial features like the eyes, mouth, and nose. The positions of these landmarks are passed to the next step. An image of a face with highlighted facial landmarks can be seen in Fig. 3.3.

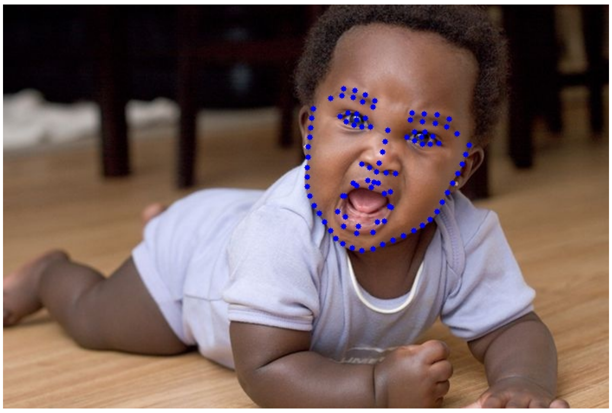


Figure 3.3. Face with highlighted facial landmarks.

To detect the positions of facial landmarks, we use the SPIGA model [10] with the `wflw` weights, which detects 98 distinct landmarks.

3.3 Binary Mask Construction

To highlight the inner part of the detected face, we construct a binary mask using the facial landmarks identified in the previous step. We begin by finding the convex hull of these landmarks. Next, we create a black image and draw a polygon that matches the shape of the convex hull onto it. This resulting image serves as the binary mask for the next step. An example of a face with the highlighted inner part and its matching mask can be seen in Fig. 3.4.

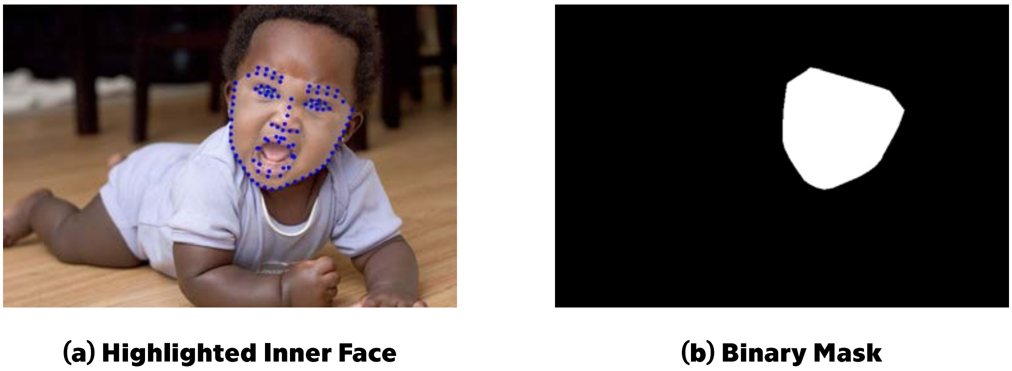

(a) Highlighted Inner Face **(b) Binary Mask**

Figure 3.4. Face with highlighted inner face and facial landmarks (a) and the corresponding binary mask (b).

3.4 Inpainting

In the final step, we use the previously generated mask and an image-generating model to anonymize the identified face through inpainting. Image inpainting, as described in [11], is the process of filling in missing parts of an image with such content, that an unknowing observer would not notice the modification. We specify the area, that is supposed to be filled in by the image generator using the generated mask. To influence the inpainting process more, we provided various parameters to the image generator. An example of the source image, the binary mask, and the resulting inpainted image is shown in Fig. 3.5. Once inpainting is complete, the output image is either used as a base for anonymizing additional faces or deemed the final anonymized result.

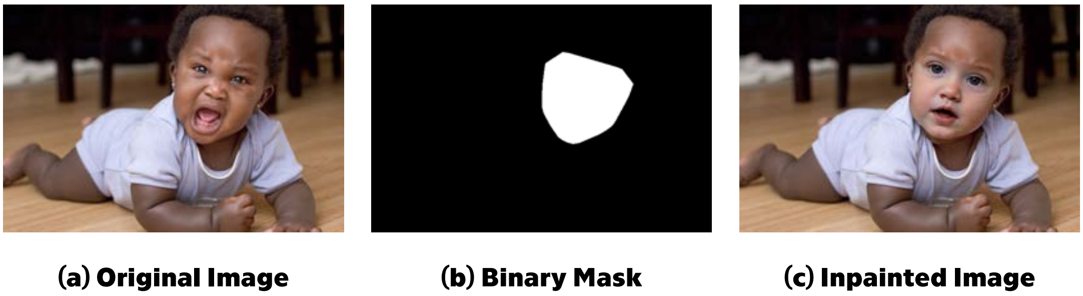

(a) Original Image **(b) Binary Mask** **(c) Inpainted Image**

Figure 3.5. Inpainting input image (a), a binary mask used for inpainting (b) and the inpainting output (c).

We employ the generative model called Stable Diffusion [12] as our image generator. Specifically its implementation called Stable Diffusion WebUI [13]. For generating images, we use two different checkpoints: `Realistic Vision V2.0`[1] and

[1] https://huggingface.co/SG161222/Realistic_Vision_V2.0

Realistic Vision V6.0 B1[2]. Typically, when inpainting we use the parameters detailed in Appendix A or slight modifications of these parameters. This model does not perform inpainting strictly as mentioned in the previous paragraph, but rather adds noise to the inpainted area which it subsequently tries to remove. This means that it usually starts with some information about what initially was in the inpainted image.

3.4.1 Parameters and Extensions

When employing Stable Diffusion for inpainting tasks, there are many parameters that influence the quality of the final generated image. Some of them influence the generation more and some less. Out of them all the most prominent ones include:

- Positive and Negative Prompts: These prompts guide the inpainting process by providing information about the desired content to be generated in the masked regions. Positive prompts encourage the model to generate content that matches the surrounding context, while negative prompts discourage the generation of undesirable artifacts.
- Classifier-Free Guidance: This parameter determines how closely the anonymization process is affected by the text prompts. Higher values of this parameter relate to a higher influence these prompts.
- Sampling Steps: This parameter is used in Stable Diffusion to control through how many steps of denoising Stable Diffifusion will go through. Where higher values usually result in higher quality images, but also longer image generations.
- Inpainting Fill: Inpainting fill refers to the type of noise used to fill in the masked regions of an image. Different types of noise are used for different reasons as they each have their advantages and disadvantages. For example *latent noise* is usually used for generating new structures into the image. But different types of noise also require different levels of denoising strength to function properly.
- Denoising Strength: Denoising strength controls how much noise is added to the inpainted image before the denoising process starts. Lower values thus make the inpainting result resemble the original image more closely.

Extensions for Stable Diffusion are additional modules or features that enhance the capabilities or performance of the Stable Diffusion algorithm. These extensions are designed to address specific challenges or improve certain aspects of the image generation process.

When creating our method we have used two separate extensions, namely:

- **ControlNet:** The ControlNet extension for Stable Diffusion [14] is a neural network used to enhance the stability and performance of diffusion-based image inpainting algorithms. It achieves this by incorporating additional information from a control network into the inpainting process.

  We have incorporated this extension into our method to guide the inpainting process so the generated faces align correctly with the original ones. For this purpose we have chosen the OpenPose model to give the image generator as much information about the body proportions of the original person.
- **API Payload Display:** The API Payload Display extension translates any image generation requests from the Web UI into a table of parameters sent to Stable Diffusion.

[2] https://civitai.com/models/4201/realistic-vision-v60-b1

We have used this extension to convert our base set of parameters, we found by manual testing in the the Web UI, into a structure usable in code. But this extension is not a directly used as a part of our anonymization method.

3.5 Videos

The straightforward nature of this method also makes it suitable for video anonymization, which can be achieved by processing each frame individually. However, since this approach does not account for continuity between frames, the anonymized identity may vary from frame to frame. To address this issue, we employ a strategy that conditions the image generator to produce consistent identities across frames. In theory this is achieved by incorporating multiple celebrity names into the positive prompt. During the generation process, the image generator partially focuses on creating faces resembling these celebrities, thereby maintaining consistency in the anonymized identity throughout the video.

3.6 Factors Affecting Anonymization Quality

The quality of the anonymized images produced by our method is influenced by several factors. By *quality*, we refer to both the effectiveness of de-identifying (anonymizing) an individual and the realism of the anonymized face. For instance, a face with significant disfigurement and numerous inpainting artifacts is considered to be of low quality.

We investigated the impact of some Stable Diffusion parameters on image quality in the experiments described in Chapter 5. Further examination of the factors discussed in the following sections, that we noticed during our experimentation, could help enhance the proposed method.

3.6.1 The Original Face

The resolution and orientation of the original face can significantly impact the resulting image. Faces not oriented towards the camera tend to exhibit more artifacts when anonymized. Additionally, obstructions such as limbs or objects in front of the face negatively affect the quality. Other factors, such as race and age, also influence the anonymization outcome. These effects could be caused by underrepresentation of certain traits in the training data of the used models.

3.6.2 Binary Mask

The shape of the binary mask plays a significant role in the resulting image quality. Well-fitting masks generally produce higher quality images, while poorly fitting masks, that are the product of incorrectly detected landmarks, result in images with numerous artifacts.

Possible cause for this could be that Stable Diffusion is trying to fit a whole face into the area indicated by the mask as the positive prompt guides it to do. But poorly fitting mask usually cover only a part of the face resulting in conflicts.

3.6.3 Image Generator Model

The choice of the image generator model significantly impacts the quality of the anonymized faces. Different models produce faces with varying degrees of artifacts and realism. Selecting the appropriate model is critical for achieving high-quality anonymized images.

3.6.4 Image Generator Parameters

The parameters of the image generator affect the output quality in various ways. Some parameters influence the effectiveness of anonymization, others affect the realism of the final face, and some impact both. Typically, there is a trade-off between the realism of a face and the degree of anonymization.

3.6.5 Randomness

An inherent randomness in the image generation process can lead to variations in the anonymized face quality. Even with unchanged parameters, the output can fluctuate due to a different realizations of noise.

Chapter 4
Benchmarking Suite

To empirically evaluate the performance of various facial anonymization methods, we have developed a comprehensive benchmarking suite. This suite is designed to assess both images and videos, drawing inspiration from AnonyBench [15]. While some of the metrics are adopted directly from AnonyBench, others have been omitted, and several new metrics have been introduced.

To measure the quality of an anonymization method, we compare the original input to its anonymized output. This comparison is performed for each image and each frame of a video. Since multiple faces can appear in each image or frame, our suite maps every face from the original image or frame to its anonymized counterpart.

To map the faces we first identify bounding boxes for each face in both the original and anonymized images. Each anonymized face is then mapped to the original face whose center is closest. Due to potential errors during anonymization, this mapping may not always be bijective. Consequently, we categorize the mappings into two states:

1. Correctly Mapped: An original face is mapped to a single anonymized face.
2. Incorrectly Mapped: An original face is either not mapped to any face or is mapped to multiple anonymized faces.

An example of an image containing correctly and incorrectly mapped faces is shown in Fig. 4.1. When evaluating the overall statistics of an image, only the statistics of correctly mapped faces are considered. This is because certain metrics cannot be accurately measured for incorrectly mapped faces.

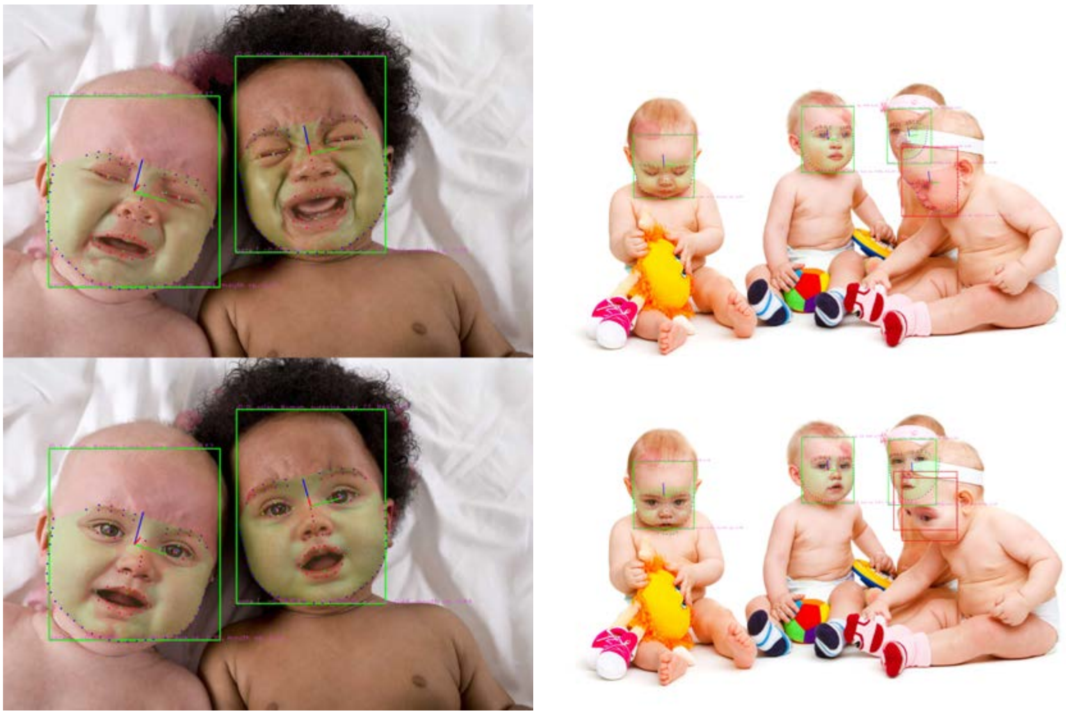
For face and facial landmark detection, we use the same libraries as described in our anonymization method in Chapter 3.

4.1 Image Statistics

During our evaluation of an image, we assess various statistics, which can be categorized into four main categories:

- Basic image statistics and mapping.
- De-identification quality.
- Preservation of facial attributes.
- Presence of anonymization artifacts.

Each statistic is either independently evaluated on both the original face and its anonymized counterpart, or it comparing corresponding traits between the original and anonymized faces. An illustration featuring an original image alongside its anonymized counterpart, with various evaluated statistics highlighted, is depicted in Fig. 4.1.



(a) First Comparison Image (b) Second Comparison Image

Figure 4.1. Two comparison images used for debugging. Each is composed of two images where the top one is an original image an the bottom one is its anonymized version. In the first image (a) both faces are correctly matched (highlighted by green bounding boxes). In the second image (b) three faces a matched correctly and one is matched incorrectly (highlighted by red bounding boxes). Besides information about the matching both images also contain highlighted areas of the inner faces (in green), detected landmarks (in colors ranging from blue to red) and facial orientations highlighted using 3 lines.

4.1.1 Basic Image Statistics

This set of statistics provides insight into the technical parameters of the evaluated image and the outcomes of mapping original faces to their anonymized counterparts. These technical parameters do not directly measure the quality of anonymization but, as discussed in Section 3.6, they can influence anonymization quality.

The statistics include:

- **Image Resolution:** The width and height of the image in pixels.
- **Bounding Box Size:** The average width and height of a face bounding box in pixels within the image.
- **Total Detected Faces:** Indicates the total number of faces detected in the original image.
- **Total Correctly Mapped Faces:** Describes how many of all the detected faces in the original image were correctly mapped to a face in the anonymized image.
- **Total Incorrectly Mapped Faces:** Describes how many of all the detected faces in the original image were incorrectly mapped.

4.1.2 De-identification Quality

To assess how well a method anonymizes identities we utilize two statistics. To be able to evaluate these statistics we first find the identity vectors of both the original and anonymized faces. To obtain these identity vectors we use the DeepFace library [16] in conjunction with the ArcFace model. Each identity vector outputted by ArcFace is 512-dimensional.

The measured statistics are as follows:

- **Identity Cosine Distance:** This metric, denoted as D_c, quantifies the cosine distance between the original identity vector $\mathbf{v}_o$ and the anonymized identity vector $\mathbf{v}_a$. It is calculated using the equation:

$$D_c = 1 - \mathbf{v}_o^T \mathbf{v}_a,$$

  where both $\mathbf{v}_o$ and $\mathbf{v}_a$ are unit vectors, and the T signifies transposition.
- **Same Identity:** By evaluating the cosine distance between identity vectors and applying an internal threshold from the DeepFace library [16], we determine whether the original and anonymized faces belong to the same identity.

4.1.3 Facial Attributes Preservation

To assess the preservation of facial attributes of different anonymization methods, we employ multiple different statistics. These include:

- **Detector Confidence:** This statistic reflects the confidence level of facial detectors in correctly identifying the searched object. We capture the output value of our face detector when detecting both original and anonymized faces.
- **Race:** the DeepFace library [16], we determine the race of both original and anonymized faces. The possible racial categories include:
  - Indian,
  - Asian,
  - Latino Hispanic,
  - Black,
  - Middle Eastern,
  - White.

