applied sciences

MDPI

Article

A Multi-Input Fusion Model for Privacy and Semantic Preservation in Facial Image Datasets

Yuanzhe Yang [1], Zhiyi Niu [2], Yuying Qiu [3], Biao Song [1,*], Xinchang Zhang [1] and Yuan Tian [4]

[1] School of Computer Science, Nanjing University of Information Science & Technology, Nanjing 210044, China
[2] Institute of Electronic and Electrical Engineering, Civil Aviation Flight University of China, Deyang 618307, China
[3] School of Data Science and Engineering, East China Normal University, Shanghai 200062, China
[4] Nanjing Institute of Technology (NJIT), Nanjing 210094, China
* Correspondence: bsong@nuist.edu.cn

Abstract: The widespread application of multimedia technologies such as video surveillance, online meetings, and drones facilitates the acquisition of a large amount of data that may contain facial features, posing significant concerns with regard to privacy. Protecting privacy while preserving the semantic contents of facial images is a challenging but crucial problem. Contemporary techniques for protecting the privacy of images lack the incorporation of the semantic attributes of faces and disregard the protection of dataset privacy. In this paper, we propose the Facial Privacy and Semantic Preservation (FPSP) model that utilizes similar facial feature replacement to achieve identity concealment, while adding semantic evaluation to the loss function to preserve semantic features. The proposed model is versatile and efficient in different task scenarios, preserving image utility while concealing privacy. Our experiments on the CelebA dataset demonstrate that the model achieves a semantic preservation rate of 77% while concealing the identities in facial images in the dataset.

Keywords: semantic preservation; face de-identification; deep learning; autoencoders

check for updates

Citation: Yang, Y.; Niu, Z.; Qiu, Y.; Song, B.; Zhang, X.; Tian, Y. A Multi-Input Fusion Model for Privacy and Semantic Preservation in Facial Image Datasets. *Appl. Sci.* **2023**, *13*, 6799. https://doi.org/10.3390/app13116799

Academic Editor: Gianluca Lax

Received: 9 May 2023
Revised: 28 May 2023
Accepted: 30 May 2023
Published: 2 June 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In today's digital age, the widespread application of multimedia technologies such as video surveillance, online meetings, and drones has resulted in the need for a large amount of personal private data. Due to the high potential for misuse of private data contained in multimedia, the protection of privacy is becoming increasingly important. Special emphasis should be placed on facial privacy protection in safeguarding personal privacy in multimedia, as facial features can reveal individuals' identities, facial semantics, and other crucial information.

Faces obtained from multimedia can be used for various purposes such as semantic research and analysis, which is useful for understanding patterns and trends related to humans. Specifically, the semantics of a face can include lots of useful information such as emotions, race, and so on. These semantics can be used for a variety of purposes such as facial recognition or to understand the impact of emotions on behavior. While this capability offers many benefits in various fields, it also raises concerns about the potential misuse of such semantics and the infringement of privacy. Taking into account the preservation of semantics, face de-identification is not equivalent to completely concealing private information in an image. After de-identification, the disruption of facial features can negatively impact the utility of the processed image for algorithms that rely on facial semantics for specific purposes, particularly in modern machine learning. As a result, ensuring privacy protection while preserving data related to specific requirements of facial-related algorithms after de-identification becomes crucial. However, specific semantics may involve privacy, and the boundary between them is vague. Therefore, it is a conundrum in facial de-identification how to achieve a balance between protecting privacy and retaining

semantics to ensure processed images in the dataset are available for subsequent machine learning algorithms.

Several existing works that aim to preserve the utility of facial images while concealing identity information are surveyed. Reference [1] suggested a technique of a generative model for face privacy protection with utility maintenance which is based on existing privacy protection technologies and innovatively adds the loss of service quality to the loss function, ensuring the generation of de-identified face images with guided quality. Reference [2] utilized conditional generative adversarial networks to remove the identifying characteristics of faces and bodies and produce high-quality images and videos. Reference [3] introduced a user-specific password and an adjustable parameter to control the direction and degree of identity variation to achieve a personalized and invertible de-identification method based on the deep generative model. Reference [4] suggested a technique of image inpainting that combines facial landmarks generated from image context and facial landmark-conditioned head inpainting for generating realistic head inpainting in the photo. However, features will be weakened or blurred after processing by existing methods. There is no de-identification model that can preserve certain aspects of semantics for images in a dataset that can still be used in subsequent machine learning tasks (such as facial expressions) [5] so far. Furthermore, the need to develop an effective face deidentification model cannot be overstated.

To address these challenges, based on previous work, we propose a new de-identification model on the dataset called Facial Privacy and Semantic Preservation (FPSP) model in this paper. The proposed generative model that is capable of producing utility-preserving de-identification images in facial datasets takes advantage of a powerful de-identification model: the Quality Maintenance-Variational AutoEncoder [1]. To generate deidentifying images contained in the dataset while maintaining features based on specific generation goals, for each image, the proposed strategy, on one hand, utilizes corresponding privacy-concealed images processed by several typical protection methods, and on the other hand, considering feature preservation, utilizes facial features extracted from other facial images that are similar to this image in the dataset. This approach ensures that the resulting image retains features and cannot be distinguished from the dataset because of the introduction of approximate facial features, achieving a desired tradeoff between utility and privacy. We transformed face images into corresponding embeddings in the first stage, then classified images into several clusters based on embeddings. Then, we used four typical protection methods, blindfold, mosaic, cartoon, and mosaic, to process the images, forming four sets of images without private information. Along with four de-identified images and a surrogate image generated in the same cluster as the input, we adjusted the loss function to set up a matched facial semantic evaluation function based on specific semantics. Finally, for generated images, we enhanced them by GFP-GAN [6].

The following are the major contributions of this work:

- We first put forward the concept of de-identification of datasets for machine learning while preserving facial semantics, proposing a novel generative model that disrupts the distribution of facial features in images as little as possible and validating that the processed dataset is available for subsequent machine learning.
- We introduce an appropriate approach that merges privacy-concealed faces corresponding to original images and similar facial feature images in the dataset, serving the purpose of enhancing the anonymity of individual identities and the usability of image features.
- We selected facial expression recognition as the machine learning task, utilizing the CelebA dataset [7]. Our model outperforms traditional de-identification and AMT-GAN [8], generating images with 5% more utility. It also maintains facial expression recognition rate, evaluated using RTCNN [9], DACL [10], and DAN [11].

The rest of the article is distributed as follows. In Section 2, we review previous work related to face de-identification. In Section 3, we present some preliminary work related to our research. In Section 4, we describe detailed information about our proposed model, the

FPSP model. In Section 5, quantitative evaluation experiments were performed to verify the performance of our proposed method, and the results are presented to demonstrate its effectiveness. In Section 6, the proposed model's capability is discussed along with topics that can be further researched. In Section 7, we give a brief conclusion of our research.

2. Related Work

Over the past few years, there has been a significant amount of research [12–17] conducted on safeguarding the privacy of facial images. Previous methods such as image blurring, color block masking [18], and pixelation [19,20] are simple but lack a formal privacy model. Although applicable to all kinds of images, the privacy of people in the image is not always guaranteed. Consequently, these methods are naive and fail to effectively preserve data utility and privacy protection. In the forthcoming sections, several novel and model techniques, which have shown promising results in protecting the privacy of individuals in images, will be introduced.