  When comparing the original and anonymized faces we evaluate whether they are of the same race, or not.
- **Gender:** Similarly, we utilize the DeepFace library to discern the gender of both original and anonymized faces. The comparison between the original and anonymized faces is done by comparing whether they are of the same gender.
- **Emotion:** Employing the DeepFace library, we ascertain the emotional state of original and anonymized faces. Possible emotion categories are:
  - sad,
  - angry,
  - surprise,
  - fear,
  - happy,
  - disgust,
  - neutral.

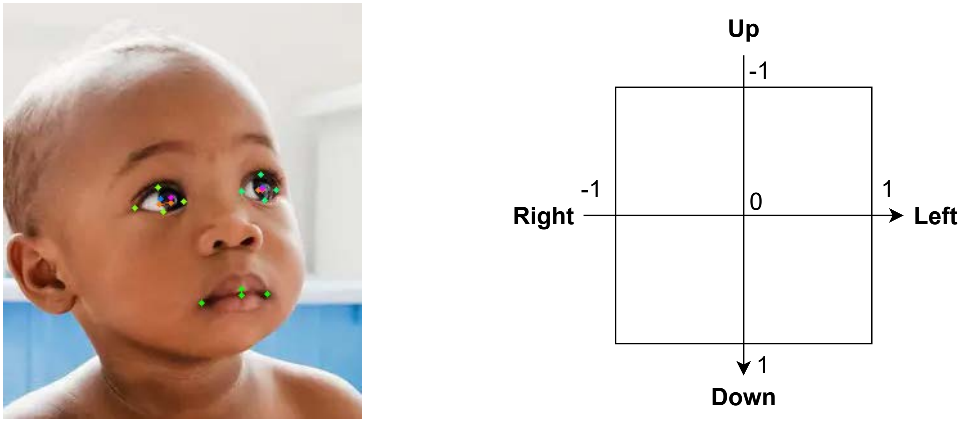
  When evaluating a pair of original and anonymized faces we compare if they have the same emotion, or not.
- **Age:** The DeepFace library is employed to estimate the age of both original and anonymized faces.
- **Eye and Mouth Openness:** For each eye and the mouth, we calculate how open it is. This involves calculating the openness value O based on the pixel distances between

the top and bottom of the eye or mouth d_{tb} and the left and right edges d_{lr}. Using these distances we compute the openness value O as:

$$O = \frac{d_{tb}}{d_{lr}}.$$

A face with highlighted eye and mouth edges can be seen in Fig. 4.2.

- **Gaze:** To try to determine which way each detected face is looking we compute a gaze vector for each eye. The gaze vector is a 2-dimensional vector whose components are in the range $\langle -1; 1 \rangle$. A diagram showing where an eye is looking based on its gaze vector can be seen if Fig. 4.2.



(a) Highlighted Face

(b) Gaze Directions Diagram

Figure 4.2. Face with highlighted edges of eyes and mouth (in green), pupils (in pink), projection of the vectors between vertical centers of eyes and pupils $\mathbf{v}_{vp}$ onto the vectors between vertical centers and the bottom edges of the eyes $\mathbf{v}_{vb}$ (in blue) and projection of the vectors between horizontal centers of eyes and pupils $\mathbf{v}_{hp}$ onto the vectors between horizontal centers and the left edges of the eyes $\mathbf{v}_{hl}$ (in orange) (a) and a diagram showing where a person is looking based on the gaze vector (b).

The gaze vector is calculated using multiple steps. First we find the pixel coordinates of the left $\mathbf{e}_l$, right $\mathbf{e}_r$, top $\mathbf{e}_t$ and bottom $\mathbf{e}_b$ edges of an eye and the pixel position of its pupil $\mathbf{p}$ from the facial landmarks. The positions of the edges are considered relative to face, not the image, meaning that the left edge is towards the left side of the face. We then find the horizontal center between the left and right edge $\mathbf{c}_h$ and the vertical center between the top and the bottom edge $\mathbf{c}_v$ using these formulas:

$$\mathbf{c}_h = \frac{\mathbf{e}_l + \mathbf{e}_r}{2},$$

$$\mathbf{c}_v = \frac{\mathbf{e}_t + \mathbf{e}_b}{2}.$$

Then we find the vectors representing the difference between pupil position and the position of each of the centers:

$$\mathbf{v}_{hp} = \mathbf{p} - \mathbf{c}_h,$$

$$\mathbf{v}_{vp} = \mathbf{p} - \mathbf{c}_v.$$

After that we find the vector between the horizontal center and the left edge. Similarly with the vertical center and the bottom edge:

$$\mathbf{v}_{hl} = \mathbf{e}_l - \mathbf{c}_h,$$

$$\mathbf{v}_{vb} = \mathbf{e}_b - \mathbf{c}_v.$$

Finally we compute the lengths of orthogonal projections of the vectors between the centers and the pupil onto the vectors between the centers and the edges:

$$g_h = \frac{\mathbf{v}_{hp} \cdot \mathbf{v}_{hl}}{\mathbf{v}_{hl} \cdot \mathbf{v}_{hl}},$$

$$g_v = \frac{\mathbf{v}_{vp} \cdot \mathbf{v}_{vb}}{\mathbf{v}_{vb} \cdot \mathbf{v}_{vb}}.$$

From these projections we construct the gaze vector:

$$\mathbf{g} = \begin{bmatrix} g_h \\ g_v \end{bmatrix}.$$

When comparing gaze vectors of original and anonymized faces first an average gaze vector of each face is calculated from the gaze vectors of individual eyes. Then as the gaze difference we pronounce the norm of the difference between the gaze vectors of both faces.

- **Orientation:** The orientation of each detected face is internally represented using a three dimensional rotation matrix. We find the facial orientation using the facial landmarks detector that outputs it as a secondary value when detecting facial landmarks. A face with its orientation highlighted can be seen in Fig. 4.3. To find the difference ΔR between the orientation of the original face R_o and the anonymized face R_a, we find the following matrix product:

$$\Delta R = R_o R_a^T,$$

where T denotes transposition, which is equivalent to inversion for rotation matrices. To be able to more easily assess the results by eye we convert the rotation matrices to euler angles for presentation.

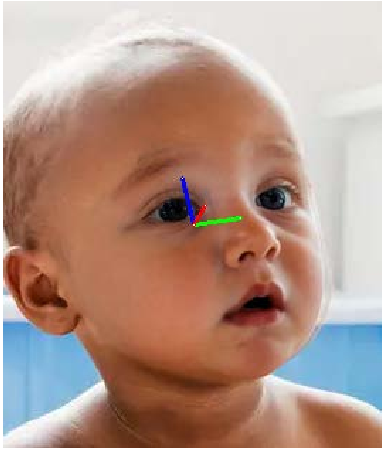


Figure 4.3. Face with highlighted basis vectors of a right handed orthonormal coordinate system representing the orientation of the face. The x-axis (in red) is pointing almost towards the camera, y-axis (in green) is pointing to the left side of the face and z-axis (in blue) is pointing above the face.

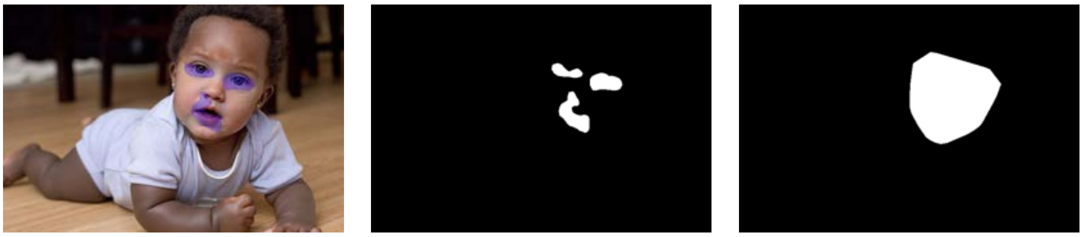
4.1.4 Anonymization Artifacts

In order to quantify the presence of artifacts resulting from the anonymization process, we opted to utilize the Perceptual Artifact Ratio PAR as outlined in PAL4Inpaint [17]. As its name implies, PAR aims to gauge the artifacts perceived by humans introduced by the inpainting process into an image. This statistic is calculated as the ratio between the area covered by a binary mask of the regions subjected to inpainting n_m and the area identified by the PAL4Inpaint model as containing artifacts n_a:

$$PAR = \frac{n_a}{n_m}.$$

Both of these areas are expressed in terms of the number of pixels they encompass. Images of an example binary mask highlighting the marked artifacts region and the corresponding binary mask utilized for inpainting can be seen in Fig. 4.4.

To avoid PAL4Inpaint model from identifying artifacts of other faces in the image we give it only a small cropped area around the bounding box of the evaluated face. This cropped area can be seen in Fig. 4.3. There is a potential that PAR will be greater than one in case PAL4Inpaint identifies an area larger than the area on the inpainting mask.



(a) Highlighted Artifacts Region **(b) Artifacts Mask** **(c) Inner Face Mask**

Figure 4.4. Anonymized face with detected inpainting artifacts region highlighted (a), binary mask of the inpainting artifacts (b) and the the binary mask of the inner face (c).

4.1.5 Statistics Averaging

When analyzing an image containing multiple faces or multiple such images, we average the statistics computed for each individual face or face pair to get an overall statistic. This averaging is done differently for different types of statistics.

Boolean statistics are averaged by tallying the occurrences where the condition is true and dividing this count by the total number of instances. An example of such a statistic is evaluating whether both the original and anonymized faces share the same gender.

Orientation requires a distinct averaging method compared to other measured values because rotation matrices cannot be directly averaged element-wise. To average orientations, we convert each rotation matrix into Euler angles. Subsequently, we determine the absolute value of each angle and aggregate all angles expressing rotation around the same axis. Although this approach does not yield the correct average orientation, it effectively captures the absolute differences in facial orientations.

For all other statistics, a straightforward scalar mean is calculated.

4.2 Video Statistics

On top of the statistics used for evaluating images, which are used for analyzing all individual frames of the evaluated video, we introduce supplementary statistics aimed at assessing the behavior of the anonymization method across consecutive frames.

To enable evaluation across multiple frames, we implement an identity tracking mechanism. This involves maintaining a repository of `identity pairs`, which track the occurrences of each identity in both the original and anonymized videos. These pairs retain the identity vector of the first appearance of each identity in the original video, facilitating matching with subsequent appearances in subsequent frames.

During the evaluation of each new frame in both the original and anonymized videos, all faces are initially mapped together. Any incorrectly mapped faces are disregarded. Subsequently, the identity vectors of all correctly matched original faces are evaluated and compared with those in the existing identity pairs. Based on the cosine distance between these vectors, the identities in the new frame are either incorporated into the existing identity pairs or formed into new ones. This process is repeated for all frames in the video.

4.2.1 Identity Consistency

In assessing the consistency of an identity throughout all frames of a video, we compute the variation in identity vector angles. This calculation is performed for both the original and anonymized identities within all identity pairs present in the video.

The process begins by extracting identity vectors from every frame in which an identity is present. Subsequently, an element-wise median vector is derived from this collection of identity vectors. This median vector is then normalized, and the angles between it and all other identity vectors are determined. Finally, the variance is computed based on these angles.

4.2.2 Face Trajectory Correlation

To assess whether an identity in the anonymized video exhibits similar movement patterns to its counterpart in the original video, we compute what we term as *normalized landmarks correlation*. This metric is computed for each identity pair within the video.

To calculate the *normalized landmarks correlation*, we begin by extracting the positions of landmarks in every frame where both the original and anonymized faces are detected. This results in two arrays of landmarks for each frame, with each array containing 98 distinct landmarks, each defined by x and y coordinates. Subsequently, we iterate through every coordinate of every landmark, extracting the corresponding coordinate from all frames in which the identity is present, for both the original and anonymized faces. This process yields two arrays of scalar values. We then compute the normalized cross-correlation of these arrays. After computing the normalized cross-correlation for each coordinate of every landmark, we average the results to obtain the *normalized landmarks correlation*.

Chapter 5 Experiments

A series of experiments has been conducted on both images and videos to assess the effectiveness of our proposed anonymization method. Initially, we conducted comparisons between our anonymization method and an established anonymization toolkit, DeepPrivacy 2 [7] with the FDF256 settings. Subsequently, we delved into an investigation aimed at uncovering the impacts of various parameters of Stable Diffusion on anonymization quality. This was done by using our base set of parameters as a starting point and then always changing one parameter to see its effects. The base parameters are listed in Appendix A. If it is not specified directly the base set of parameters is denoted by "*" in the following sections. Throughout these experiments, our benchmarking suite served as the primary tool for quantifying anonymization quality.

5.1 Images

This section is dedicated to assessing the performance of both our method and DeepPrivacy 2 exclusively on images. A dataset comprising 128 images of babies was utilized as the foundation for all ensuing experiments. This dataset was curated by sourcing images from the internet, with a deliberate effort made to ensure diversity to test the anonymizers under varied conditions.

Besides the experiments described in the following subsections we have conducted a few additional ones. Unfortunately these experiments did not lead to as conclusive results, so we decided not to include them in the main text of the thesis. But comparisons of images resulting from these experiments can be seen in Appendix B.

5.1.1 Method Comparison

In this experiment we compare the results of our anonymization method, using the base set of parameters, to the results of DeepPrivacy 2 and a set of 128 random image pairs taken from our dataset. An example of a few of these random pairs can be seen in Fig. 5.1. The random image pairs are meant to serve as a baseline for comparing the quality of the anonymization methods. Values measured on these pairs can serve as attainable higher or lower bounds of some of the measured metrics.

Results of our method, using the base parameters, and DeepPrivacy 2, illustrated on a few images picked from the dataset, can be seen in Fig. 5.2. From these images it is apparent that neither method is perfect. Both of them sometimes introduce artifacts into the anonymized images. Especially when it comes to anonymization of faces that are not oriented directly towards the camera. It is also visible that DeepPrivacy 2 sometimes struggles with generating realistic shadows. Another aspect in which our method seems to perform better is the age of the anonymized faces. This is thanks to the positive prompt we use. If we would not specify to Stable Diffusion that it should inpaint faces of babies it would most likely also struggle with age. As can partially be seen in Section 5.1.2, where we test the performance without a positive prompt.

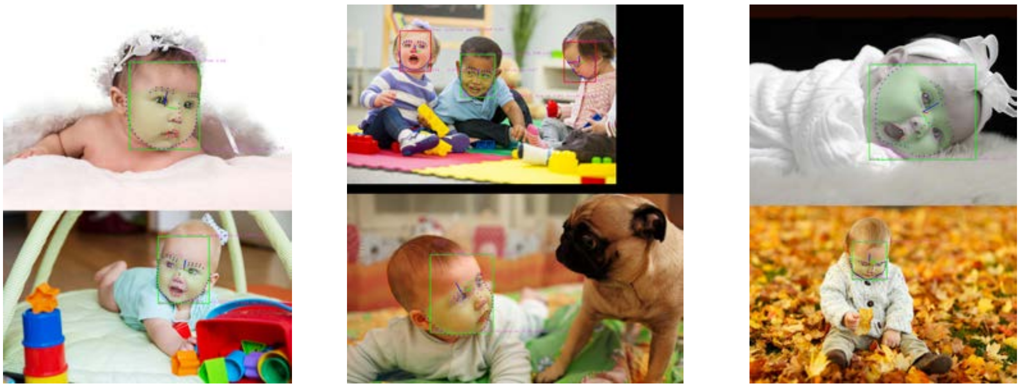


Figure 5.1. A few of the random image pairs from our dataset of children faces, that were used to define a baseline to compare our method to. The image pairs are stacked on top of each other and they have highlighted mapping and some additional metrics.

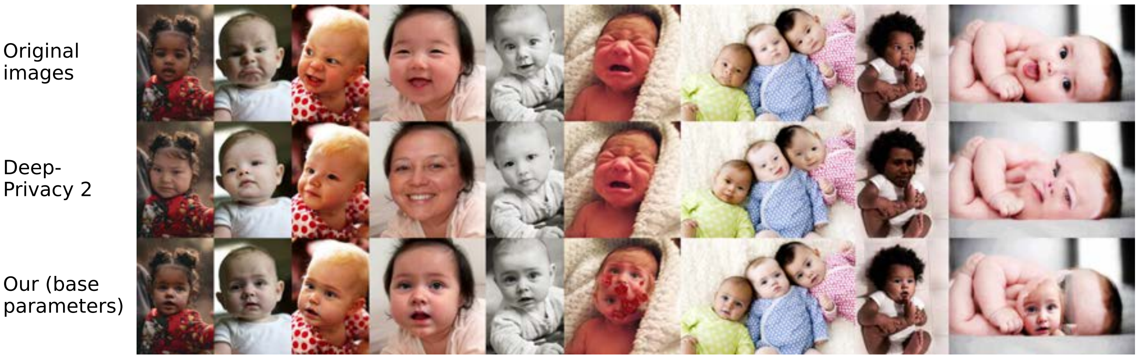


Figure 5.2. A few images anonymized by our method using the base parameters and DeepPrivacy 2 compared to the original images.

The statistics of mapping between the original and anonymized images of the random image pairs and the compared methods can be seen in Tab. 5.1. Related to the mapping average confidence of the face detector on the original and anonymized images can be seen in Tab. 5.2. From these tables we can see, that our method anonymizes faces in way, where they are more likely to be detected by our face detector.

Method	Total Detected Faces	Total Correctly Mapped Faces	Total Incorrectly Mapped Faces
Random Image Pairs	169	79	90
DeepPrivacy 2	199	183	16
Our (Base Parameters)	199	190	9

Table 5.1. Comparison of mapping between random image pairs, DeepPrivacy 2 and our method using base parameters.

De-identification quality comparison between the two methods and the random image pairs can be seen in Tab. 5.3. In this cathegory DeepPrivacy 2 performs way better than our method. Both when it comes to determining whether the original and anonymized faces belong to the same identity and the cosine distance between the two identities. It is interesting that so many of the random image pairs were considered to be belong to the same identity. It could be that DeepFace has problems finding the identity vector

Method	Original Detector Confidence	Anonymized Detector Confidence	Detector Confidence Difference
Random Image Pairs	0.736 ± 0.075	0.718 ± 0.086	0.090 ± 0.070
DeepPrivacy 2	0.746 ± 0.071	0.747 ± 0.070	0.034 ± 0.038
Our (Base Parameters)	0.742 ± 0.076	0.749 ± 0.072	0.025 ± 0.029

Table 5.2. Comparison of detector confidence between random image pairs, DeepPrivacy 2 and our method using base parameters.

Method	Same Identity Ratio	Cosine Distance
Random Image Pairs	50.63%	0.70 ± 0.14
DeepPrivacy 2	70.49%	0.55 ± 0.24
Our (Base Parameters)	90.53%	0.45 ± 0.17

Table 5.3. Comparison of de-identification quality between random image pairs, DeepPrivacy 2 and our method using base parameters.

of infants, or its threshold for deciding when the identities are different is too low for them.

The comparison of anonymization artifacts introduced by the two methods can be seen in Tab. 5.4. Here DeepPrivacy 2 also seems to perform better than our method, even though there are many visible artifacts in the images comparing the methods, shown in Fig. 5.2. That could be caused by multiple different factors. It could be that we are using our mask of inner face for evaluating PAR meanwhile DeepPrivacy 2 uses a different for its inpainting. Or PAL4Inpaint could be having trouble identifying the types of artifacts DeepPrivacy 2 introduces. This is quite probable because when we look at the values corresponding to the original images, PAL4Inpaint identified almost as many artifacts in them as it did in the anonymized ones.

Method	Original PAR	Anonymized PAR	PAR Difference
Random Image Pairs	0.19 ± 0.24	0.23 ± 0.25	0.24 ± 0.22
DeepPrivacy 2	0.17 ± 0.22	0.17 ± 0.23	0.09 ± 0.15
Our (Base Parameters)	0.17 ± 0.22	0.25 ± 0.25	0.11 ± 0.15

Table 5.4. Comparison of anonymization artifacts between random image pairs, DeepPrivacy 2 and our method using base parameters.

The comparison of mouth an eye openness between the two methods can be seen in Tab. 5.5 and Tab. 5.6 respectively. For both the eyes and mouth our method seems to preserve their opennness a bit better. Although it seems to have a tendency to make the eyes in anonymized images more open than in the original ones. DeepPrivacy 2 seems to have an opposite effect when it comes to mouth openness, so original faces with open mouths tend to have them closed when anonymized.

Comparison of gender, race and emotion preservation between the two methods can be seen in Tab. 5.7. From the measured values it seems like both our method and DeepPrivacy 2 perform similarly when it comes to keeping the gender and race on the anonymized infants. DeepPrivacy 2 performs slightly worse when it comes to preserving

Method	Original Mouth Openness	Anonymized Mouth Openness	Mouth Openness Difference
Random Image Pairs	0.29 ± 0.23	0.24 ± 0.24	0.27 ± 0.21
DeepPrivacy 2	0.25 ± 0.21	0.15 ± 0.13	0.16 ± 0.15
Our (Base Parameters)	0.25 ± 0.22	0.27 ± 0.18	0.10 ± 0.12

Table 5.5. Comparison of mouth openness between random image pairs, DeepPrivacy 2 and our method using base parameters.

Method	Average Orig. Eye Openness	Average Anon. Eye Openness	Average Eye Openness Diff.
Random Image Pairs	0.35 ± 0.16	0.35 ± 0.15	0.19 ± 0.14
DeepPrivacy 2	0.39 ± 0.16	0.39 ± 0.11	0.12 ± 0.10
Our (Base Parameters)	0.39 ± 0.16	0.44 ± 0.09	0.11 ± 0.11

Table 5.6. Comparison of average eye openness between random image pairs, DeepPrivacy 2 and our method using base parameters.

Method	Same Gender Ratio	Same Race Ratio	Same Emotion Ratio
Random Image Pairs	49.37%	43.04%	21.52%
DeepPrivacy 2	63.39%	57.92%	27.87%
Our (Base Parameters)	62.63%	56.84%	32.11%

Table 5.7. Comparison of gender, race and emotion consistency between random image pairs, DeepPrivacy 2 and our method using base parameters.

emotions. But it is quite questionable if DeepFace is able to correctly identify these values on infants, as even for humans it can be hard to distinguish them. Especially when it comes to babies.

The comparison of age preservation between the two methods can be seen in Tab. 5.8. It is clear that these age estimates are wrong, as none of the babies should be older than 10 years. This means that DeepFace is unable to correctly estimate the age of babies, and so we will refrain for using these values for comparison in the other experiments.

Method	Age Difference	Original Age	Anonymized Age
Random Image Pairs	6.6 ± 5.8	27.7 ± 6.8	29.1 ± 6.1
DeepPrivacy 2	5.1 ± 5.7	27.2 ± 7.2	25.4 ± 7.4
Our (Base Parameters)	4.6 ± 4.9	27.1 ± 7.1	24.6 ± 5.1

Table 5.8. Comparison of average age between random image pairs, DeepPrivacy 2 and our method using base parameters.

Comparison of average orientation differences between the original and anonymized faces, generated by our method and DeepPrivacy 2, can be seen in Tab. 5.9. As described in Chapter ?? these orientation differences are not exact averages, but they show how well the methods preserve orientation when it comes to rotations around

Method	Orientation Difference Z-axis Angle [rad]	Orientation Difference Y-axis Angle [rad]	Orientation Difference X-axis Angle [rad].
Random Image Pairs	0.24 ± 0.22	0.18 ± 0.16	0.46 ± 0.44
DeepPrivacy 2	0.08 ± 0.10	0.11 ± 0.11	0.13 ± 0.30
Our (Base Parameters)	0.05 ± 0.10	0.053 ± 0.059	0.07 ± 0.19

Table 5.9. Comparison of average orientation difference between random image pairs, DeepPrivacy 2 and our method using base parameters. The orientation is in the format of Euler angles where the order of axes of rotation is z, y and x.

different axes. We can see from the table that our method outperforms DeepPrivacy 2, when it comes to orientation preservation.

5.1.2 Prompts

In this experiment we compare the effects of positive and negative text prompts on our method. In total four different measurements were conducted as part of this experiment. They were:

- **Full Positive Prompt:** This measurement used the base set of parameters where both the positive and negative prompts were unchanged.
- **No Celebrities**: In this measurement the celebrity names were removed from the positive prompt. The negative prompt was unchanged.
- **No Positive Prompt**: In this measurement the whole positive prompt was removed. The negative prompt was unchanged.
- **No Negative Prompt**: In this measurement the positive prompt was unchanged from the base set of parameters. The negative prompt was completely removed.

The used positive prompt:

```
(Daryl Sabara: 0.2), (Macaulay Culkin: 0.1), (Thomas Sangster: 0.1),
(Kelly Macdonald: 0.1), (Taylor Swift: 0.2), (Sydney Sweeney: 0.2),
(photo of a little baby face: 1.2)
```

The positive prompt has 2 parts. First part consists of 6 celebrity names. Then the second part describes what we want to inpaint into the image, which is a face of a baby. Each part of the prompt has a number next to it, that represents the weight of its accompanying part. Based on these weights Stable Diffusion will follow each part of the prompt more, or less closely.

The negative prompt can be seen in the Appendix A. We based it on the suggested negative prompt of our main used model `Realistic Vision V6.0 B1`.

Comparison of images anonymized using different positive and negative prompts can be seen in Fig. 5.3. It is clear from the images, that the main part of the positive prompt, has a big effect on the anonymization quality. On the other hand the absence of celebrity names in the positive prompts is almost unnoticable. Similarly it is visible, that without the use of the negative prompt, the amount of anonymization artifacts rises.

Evaluation of matching between the original images and the ones anonymized using different CFG scales can be seen in Tab. 5.10. There do not seem to be great differences in the how many of the anonymized faces can be detected.

Comparison of de-identification quality between the anonymizations using different text prompts can be seen in Tab. 5.11. The use of celebrities in the positive prompt has

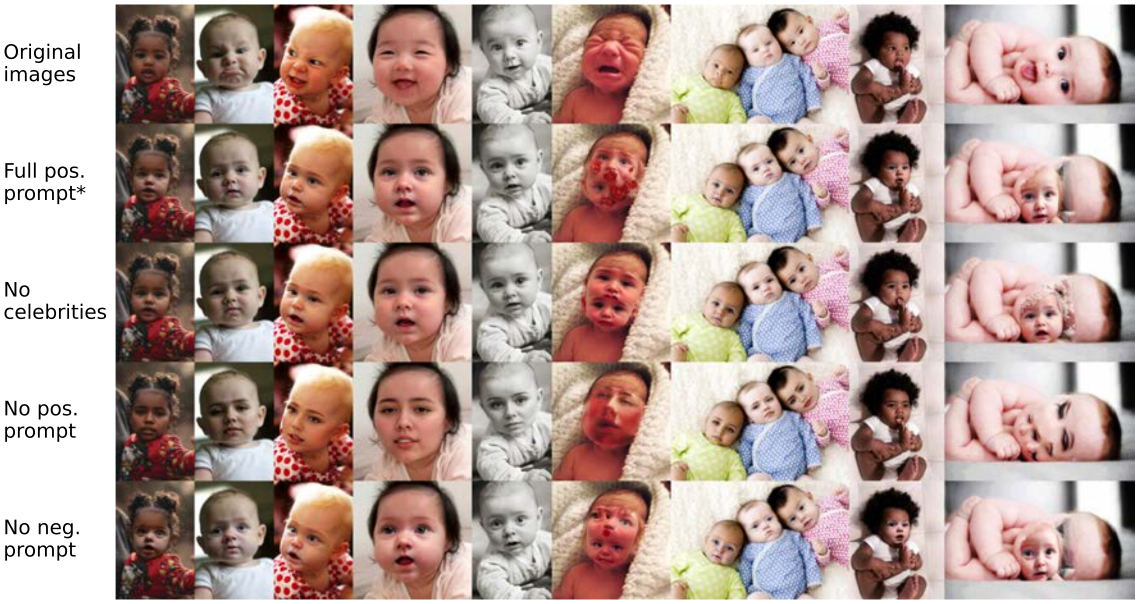


Figure 5.3. Comparison of images anonymized using our method with different positive and nagative text prompts.

Used Prompts	Total Correctly Mapped Faces	Correctly Mapped Faces Ratio	Total Incorrectly Mapped Faces
Full Pos. Prompt*	190	95.48%	9
No Celebrities	191	95.98%	8
No Pos. Prompt	189	94.97%	10
No Neg. Prompt	189	94.97%	10

Table 5.10. Comparison of mapping for different prompts used for conditioning our method.

almost no effect on the de-identification quality. Unlike the main part of the positive prompt. Omitting it seems so greatly increase the de-identification quality, but at the const of how realistic looking the anonymized faces are. Surprisingly not using the negative prompt slightly decreases the de-identification quality. It would be logical that the artifacts caused by the absence of the negative would increase the de-identification quality, but maybe the DeepFace is able to ignore them.

Used Prompts	Same Identity Ratio	Cosine Distance
Full Pos. Prompt*	90.53%	0.45 ± 0.17
No Celebrities	90.58%	0.45 ± 0.19
No Pos. Prompt	66.67%	0.56 ± 0.24
No Neg. Prompt	92.06%	0.47 ± 0.17

Table 5.11. Comparison of de-identification quality for different text prompts.

Comparison of anonymization artifacts between the anonymizations using different CFG scales can be seen in Tab. 5.12. From the measured values it seems like the different prompts do not influence the amount of introduced artifacts. This is suspicious as it would be expected, that the absence of negative prompt, would introduce more artifacts into the anonymized images. This effect is even slightly visible in the example images in Fig. 5.3. What could be causing this is the inability of PAL4Inpaint to detect this type of introduced artifacts.

Used Prompts	Original *PAR*	Anonymized *PAR*	*PAR* Difference
Full Pos. Prompt*	0.17 ± 0.22	0.25 ± 0.25	0.11 ± 0.15
No Celebrities	0.16 ± 0.22	0.26 ± 0.26	0.11 ± 0.16
No Pos. Prompt	0.16 ± 0.22	0.26 ± 0.28	0.12 ± 0.19
No Neg. Prompt	0.17 ± 0.22	0.26 ± 0.27	0.11 ± 0.17

Table 5.12. Comparison of anonymization artifacts for different text prompts.

Comparison of differences in mouth and eye openness and in the norm of gaze vector between the anonymizations using different CFG scales can be seen in Tab. 5.13. Similarly to the previously mentioned anonymization artifacts, the different prompts do not seem to effect them.

Used Prompts	Mouth Openness Difference	Average Eye Openness Diff.	Average Gaze Difference Norm
Full Pos. Prompt*	0.10 ± 0.12	0.11 ± 0.11	0.32 ± 0.53
No Celebrities	0.11 ± 0.13	0.12 ± 0.11	0.28 ± 0.33
No Pos. Prompt	0.12 ± 0.14	0.11 ± 0.10	0.30 ± 0.34
No Neg. Prompt	0.13 ± 0.12	0.12 ± 0.12	0.28 ± 0.32

Table 5.13. Comparison of average mouth and eye openness and haze norm for different text prompts.

5.1.3 Denoising Strength

In this experiment we focus on the effects of a parameter of Stable Diffusion called **denoising strength**. It controlls how aggresive is the denoising process. This allows to control how similar to the original image will the anonymized one be. With lower values of denoising strength increasing the similarity and higher values decreasing it.

Comparison of a few images anonymized by our method utilizing different denoising strengths can be seen in Fig. 5.4. From these images one can see that the lower the denoising strength is the more similar the anonyized image is to its original one. Another noticeable characteristic is that anonymizations with higher denoising strengths tend to have more difference in facial attributes. From these two observations we can deduce that there will always be a trade-off between the de-identification quality and the preservation of facial features. From the values corresponding to the de-identification quality, shown in Tab. 5.15, we can see the same trend as is visible in the images and that is that the higher the denoising strength the more is the identity changed..

5.1.4 Mask Padding

In this experiment we evaluate the effects of a parameter of Stable Diffusion we call **mask padding**. It influences the region Stable Diffusion takes into consideration when

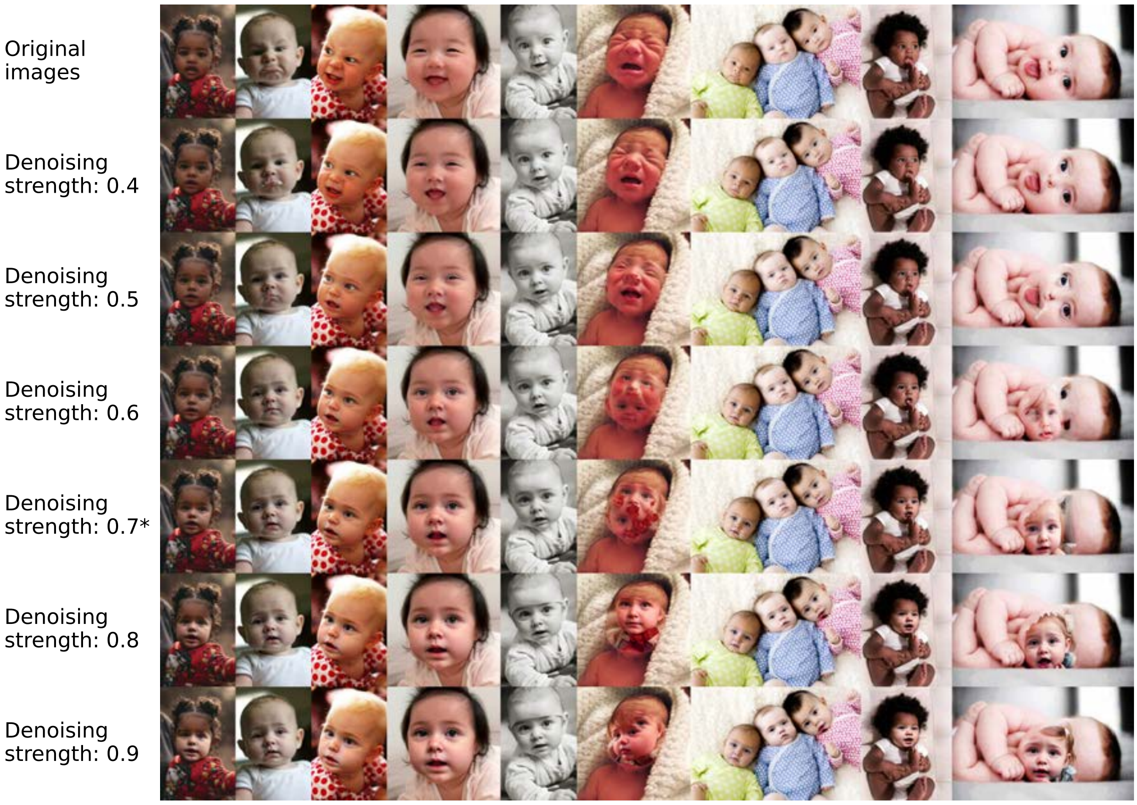


Figure 5.4. Comparison between images anonymized using different denoising strengths.

Denoising Strength	Total Correctly Mapped Faces	Correctly Mapped Faces Ratio	Total Incorrectly Mapped Faces
0.4	195	97.99%	4
0.5	194	97.49%	5
0.6	191	95.98%	8
0.7*	190	95.48%	9
0.8	192	96.48%	7
0.9	188	94.47%	11

Table 5.14. Comparison of mapping for different values of denoising strength.

Denoising Strength	Same Identity Ratio	Cosine Distance
0.4	97.44%	0.30 ± 0.15
0.5	94.33%	0.38 ± 0.17
0.6	92.67%	0.43 ± 0.18
0.7*	90.53%	0.45 ± 0.17
0.8	90.10%	0.48 ± 0.19
0.9	88.83%	0.48 ± 0.18

Table 5.15. Comparison of de-identification quality for different values of denoising strength.

Denoising Strength	Original PAR	Anonymized PAR	PAR Difference
0.4	0.17 ± 0.22	0.22 ± 0.24	0.07 ± 0.11
0.5	0.17 ± 0.22	0.24 ± 0.25	0.09 ± 0.13
0.6	0.17 ± 0.22	0.24 ± 0.25	0.09 ± 0.14
0.7*	0.17 ± 0.22	0.25 ± 0.25	0.11 ± 0.15
0.8	0.17 ± 0.22	0.25 ± 0.25	0.11 ± 0.15
0.9	0.17 ± 0.22	0.25 ± 0.26	0.11 ± 0.16

Table 5.16. Comparison of anonymization artifacts for different values of denoising strength.

Denoising Strength	Detector Confidence Difference	Mouth Openness Difference	Average Eye Openness Diff.	Average Gaze Difference Norm
0.4	0.018 ± 0.019	0.06 ± 0.07	0.06 ± 0.08	0.27 ± 0.58
0.5	0.023 ± 0.031	0.08 ± 0.08	0.09 ± 0.14	0.27 ± 0.37
0.6	0.025 ± 0.033	0.09 ± 0.09	0.11 ± 0.15	0.30 ± 0.55
0.7*	0.025 ± 0.029	0.10 ± 0.12	0.11 ± 0.11	0.32 ± 0.53
0.8	0.025 ± 0.028	0.12 ± 0.13	0.12 ± 0.11	0.28 ± 0.34
0.9	0.027 ± 0.033	0.11 ± 0.14	0.12 ± 0.13	0.29 ± 0.35

Table 5.17. Comparison of denoising strength, mouth and eye openness and gaze norm for different values of denoising strength.

Denoising Strength	Same Gender Ratio	Same Race Ratio	Same Emotion Ratio
0.4	74.36%	67.69%	51.79%
0.5	68.04%	63.4%	38.66%
0.6	65.45%	57.59%	34.55%
0.7*	62.63%	56.84%	32.11%
0.8	55.73%	51.04%	26.56%
0.9	58.51%	52.66%	29.26%

Table 5.18. Comparison of gender, race and emotion consistency for different values of denoising strength.

inpainting an image, where the **mask padding** influences how many pixels will be add to each side of the masked region.

We have experimented with these two different ways of setting this parameter:

1. Constant Padding: A fixed value of mask padding has been you to generate all the anonymized images.
2. Padding Coefficient: The value of mask padding is set based on the size of the bounding box of each original face. Exact value of the mask padding p is the product of padding coefficient c_p and either the width b_w or the height b_h of the bounding box based on whichever is higher:

$$p = c_p \cdot max\left(b_w; b_h\right).$$

Denoising Strength	Orientation Difference Z-axis Angle [rad]	Orientation Difference Y-axis Angle [rad]	Orientation Difference X-axis Angle [rad].
0.4	0.030 ± 0.058	0.040 ± 0.061	0.06 ± 0.25
0.5	0.033 ± 0.049	0.048 ± 0.076	0.042 ± 0.097
0.6	0.034 ± 0.046	0.053 ± 0.068	0.06 ± 0.16
0.7*	0.05 ± 0.10	0.053 ± 0.059	0.07 ± 0.19
0.8	0.05 ± 0.11	0.056 ± 0.066	0.07 ± 0.19
0.9	0.053 ± 0.094	0.059 ± 0.068	0.08 ± 0.22

Table 5.19. Comparison of average orientation difference for different values of denoising strength. The orientation is in the format of Euler angles where the order of axes of rotation is z, y and x.

Comparison of faces, anonymized using both approaches to getting mask padding with different values, can be seen in Fig. 5.5. From the images it seems that setting mask padding using a padding coefficient results in less anonymization artifacts. It also seems to help with aligning of the anonymized face as can be seen in Tab. 5.22. Where anonymizations using padding based on the bounding box size performed way better. On the other hand it seems to perform worse when it comes to de-identification, as can be seen in Tab. 5.21. It makes sense that an adaptive value performs better than a fixed one as each image has a different resolution.

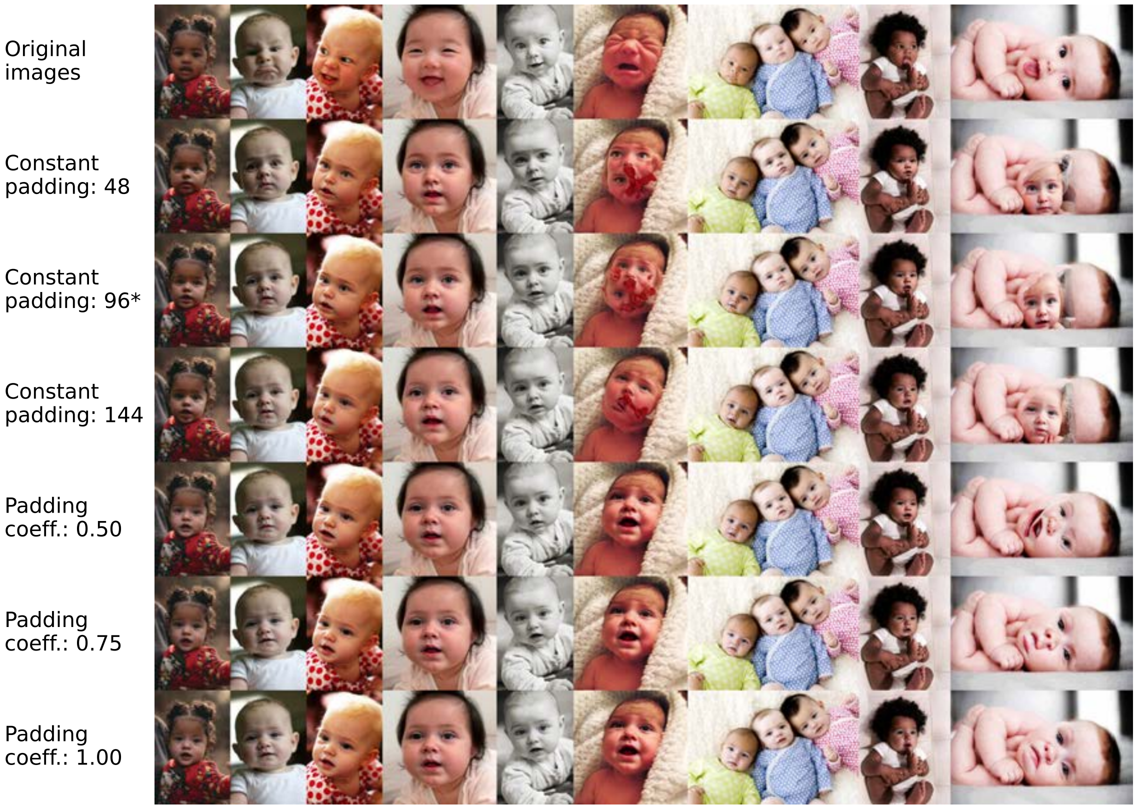


Figure 5.5. Comparison of images anonymized using different mask padding values.

Padding Type	Total Correctly Mapped Faces	Correctly Mapped Faces Ratio	Total Incorrectly Mapped Faces
Const. Padding 48	190	95.48%	9
Const. Padding 96*	190	95.48%	9
Const. Padding 144	189	94.97%	10
Padding Coeff. 0.50	188	94.47%	11
Padding Coeff. 0.75	189	94.97%	10
Padding Coeff. 1.00	191	95.98%	8

Table 5.20. Comparison of mapping for different mask paddings.

Padding Type	Same Identity Ratio	Cosine Distance
Const. Padding 48	86.32%	0.49 ± 0.19
Const. Padding 96*	90.53%	0.45 ± 0.17
Const. Padding 144	90.48%	0.45 ± 0.18
Padding Coeff. 0.50	93.09%	0.44 ± 0.18
Padding Coeff. 0.75	92.06%	0.42 ± 0.18
Padding Coeff. 1.00	93.72%	0.42 ± 0.17

Table 5.21. Comparison of de-identification for different mask paddings.

Padding Type	Orientation Difference Z-axis Angle [rad]	Orientation Difference Y-axis Angle [rad]	Orientation Difference X-axis Angle [rad].
Const. Padding 48	0.05 ± 0.13	0.054 ± 0.066	0.09 ± 0.27
Const. Padding 96*	0.05 ± 0.10	0.053 ± 0.059	0.07 ± 0.19
Const. Padding 144	0.040 ± 0.066	0.052 ± 0.061	0.07 ± 0.21
Padding Coeff. 0.50	0.037 ± 0.074	0.053 ± 0.055	0.05 ± 0.18
Padding Coeff. 0.75	0.037 ± 0.062	0.046 ± 0.048	0.06 ± 0.16
Padding Coeff. 1.00	0.034 ± 0.048	0.045 ± 0.048	0.05 ± 0.17

Table 5.22. Comparison of average orientation difference for different mask paddings. The orientation is in the format of Euler angles where the order of axes of rotation is z, y and x.

5.1.5 Classifier-Free Guidance

This experiment focused on assessing the influence of **Classifier-Free Guidance** scale on the quality of images anonymized using our method. The parameter influences how closely the inpainting process follows the text prompts.

Comparison of a few images anonymized using different CFG scales can be seen in Fig. 5.6. It is apparent from these images that higher values of CFG introduce more inpainting artifacts to the image. Especially he skin of anonymized faces looks very unnatural as if its contrast was increased.

5.1.6 ControlNet

This experiments focuses on evaluating the effects of the ControlNet extension on the anonymization quality of our method. ControlNet is a neural network used for conditioning the image generation of Stable Diffusion. In our case we use a model that conditions Stable Diffusion based on the posture of the person to whom the original

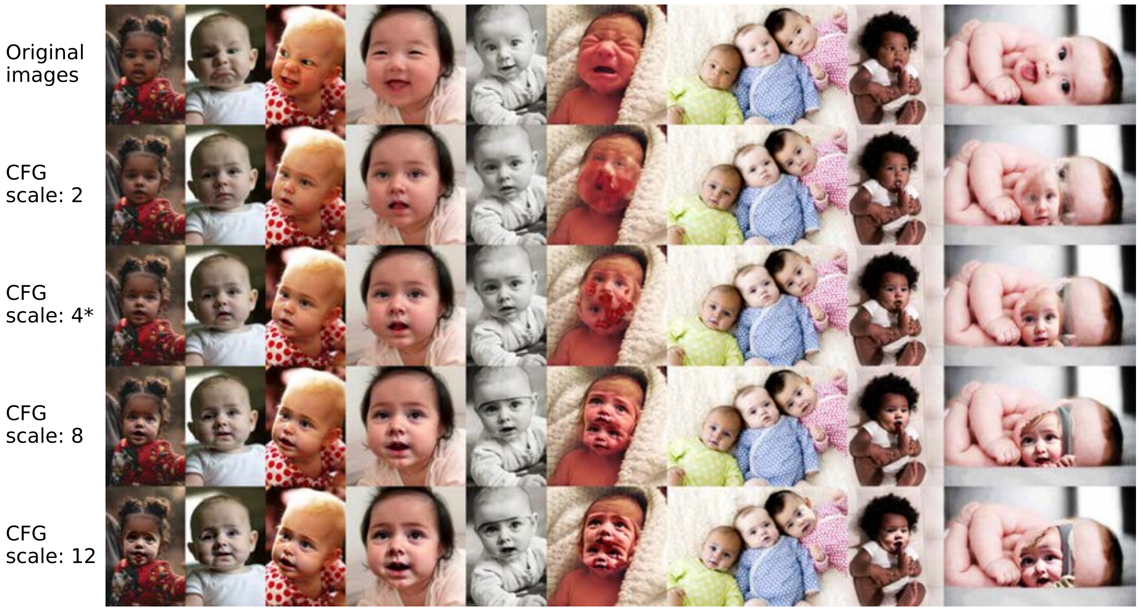


Figure 5.6. Comparison of images anonymized using different Classifier-Free Guidance scales.

Classifier-Free Guidance Scale	Total Correctly Mapped Faces	Correctly Mapped Faces Ratio	Total Incorrectly Mapped Faces
2	190	95.48%	9
4*	190	95.48%	9
8	187	93.97%	12
12	191	95.98%	8

Table 5.23. Comparison of mapping for different Classifier-Free Guidance scales.

face belongs to. This should mean that the anonymized face matches the original one more closely.

Basic statistics from this experiment can be seen in Tab. 5.24. Comparison of a few images anonymized using or not using ControlNet can be seen in Fig. 5.6. From the images it seems like ControlNet greatly helps Stable Diffusion with aligning the anonymized face with the original one. As the data regarding the orientation, that can be seen in Tab. 5.26 differences also suggests. It also seems to have a negative effect on quality of de-identification, as can be seen in Tab. 5.25.

ControlNet State	Total Correctly Mapped Faces	Correctly Mapped Faces Ratio	Total Incorrectly Mapped Faces
On*	190	95.48%	9
Off	187	93.97%	12

Table 5.24. Comparison of mapping based on the use of ControlNet.

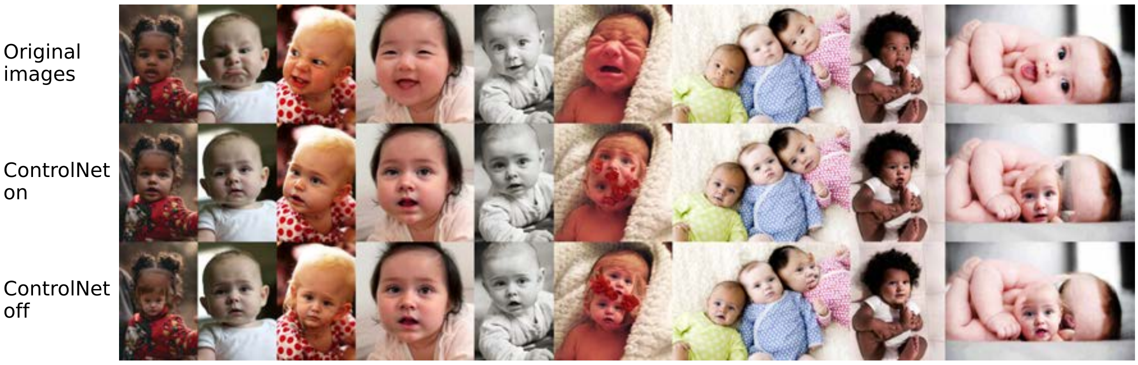


Figure 5.7. Comparison of images anonymized using or not using ControlNet.

ControlNet State	Same Identity Ratio	Cosine Distance
On*	90.53%	0.45 ± 0.17
Off	72.73%	0.55 ± 0.21

Table 5.25. Comparison of de-identification quality based on the use of ControlNet.

ControlNet State	Orientation Difference Z-axis Angle [rad]	Orientation Difference Y-axis Angle [rad]	Orientation Difference X-axis Angle [rad].
On*	0.05 ± 0.10	0.053 ± 0.059	0.07 ± 0.19
Off	0.12 ± 0.16	0.11 ± 0.12	0.17 ± 0.34

Table 5.26. Comparison of average orientation difference based on the use of ControlNet. The orientation is in the format of Euler angles where the order of axes of rotation is z, y and x.

5.2 Videos

The experiments described in this section focus on comparing the anonymization quality of videos anonymized by our method and by DeepPrivacy 2. Experiments were conducted on 3 videos in total. Each video contained a single identity for easier evaluation. From each video only the first 480 frames were evaluated.

Besides the method comparison we also try to show the effects of two different approaches used for increasing the the anonymization quality, namely:

- ControlNet: An extension of Stable Diffusion used for conditioning of inpainting. It is used in our method to help correctly align the inpainted face with the original one. In the videos this should manifest by higher correlation between the trajectories of original and anonymized landmarks.
- Celebrity Names in Positive Prompt: As a part of the positive prompt we include six celebrity names. These are supposed to help keep a consistent identity when anonymizing faces. This should keep the identity in a video more consistent across different frames.

To measure the effects of these two approaches we anonymized each video using them as they are included in the base parameters. And we also anonymized each video first without one and then without the other.

5.2.1 First video

A few frames from the original first video and its anonymized versions can be seen in Fig. 5.8. From these few frames alone it can be seen that not all the anonymizations turned out well. This was probably caused by the rotation of the face relative to the camera. Especially our method without the use of ControlNet seems to have had problems with correctly aligning the face.

The measured statistics of the first video can be seen in Tab. 5.27. Normalized landmarks correlation confirms what can be seen in the example frames and that is that without the use ControlNet our method poorly matches the anonymized face to its original source. DeepPrivacy 2 also seems to have somewhat struggled with correctly aligning the anonymized face in this video. On the other hand from the anonymized identity variance it seems that DeepPrivacy 2 has been anonymizing the face to the same identity the most consistently. It is quite surprising considering the not very human looking faces in the example frames.

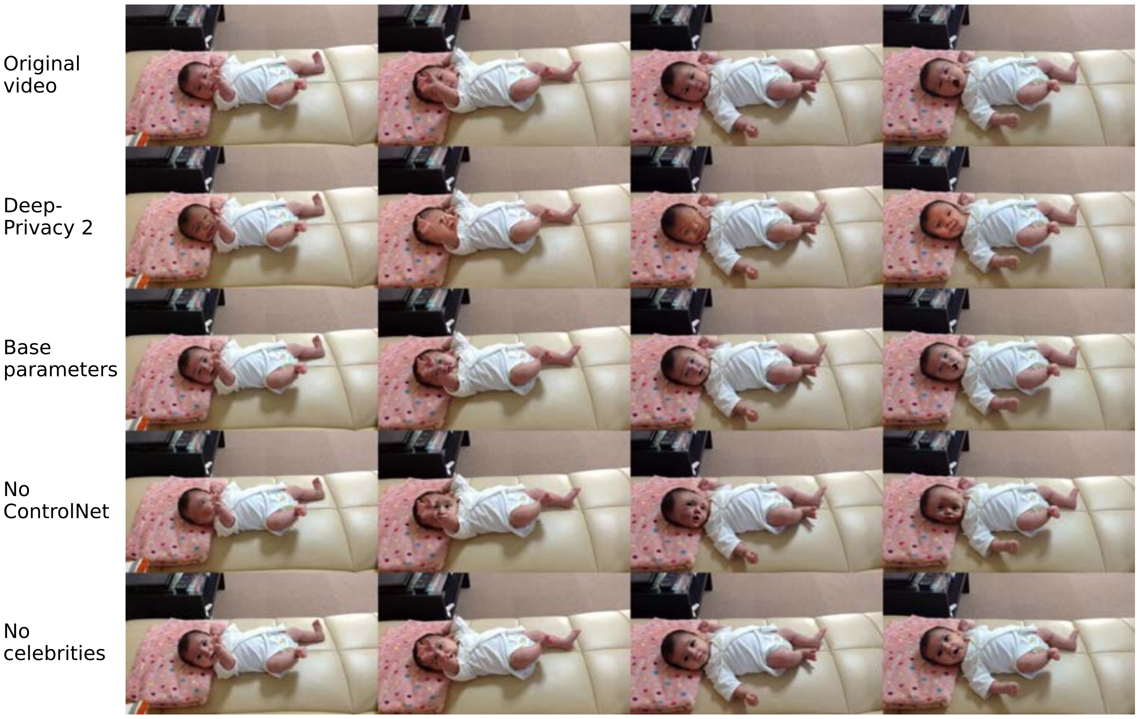


Figure 5.8. Comparison of a few frames from the first video.

5.2.2 Second Video

A few frames from the original second video and its anonymized versions can be seen in Fig. 5.9. The measured statistics of the second video can be seen in Tab. 5.28. The results are similar to the first video. ControlNet again seems to greatly help with the face alignment. Even though not as might as in the last example. This is most likely due to the face being upright and not sideways. All the methods again somehow keep the identity more consistent than in the original video.

Anonymization method	DeepPrivacy 2	Base parameters	No ControlNet	No celebrities
Original identity variance	0.0412	0.0412	0.0412	0.0412
Anonymized identity variance	0.0098	0.0398	0.0201	0.0346
Normalized landmarks correlation	0.8741	0.9716	0.7303	0.9705

Table 5.27. Evaluation of the anonymization quality of different methods. The measurements are done on the only identity in the first anonymized video.

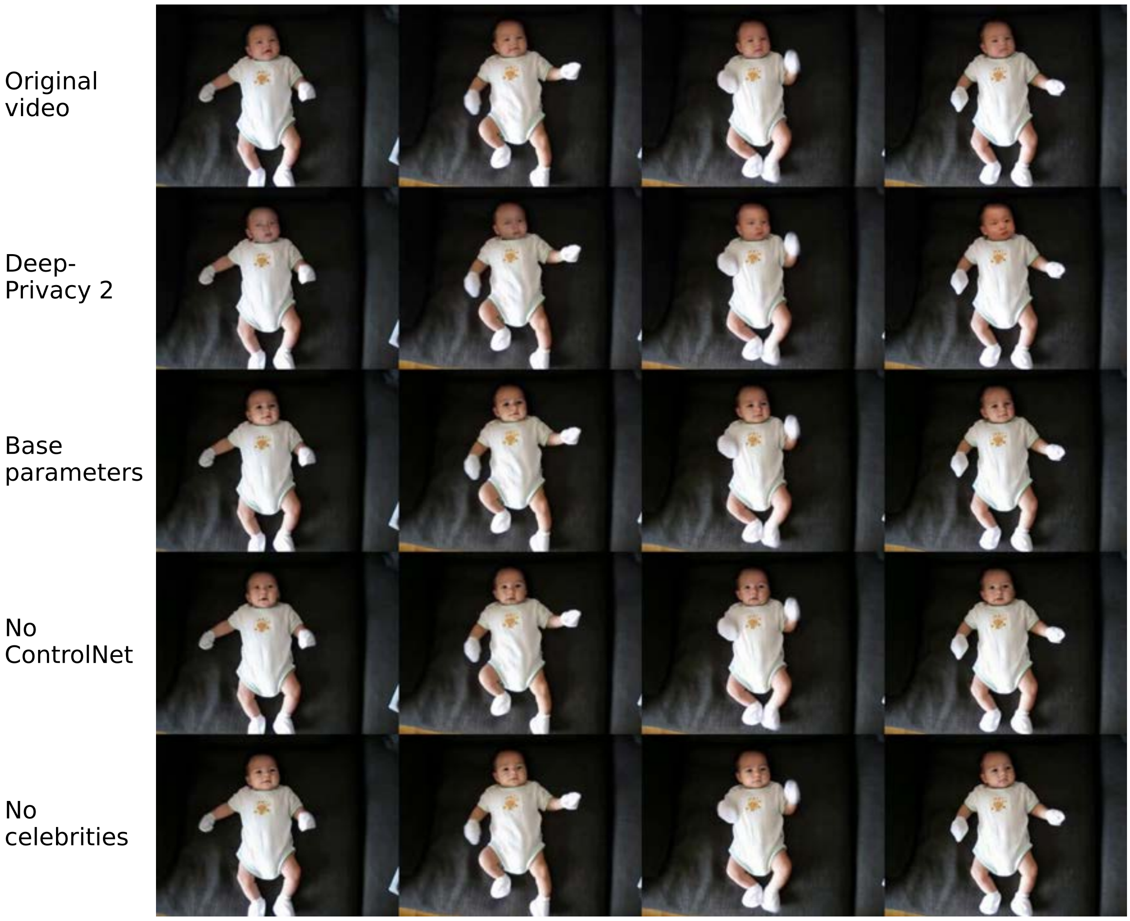


Figure 5.9. Comparison of a few frames from the second video.

5.2.3 Third Video

A few frames from the original third video and its anonymized versions can be seen in Fig. 5.10. The measured statistics of the third video can be seen in Tab. 5.29. When it comes to landmark correlation of this video all methods except the one not using ControlNet perform better than in the previous examples.

Anonymization method	DeepPrivacy 2	Base parameters	No ControlNet	No celebrities
Original identity variance	0.0281	0.0281	0.0281	0.0281
Anonymized identity variance	0.0144	0.0113	0.0131	0.0057
Normalized landmarks correlation	0.9322	0.9754	0.8559	0.9718

Table 5.28. Evaluation of the anonymization quality of different methods. The measurements are done on the only identity in the second anonymized video.

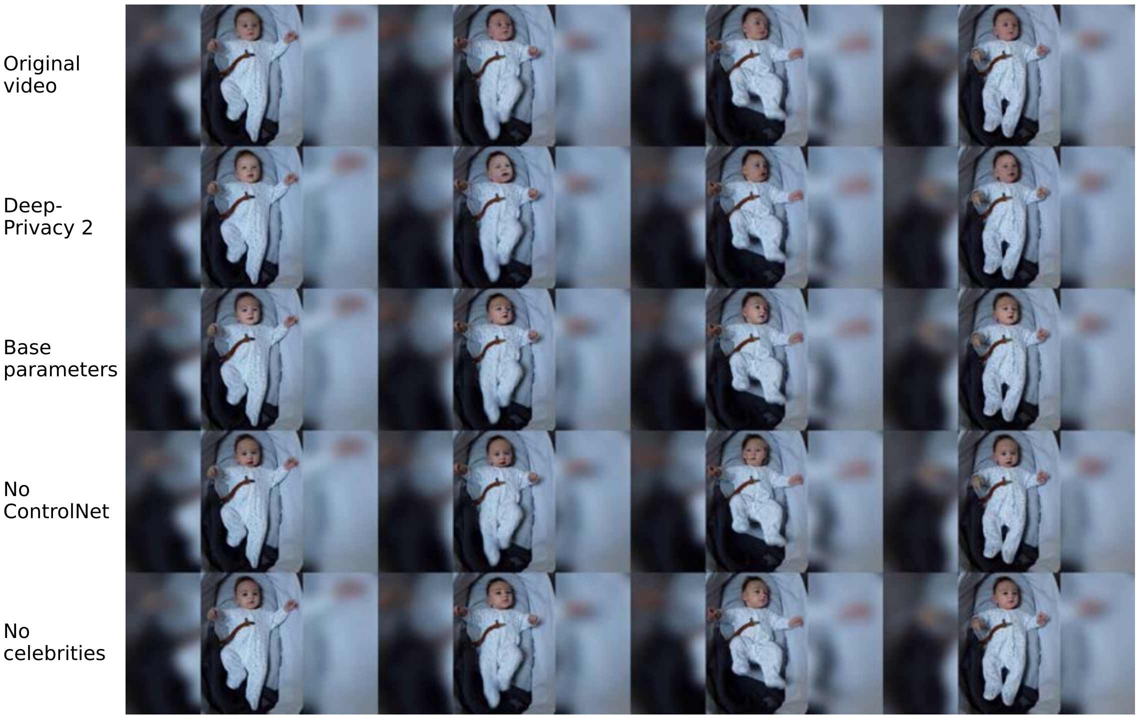


Figure 5.10. Comparison of a few frames from the third video.

Anonymization method	DeepPrivacy 2	Base parameters	No ControlNet	No celebrities
Original identity variance	0.0171	0.0171	0.0171	0.0171
Anonymized identity variance	0.0232	0.0116	0.0287	0.0135
Normalized landmarks correlation	0.9467	0.9794	0.5324	0.9793

Table 5.29. Evaluation of the anonymization quality of different methods. The measurements are done on the only identity in the third anonymized video.

5.3 User Study

To confirm how well Perceptual Artifact Ratio corresponds with the human perception of anonymization artifacts and to see how humans would react to the anonymized images, we decided to conduct a user study focused on the human perception of anonymization artifacts.

There were 2 versions of a questionnaire. Both of them consisted of 10 images. 5 of these 10 images were original unedited images and the other 5 were anonymized. In both version of the questionnaire the original images were the same. But the questionnaires differed in the 5 anonymized images. Even though the source images were the same in both versions, in one version they were anonymized using DeepPrivacy 2 toolbox and in the other they were anonymized using our method with the base parameters. The order in which the images were presented was randomly determined when creating the questionnaire and it was the same in both versions of the study.

In both versions we initially explained the goal of the study and what inpainting artifacts are, so the participants know what to look for. Then the participants were asked the same question about all the images. That question was "Is this image anonymized (edited)?" At the end of both versions of the questionnaire the participants were asked to rank their performance at distinguishing the anonymized images and how discomfortable they were when looking at the anonymization artifacts.

5.3.1 Overall Results

The overall results of the user study can be seen in Tab. 5.30. The accuracy metrics describe how often were the participants correct in their judgement of each image. Further details can be seen in Section 5.3.2 and Section 5.3.3.

Difficulty score is the average value from all the responses where participants evaluated the difficulty of distinguishing anonymized images. They were supposed to rank the difficulty on the scale from 1 to 5, where 1 means that distinguishing them was very easy and 5 means that it was very hard. Further details can be seen in Section 5.3.4.

Discomfort score is the average value from all the responses where participants evaluated how discomfortable they were looking at the anonymization artifacts. Participants ranked their dicomfort on a scale from 1 to 5, where 1 means they were not bothered by looking at the artifacts and 5 means that it was very hard for them to look. Further details can be seen in Section 5.3.5.

Anonymization Method	Our (Base Parameters)	DeepPrivacy 2
Total Respondents	29	27
Overall Accuracy	70.00%	69.00%
Original Accuracy	68.97%	69.63%
Anonymized Accuracy	71.03%	67.41%
Difficulty Score	3.34 ± 1.09	2.63 ± 0.91
Discomfort Score	3.17 ± 1.26	3.22 ± 0.96

Table 5.30. Overall results of the user study.

The participants were a little worse at distinguishing images anonymized using DeepPrivacy 2. Even though they felt like it was easier for them to distinguish these anonymized images. But the overall results of the study seem quite inconclusive.

5.3.2 Original Images

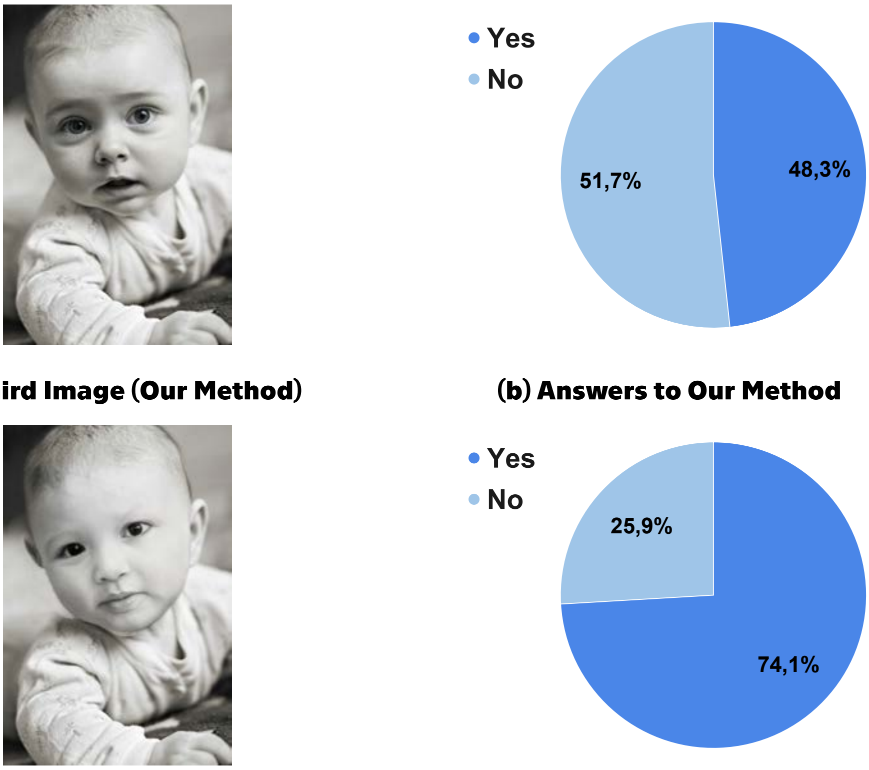
The distribution of answers to the question “Is this image anonymized (edited)?” regarding the original images in the study can be seen in Appendix C. The correct answer to the question in this case was “No.”

All the original images have the distribution of answers quite similar inbetween the two version of the study. Except maybe the fourth image. But the difference still is not significant enough to draw any conclusions. Overall the participants have been able to distinguish that the images are not anonymized.

5.3.3 Anonymized Images

Compared to the distributions of responses to the original images, the distributions of responses to the anonymized images differ more between both versions of the study.

The distribution of responses to the third image in both studies and the first anonymized one can be seen together with each image in Fig. 5.11. From the results it seems that the version of the image anonymized by DeepPrivacy 2 was easier to as being anonymized.



(a) Third Image (Our Method)

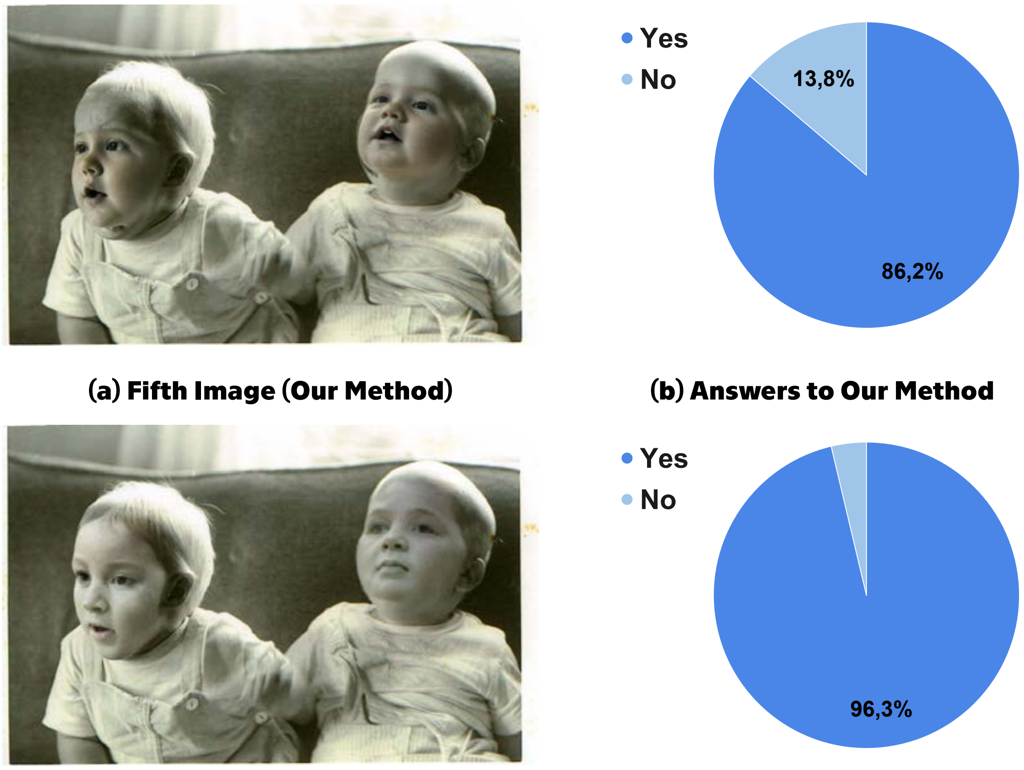
(b) Answers to Our Method

(c) Third Image (DeepPrivacy 2)

(d) Answers to DeepPrivacy 2

Figure 5.11. Comparison of the results of the third image of conducted user studies. The particapants were asked “Is this image anonymized (edited)?”. Third image in the version of the study using our method of anonymization (a) and corresponding distribution of answers (b). Third image in the version of the study using DeepPrivacy 2 for anonymization (c) and corresponding distribution of answers (d).

The fifth images and the second anonymized ones can be seen together with their response distributions in Fig. 5.12. This image seemed to have been distinguished as anonymized very easily in both versions of the study.



(a) Fifth Image (Our Method)

(b) Answers to Our Method

(c) Fifth Image (DeepPrivacy 2)

(d) Answers to DeepPrivacy 2

Figure 5.12. Comparison of the results of the fifth image of conducted user studies. The particapants were asked “Is this image anonymized (edited)?”. Fifth image in the version of the study using our method of anonymization (a) and corresponding distribution of answers (b). Fifth image in the version of the study using DeepPrivacy 2 for anonymization (c) and corresponding distribution of answers (d).

The sixth images used in the two versions of the study can be seen together with their response distributions in Fig. 5.13. DeepPrivacy 2 has performed way better on this image. Tricking almost all participants into believing that it was not edited. The image anonymized by our method on the other hand contained artifacts on the edge of the face, which most participants spotted.

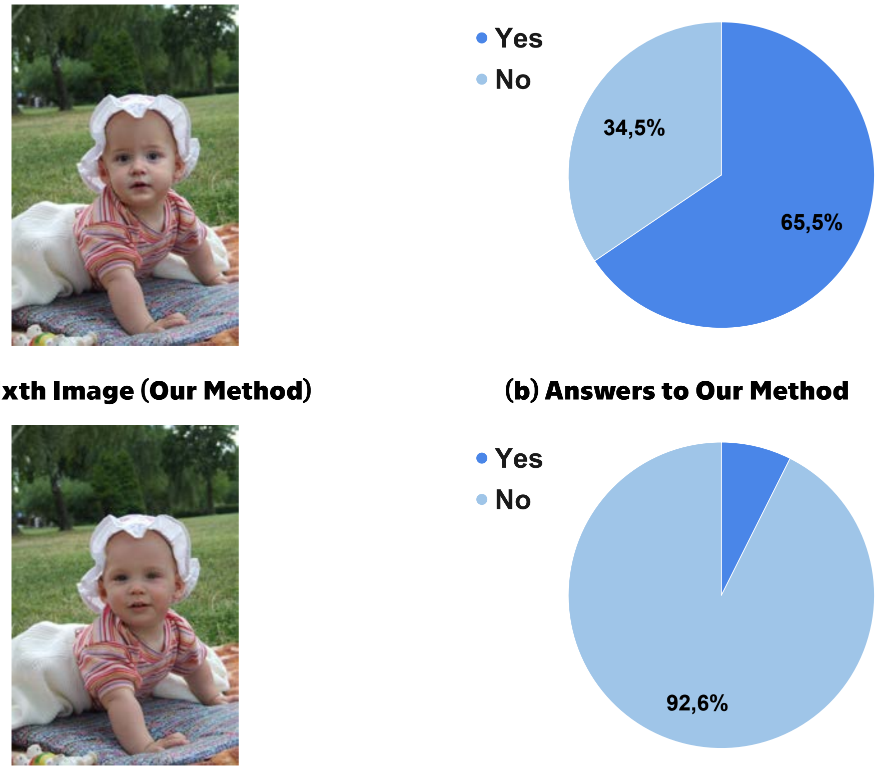
The ninth images and the fourth anonymized ones can be seen together with their response distributions in Fig. 5.14. Here DeepPrivacy 2 also outperformed our method, even though the face it produced looks quite old compared to the rest of the body. This might have been caused by some of the other anonymized images containing way more artifacts, which confused the participants.

The tenth and final images of the study can be seen together with their response distributions in Fig. 5.15. In this pair our image outperformed DeepPrivacy 2, which looked so disfigured that no participant considered it to be unedited.

5.3.4 Difficulty Evaluation

When evaluating the difficulty of distinguishing anonymized images the participants were asked the question: “How hard was it for you to distinguish the edited images?” They were then supposed to rank their experience on a scale from 1 to 5, where 1 meant that distinguishing the anonymized images was very easy and 5 meant that distinguishing them was very hard.

The distibution of responses from both versions of the study can be seen in Fig. 5.16.



(a) Sixth Image (Our Method)

(b) Answers to Our Method

(c) Sixth Image (DeepPrivacy 2)

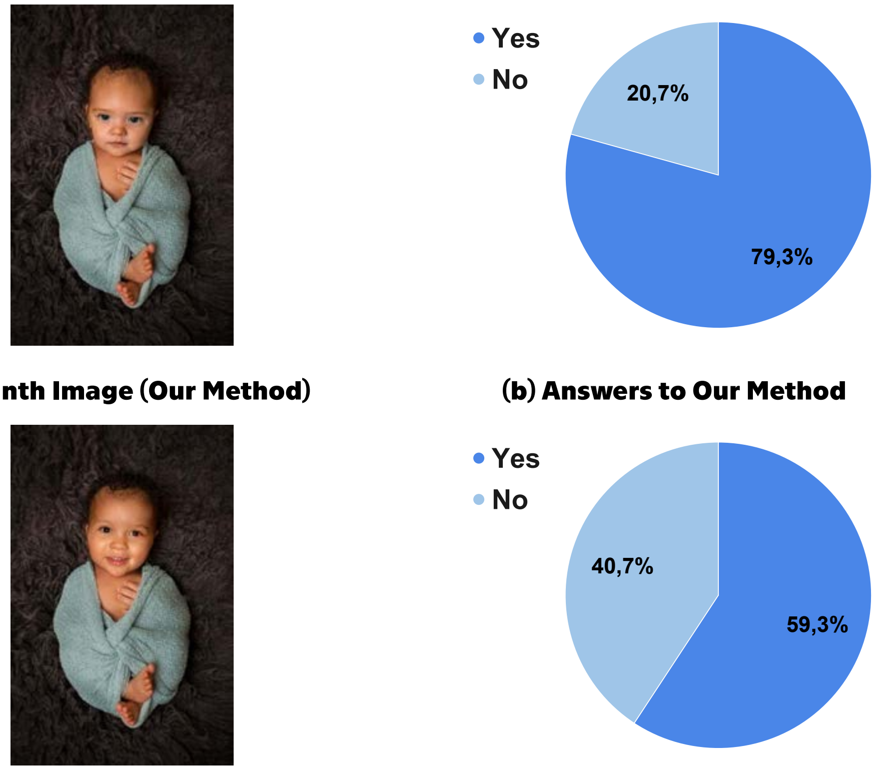
(d) Answers to DeepPrivacy 2

Figure 5.13. Comparison of the results of the sixth image of conducted user studies. The particapants were asked “Is this image anonymized (edited)?”. Sixth image in the version of the study using our method of anonymization (a) and corresponding distribution of answers (b). Sixth image in the version of the study using DeepPrivacy 2 for anonymization (c) and corresponding distribution of answers (d).

5.3.5 Discomfort Evaluation

When evaluating their discomfort the participants were asked the question: “How annoying did you find the artifacts in the edited images?” They were then supposed to rank their experience on a scale from 1 to 5, where 1 meant they were not bothered at all by looking at the artifacts and 5 meant that it was very hard for them to look at the artifacts.

The distibution of responses from both versions of the study can be seen in Fig. 5.17.



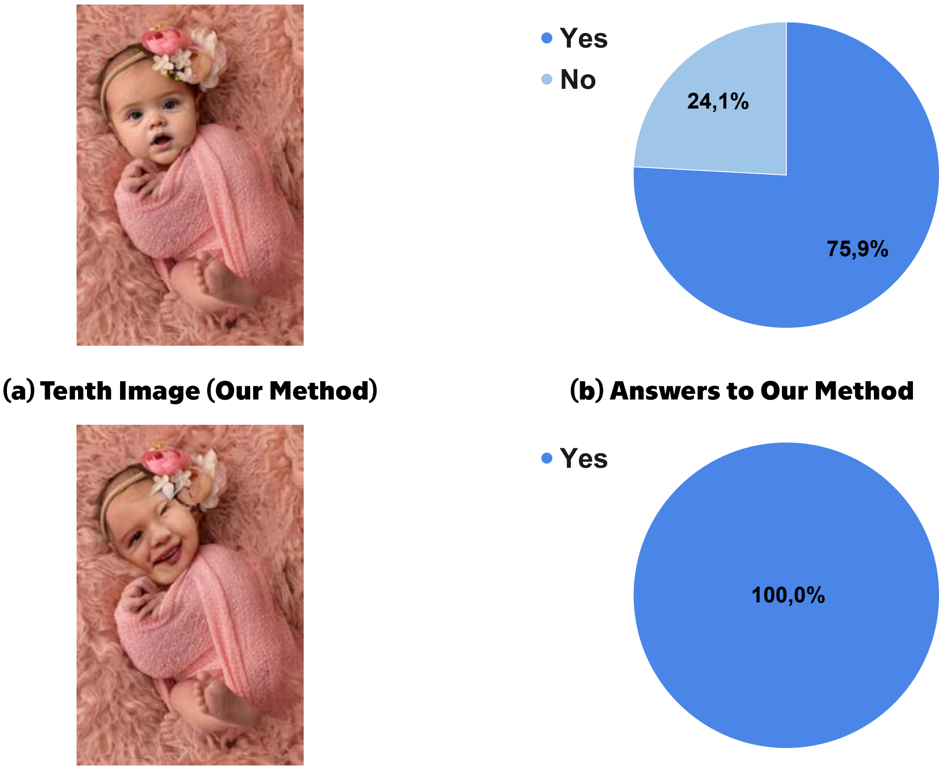
(a) Ninth Image (Our Method)

(b) Answers to Our Method

(c) Ninth Image (DeepPrivacy 2)

(d) Answers to DeepPrivacy 2

Figure 5.14. Comparison of the results of the ninth image of conducted user studies. The particapants were asked “Is this image anonymized (edited)?”. Ninth image in the version of the study using our method of anonymization (a) and corresponding distribution of answers (b). Ninth image in the version of the study using DeepPrivacy 2 for anonymization (c) and corresponding distribution of answers (d).



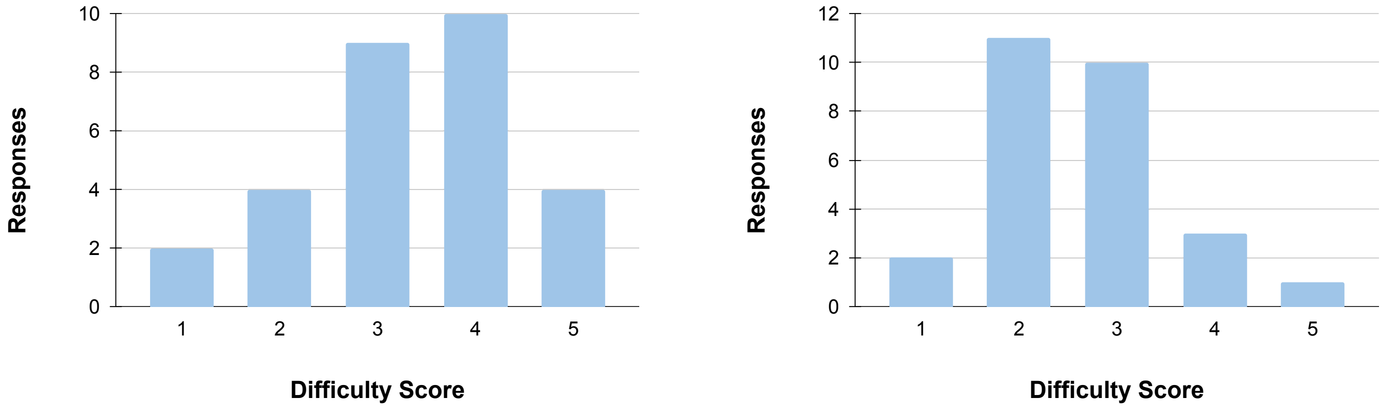
(a) Tenth Image (Our Method)

(b) Answers to Our Method

(c) Tenth Image (DeepPrivacy 2)

(d) Answers to DeepPrivacy 2

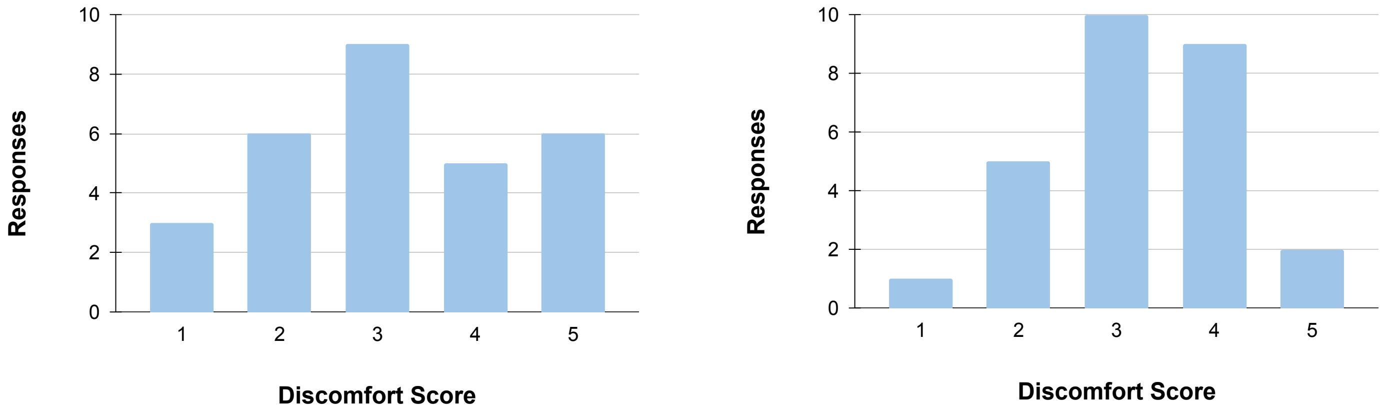
Figure 5.15. Comparison of the results of the tenth image of conducted user studies. The particapants were asked “Is this image anonymized (edited)?”. Tenth image in the version of the study using our method of anonymization (a) and corresponding distribution of answers (b). Tenth image in the version of the study using DeepPrivacy 2 for anonymization (c) and corresponding distribution of answers (d).



(a) Our Method

(b) DeepPrivacy 2

Figure 5.16. Comparison of anwers to the question “How hard was it for you to distinguish the edited images?” between the version using our method of anonymization (a) and the version of the study using DeepPrivacy 2 (b). Where 1 corresponds to lowest difficutly and 5 to highest difficulty.



(a) **Our Method**

(b) **DeepPrivacy 2**

Figure 5.17. Comparison of anwers to the question “How annoying did you find the artifacts in the edited images?” between the version using our method of anonymization (a) and the version of the study using DeepPrivacy 2 (b). Where 1 corresponds to not at all and 5 to it was hard to look at them.

Chapter 6
Limitations and Future Work

Both our anonymization model and benchmarking suite functioned effectively, yet they also exhibited several flaws and limitations. In this section, we identify these limitations and suggest potential solutions for some of them, which could be explored in future work.

Many of the limitations of both our model and our benchmarking suite were due to the libraries we used to perform various parts of their function.

6.1 Face Detection

The face detector we used sometimes struggled with detecting faces. A few examples of faces that were not detected can be seen in Fig. 6.1. It also occasionally falsely identified non-face objects as faces. Examples of these false positives can be seen in Fig. 6.2.

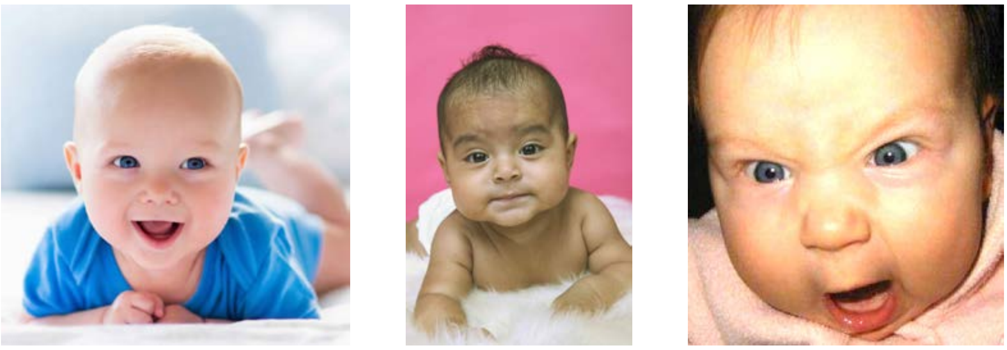


Figure 6.1. Examples of faces not detected by our face detector YOLOv8.

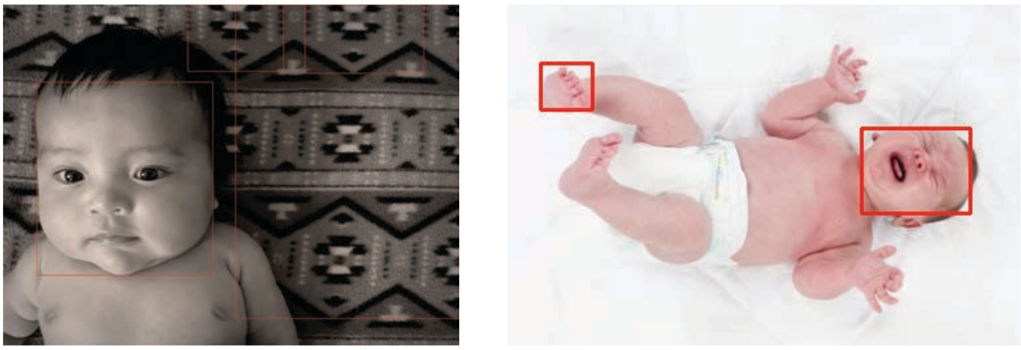


Figure 6.2. Examples of false positives of our face detector YOLOv8.

To improve the face detection component of our method, we could utilize multiple face detectors and compare their results. This approach could reduce both the number of undetected faces and the number of false positives.

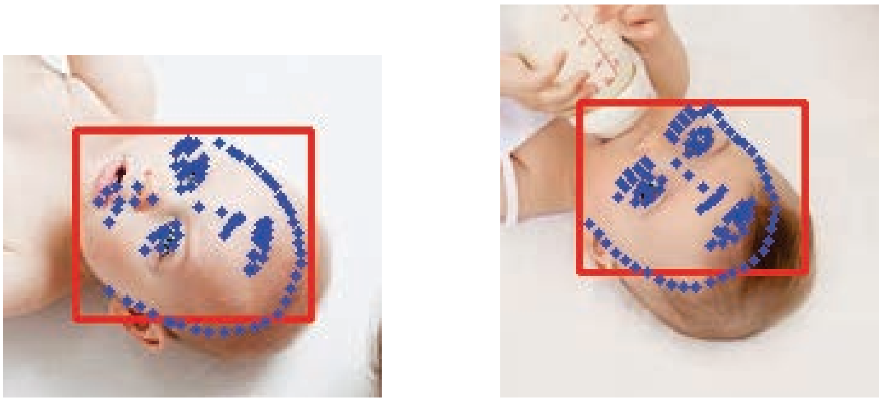


Figure 6.3. Examples of facial landmarks incorrectly detected by SPIGA.

6.2 Facial Landmarks Detection

Our facial landmarks detector sometimes struggled with accurately detecting the landmarks, particularly for faces that were rotated relative to the camera or had obstructions in front of them. Examples of incorrectly detected landmarks are shown in Fig. 6.3.

Similar to face detection, using multiple facial landmarks detectors could help determine the most likely positions of distinct facial features.

6.3 Face Mask

The method for obtaining the binary mask we use for inpainting only allows for masks of the inner face. It could be beneficial to test the behavior of our anonymization method with different shapes of face masks, such as rectangles obtained from the bounding box of each face.

6.4 Metrics of Benchmarking Suite

Many of the metrics utilized by our benchmarking suite did not function as expected. This issue likely arises because babies typically constitute a small part of the datasets used to train the methods for evaluating these metrics. Separate testing to verify the correctness of these metrics should be conducted.

Our model contained many different metrics, making the overall evaluation of our experiments cumbersome. For further testing, it would be beneficial to combine the statistics in each category into a single scalar value that could be compared more easily.

6.5 Anonymization Time

Diffusion models are time-consuming when generating images, making our anonymization method quite slow. Running our method on a computation server with four Tesla A 100 graphics cards, anonymizing a single face usually took around ten seconds, though times varied based on parameters and extensions used. For this reason real-time anonymization using our method is currently infeasible. Anonymizing our entire dataset of 128 images typically took around an hour. The primary factor influencing the time required to anonymize all the images was the number of faces they contained.

6.6 Sampling-and-Testing-Anonymization

A simple way in which our anonymization method could be improved is the utilization of inherent randomness of the Stable Diffusion image generation. To do this we could try generate a batch of images from the same inpainting input and then select the best one based on the metrics of our benchmarking suite. The disadvantage of this approach is that it would increas the anonymization time as many times as we would inpaint each face.

6.7 Identity Constistency in Videos

Using celebrity names in the positive prompt did not achieve the desired identity consistency across videos. Another approach worth testing is using ControlNet to condition the inpainting process based on the anonymized faces from previous frames.

Chapter 7
Conclusion

This thesis has centered on the crucial task of facial anonymization, particularly focusing on infants, aiming to preserve facial features while effectively de-identifying faces. So their anonymized faces can published while not compromising their identity and still allowing for their use in downstream tasks

Initially, we introduced a straightforward method for anonymizing both images and videos, leveraging Stable Diffusion.

Subsequently, we developed a comprehensive benchmarking suite capable of assessing anonymization quality across images and videos. Our evaluation criteria encompass the quality of de-identification, preservation of facial features, and the degree of introduced artifacts.

A series of experiments were then conducted to gauge the efficiency of our anonymization approach. Leveraging our benchmarking suite, we meticulously analyzed the performance across various scenarios.

Specifically, we conducted experiments centered on image anonymization, utilizing a dataset comprising 128 images of infants. Here, we compared our method to DeepPrivacy 2 and explored the impact of different parameters within Stable Diffusion on anonymization quality.

Furthermore, we extended our analysis to videos, conducting comparative evaluations with DeepPrivacy 2 and investigating the effects of various parameters within our method.

Additionally, we conducted a human perception study to gauge reactions to images anonymized using our method, providing a comparative analysis with DeepPrivacy 2.

Based on our evaluation, both our method and DeepPrivacy 2 exhibit comparable overall quality. Our method slightly outperforms DeepPrivacy 2 in preserving the facial features of individuals, while DeepPrivacy 2 performs better when it comes to de-identification. Both methods introduce a similar amount of artifacts during the anonymization process.

It is important to note that some metrics where our method underperformed may have been influenced by the presence of infants in our dataset. Infants typically represent only a small portion of the datasets used to develop the tools these metrics are based on, potentially skewing the results.

Lastly, we outlined the limitations of our anonymization method and proposed potential avenues for improvement.

In summary, we have developed a facial anonymization method, accompanied by an evaluation tool. Through rigorous experimentation and comparison with an existing method, we have effectively assessed its efficiency and identified areas for enhancement.

References

[1] Karen Adolph, Rick Gilmore, and John Kennedy. Video data and documentation will improve psychological science. *Psychological Science Agenda*. 2017, 31 (10).

[2] Karen E Adolph, and Susan R Robinson. Sampling development. *Journal of Cognition and Development*. 2011, 12 (4), 411–423.

[3] Slobodan Ribaric, and Nikola Pavesic. *An overview of face de-identification in still images and videos*. In: *2015 11th IEEE International conference and workshops on automatic face and gesture recognition (FG)*. 2015. 1–6.

[4] E.M. Newton, L. Sweeney, and B. Malin. Preserving privacy by de-identifying face images. *IEEE Transactions on Knowledge and Data Engineering*. 2005, 17 (2), 232-243. DOI 10.1109/TKDE.2005.32.

[5] Jiří Moravčík. *Face anonymizer*.
`http://hdl.handle.net/10467/109451`. 2023.

[6] Tianlei Zhu, Junqi Chen, Renzhe Zhu, and Gaurav Gupta. *StyleGAN3: Generative Networks for Improving the Equivariance of Translation and Rotation*. 2024.

[7] Håkon Hukkelås, and Frank Lindseth. *DeepPrivacy2: Towards Realistic Full-Body Anonymization*. In: *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 2023. 1329-1338.

[8] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. *YOLO by Ultralytics*. 2023.
`https://github.com/ultralytics/ultralytics`.

[9] Azamat Kanametov. *YOLOv8 Face*.
`https://github.com/akanametov/yolov8-face`. 2023.

[10] Andrés Prados-Torreblanca, José M Buenaposada, and Luis Baumela. *Shape Preserving Facial Landmarks with Graph Attention Networks*. In: *33rd British Machine Vision Conference 2022, BMVC 2022, London, UK, November 21-24, 2022*. BMVA Press, 2022.
`https://bmvc2022.mpi-inf.mpg.de/0155.pdf`.

[11] Marcelo Bertalmío, Guillermo Sapiro, Vicent Caselles, and C. Ballester. *Image inpainting*. In: 2000. 417-424.

[12] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. *High-Resolution Image Synthesis with Latent Diffusion Models*. 2021.

[13] AUTOMATIC1111. *Stable Diffusion web UI*. 2023.
`https://github.com/AUTOMATIC1111/stable-diffusion-webui`.

[14] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. *Adding Conditional Control to Text-to-Image Diffusion Models*.

[15] Jiří Moravčík. *AnonyBench*.
`https://github.com/jirimoravcik/AnonyBench`. 2023.

[16] Sefik Ilkin Serengil, and Alper Ozpinar. *HyperExtended LightFace: A Facial Attribute Analysis Framework*. In: *2021 International Conference on Engineering and*

Emerging Technologies (ICEET). 2021. 1-4. https://ieeexplore.ieee.org/document/9659697.

[17] Lingzhi Zhang, Yuqian Zhou, Connelly Barnes, Sohrab Amirghodsi, Zhe Lin, Eli Shechtman, and Jianbo Shi. Perceptual Artifacts Localization for Inpainting. *arXiv preprint arXiv:2208.03357*. 2022.

Appendix A
Base Set of Stable Diffusion Parameters

Base set of parameters used for anonymization of images by our method. It is formatted as a Python dictionary:

```
parameters = {
    "prompt": "(Daryl Sabara: 0.2), (Macaulay Culkin: 0.1),
    (Thomas Sangster: 0.1), (Kelly Macdonald: 0.1), (Taylor Swift: 0.2),
    (Sydney Sweeney: 0.2), (photo of a little baby face: 1.2)",
    "negative_prompt": "(deformed iris, deformed pupils, semi-realistic,
    cgi, 3d, render, sketch, cartoon, drawing, anime, painting,
    black and white, bubble gum, face mask), text, cropped, out of frame,
    worst quality,     low quality, jpeg artifacts, ugly, duplicate,
    morbid, mutilated, extra fingers, mutated hands, poorly drawn hands,
    poorly drawn face, mutation, deformed, blurry, dehydrated,
    bad anatomy, bad proportions, extra limbs, cloned face, disfigured,
    gross proportions, malformed limbs, missing arms, missing legs,
    extra arms, extra legs, fused fingers, too many fingers, long neck",
    "steps": 30,
    "width": 896,
    "height": 896,
    "resize_mode": 0,
    "sampler_name": "DPM++ SDE Karras",
    "cfg_scale": 4,
    "initial_noise_multiplier": 1,
    "denoising_strength": 0.7,
    "n_iter": 1,
    "init_images": [<image_byte64>],  # image in a base64 encoding
    "batch_size": 1,
    "mask": <mask_byte64>,  # mask in a base64 encoding
    "mask_blur": 4,
    "mask_blur_x": 4,
    "mask_blur_y": 4,
    "mask_mode": 1,  # 'Inpaint masked', 'Inpaint not masked'
    "inpainting_fill": 1,  # 'fill', 'original', 'latent noise',
        'latent nothing'
    "inpaint_full_res": 1,  # "Whole picture", "Only masked"
    "inpaint_full_res_padding": 96,
    "inpainting_mask_invert": 0,
    "override_settings": {  # this can be used to switch SD model
        'sd_model_checkpoint': "RealisticVisionV20.safetensors",
    },
    "seed": 1,
    "subseed": -1,
    "subseed_strength": 0,
    "seed_enable_extras": True,
    "seed_resize_from_h": -1,
    "seed_resize_from_w": -1,
    "tiling": False,
    "styles": [],
    "restore_faces": False,
    "script_args": [],
    "script_name": None,
    "refiner_switch_at": 0.4,
    "alwayson_scripts": {
        "API payload": {"args": []},
        "CodeFormer": {"args": [0, 0]},
        "ControlNet": {"args": [
            {"advanced_weighting": None,
             "batch_images": "",
             "control_mode": "ControlNet is more important",
             "enabled": True,
             "guidance_end": 1,
             "guidance_start": 0,
             "hr_option": "Both",
             "image": None,
             "inpaint_crop_input_image": True,
             "input_mode": "simple",
             "is_ui": True,
             "loopback": False,
             "low_vram": False,
             "model": "control_v11p_sd15_openpose",
             "module": "openpose_full",
             "output_dir": "",
             "pixel_perfect": True,
             "processor_res": -1,
             "resize_mode": "Just Resize",
             "save_detected_map": True,
             "threshold_a": -1,
             "threshold_b": -1,
             "weight": 1}]},
        "Extra options": {"args": []},
        "GFPGAN": {"args": [0]},
        "Refiner": {"args":
            [True, "RealisticVisionV60B1.safetensors", 0.4]},
        "Seed": {"args": [-1, False, -1, 0, 0, 0]}
    },
    "comments": {},
    "disable_extra_networks": False,
    "image_cfg_scale": 1.5,
    "refiner_checkpoint": "RealisticVisionV60B1.safetensors",
    "s_churn": 0.0,
    "s_min_uncond": 0.0,
    "s_noise": 1.0,
    "s_tmax": None,
    "s_tmin": 0.0
}
```

Appendix B
Comparison Images of Additional Conducted Experiments

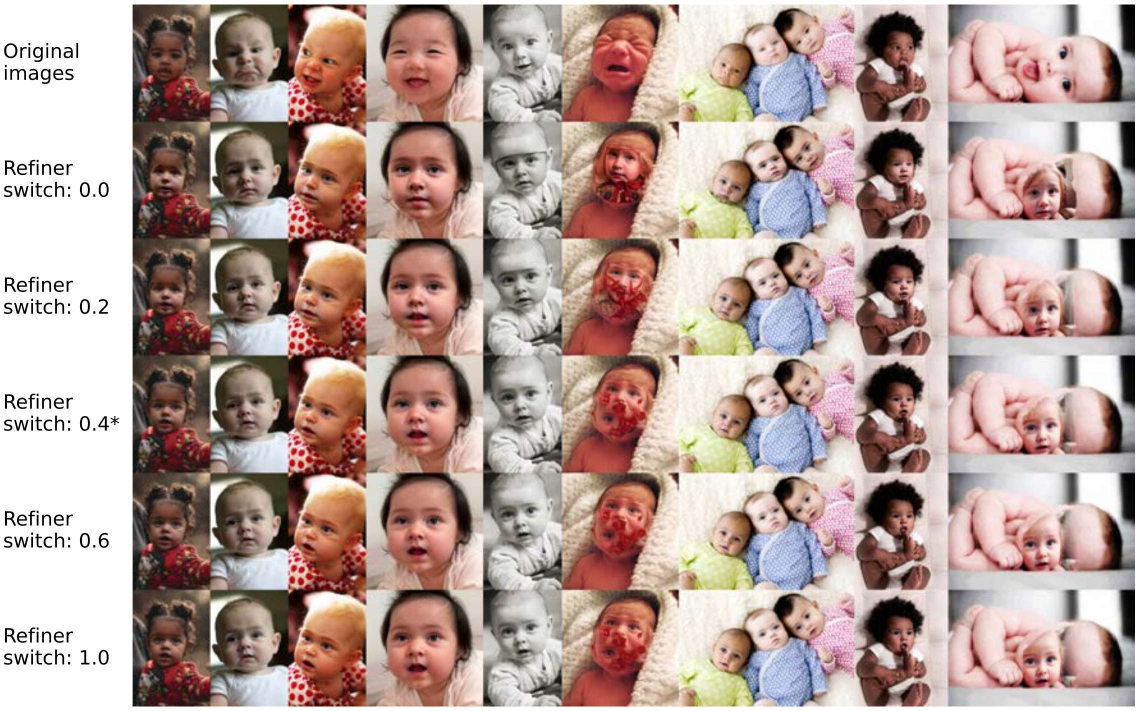


Figure B.1. Comparison of results of different times of switching between the used Stable Diffusion models.

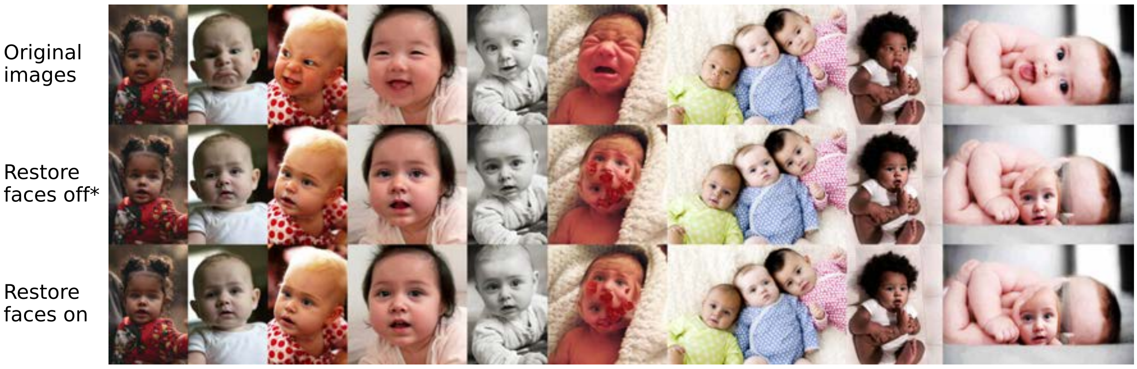


Figure B.2. Comparison of the effects of em restore faces parameter.

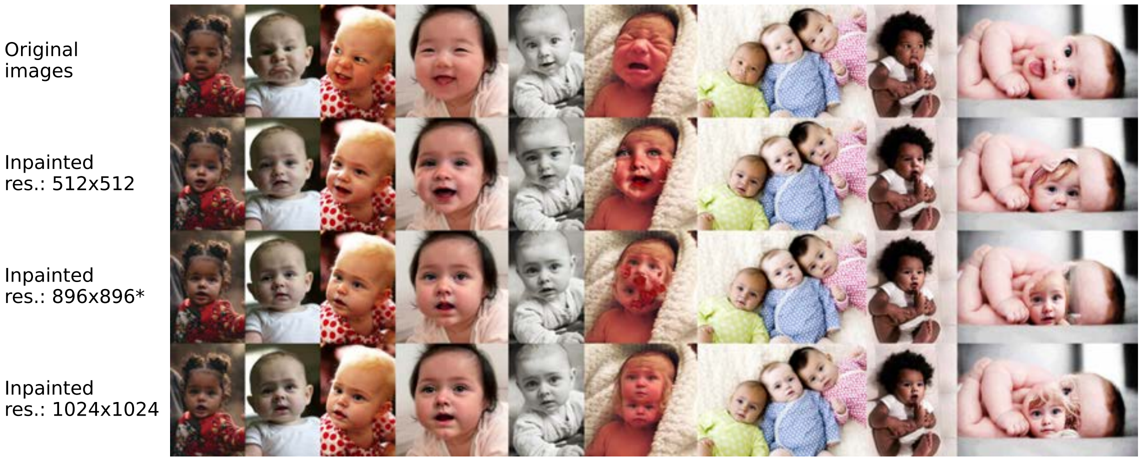


Figure B.3. Comparison of images anonymized using different inpainted resolutions.

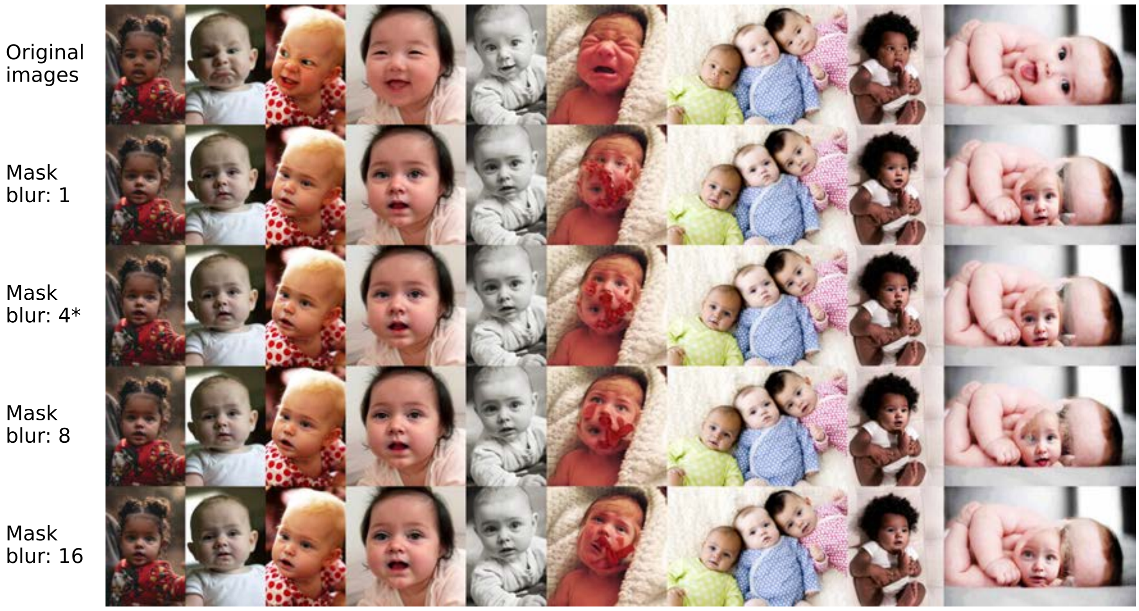
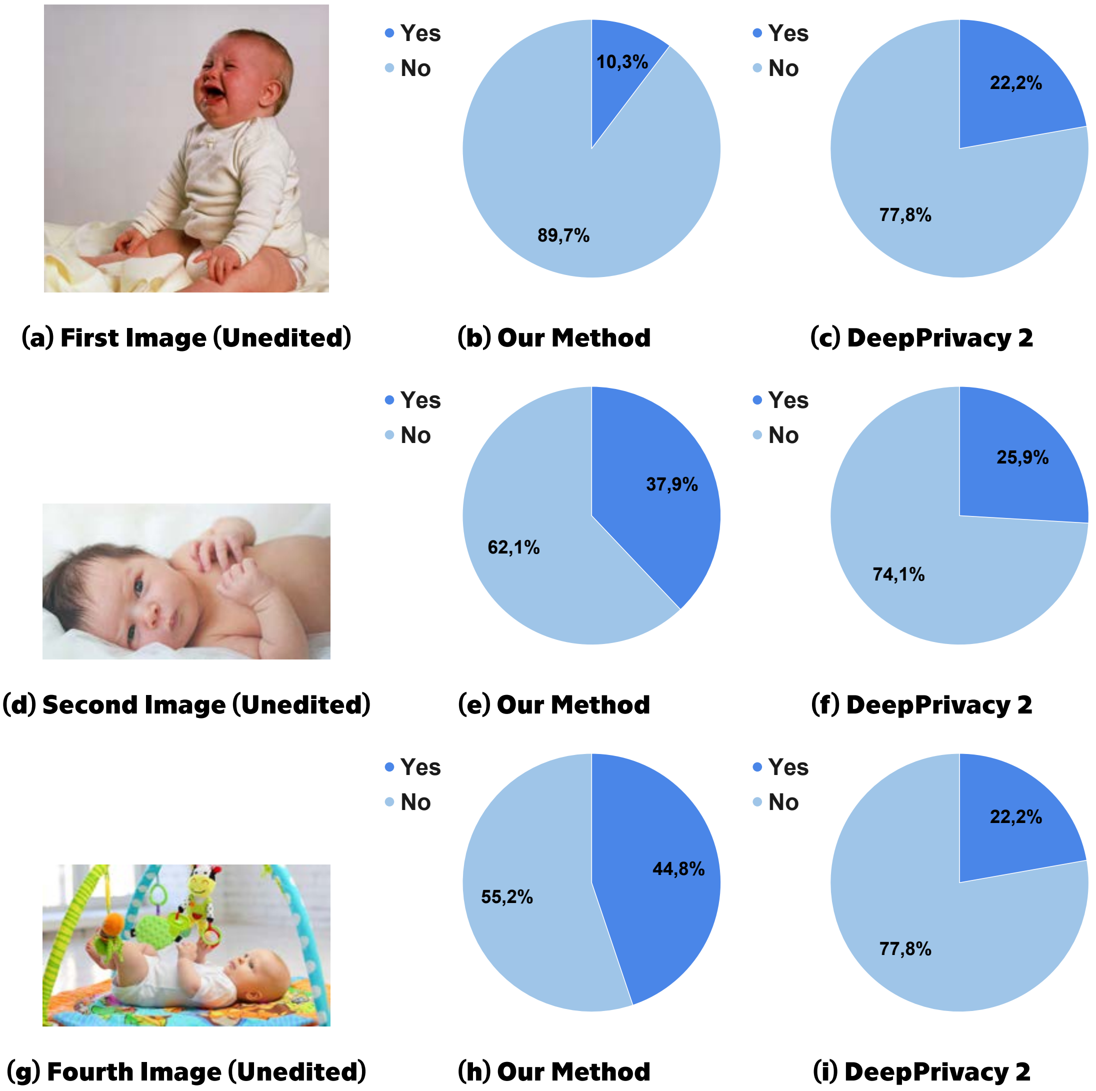


Figure B.4. Comparison of images anonymized using different mask blur values.

Appendix C
Answear Distributions of Original Images from User Study

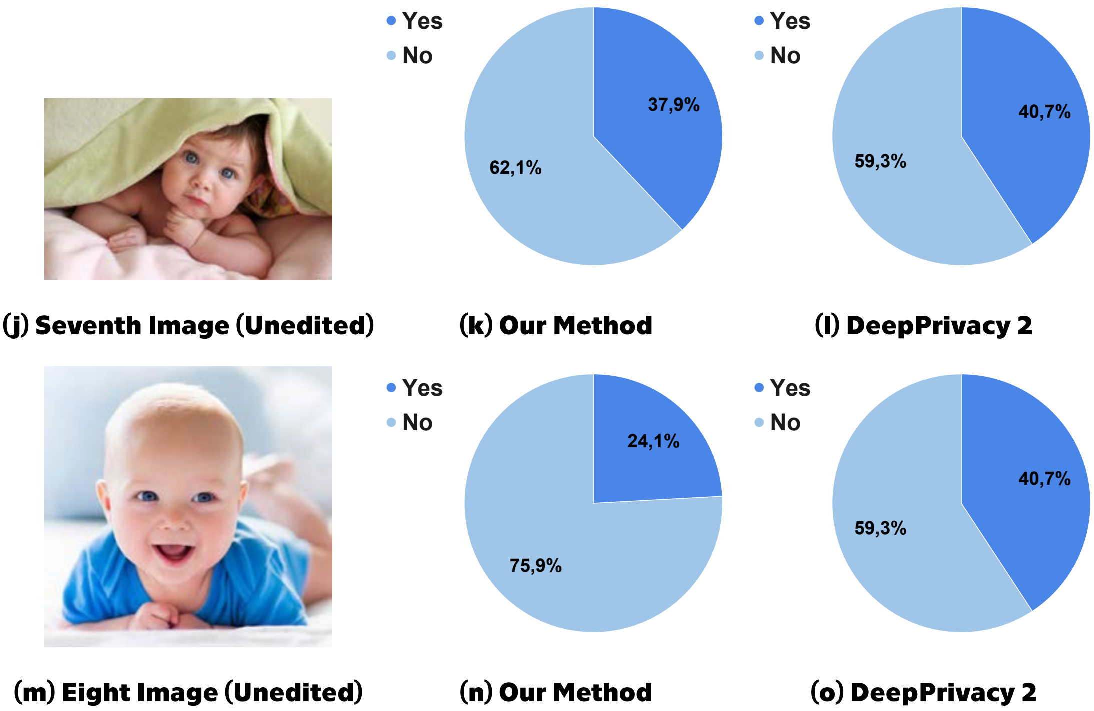


(a) First Image (Unedited) **(b) Our Method** **(c) DeepPrivacy 2**

(d) Second Image (Unedited) **(e) Our Method** **(f) DeepPrivacy 2**

(g) Fourth Image (Unedited) **(h) Our Method** **(i) DeepPrivacy 2**

Figure C.5. First three of the original images used in both versions of the user study and their their right the relative responses to the question "Is this image anonymized (edited)?" for our method and DeepPrivacy 2.



(j) Seventh Image (Unedited) **(k) Our Method** **(l) DeepPrivacy 2**

(m) Eight Image (Unedited) **(n) Our Method** **(o) DeepPrivacy 2**

Figure C.6. First two original images used in both versions of the user study and their their right the relative responses to the question “Is this image anonymized (edited)?” for our method and DeepPrivacy 2.