2.1. K-Same

The algorithm family of k-same [21–25] that adopts the average face of similar faces in the whole image set to represent the resulting face provides a sequence of techniques with theoretical anonymity guarantees. In this regard, various k-same algorithms have been developed to address this challenge. Sweeney [26] was the first to introduce the k-anonymity idea and successfully implemented it in a relational database. The idea of k-anonymity subsequently gave rise to a set of k-same algorithms. Newton et al. [27] suggested a k-same algorithm that preserves the visual features of all images in a cluster by working directly in the pixel space. To deal with the problem that generated de-identified images suffer from ghosting effects, the k-Same-Model [22] was offered. As proposed by Gross et al., the k-Same-Model adopts the idea of Active Appearance Models [28], which achieves better alignment between images and synchronized images appear more realistic. Proposed by Sun et al., k-Diff-furthest [29] utilized a novel algorithm to address the issue that the generated de-identified facial image has minor differences. Reference [30] proposed k-Same-furthest-FST. It completes the goal of identity protection by morphing the privacy-free face region and original background. What is more, this method has been proven to deliver substantial safeguarding of facial privacy within the context of the FERET database [31]. Meden et al. [32] proposed the k-Same-Net scheme. By combining the principle of anonymity with GNN architectures, this approach achieves impressive visual outcomes. Reference [33] explores a novel model based on AAM. It divides the face space into two subspaces and uses the appropriate k-anonymity technique to process utility to achieve face de-identification. Although facial privacy is, to some extent, well protected by k-anonymity family algorithms, these algorithms come with substantial limitations, as each individual can only be represented once in the dataset. This may cause a decrease in identity diversity in the whole dataset.

2.2. GAN

Currently, generative models that synthesize synthetic but natural-looking images provide novel ideas for de-identification. Generative Adversarial Networks [34] (GANs) are the most relatively well-known and commonly used model among all generative models and have paved a novel way for research on face de-identification [35–38]. The components of it are two competing deep models: a generative model and a second discriminator network. Qi et al. [39] put forward a novel Loss-Sensitive Generative Adversarial Network (LS-GAN). It combines the set of mathematical principles and computational techniques, which involve constraining the Lipschitz constant and measuring the rate of change of the function or model with the utilization of the idea of game theory to promote the generator to generate the most realistic samples. The authors of reference [40] introduced a novel model based on the architecture of a Variational Generative Adversarial Network (VGAN), combining the Variational Autoencoder (VAE) [41] and CGAN [42]. This approach is

capable of extracting image representations that are specifically disentangled from identity information. CGAN-based PPGAN was proposed by Wu et al. [43], which utilized the discriminator in the pre-trained model to output a structurally similar image, using the extracted feature space related to identity privacy. To boost the effectiveness of concealing personal information, Li et al. [44] proposed a novel GAN called SF-GAN. This method combines both a geometry synthesizer and an appearance synthesizer to construct various external mechanisms, achieving the goal of facial-related feature concealing. An au-to-GAN-based identity protection method, which lowers the data dimensionality for the database utilized to facilitate machine learning operations, was proposed in [45]. It generates deidentified data through confrontation. Reference [46] put forward FPGAN, an end-to-end method. It uses an improved convolutional neural network that involves both an encoder and a decoder pathway and two discriminators, and its loss function is devised depending on the requirement of the specific scenario of service. GAN-based methods lack the powerful grasping ability for facial features; thus, each of these methods has a common drawback in that the details of the generated facial images are not perfect.

2.3. Differential Privacy

Various types of privacy-preserving techniques are utilized in machine learning operations, with differential privacy being the most renowned approach. Dwork [47] et al. proposed it in 2006. Its background is based on the assumption that attackers have access to all non-target information but cannot discern whether the data of a specific individual are included in the dataset. This approach offers a methodical and numerically measured means of evaluating the potential probability of disclosing private information. The concept of Differential Privacy (DP) [48] is commonly employed in numerous identity-concealing technologies [49] for facial images. In order to address the issue of inconsistent noise and uniform errors in multimedia datasets, a solution known as the D-noise-mean algorithm is proposed in reference [50]. This algorithm utilizes a combination of the KD-tree [51] and multi-party secure computation techniques [52], and replaces the median with an approximate mean of noise. Reference [53] proposed a new model called PEEP. In order to safeguard against privacy attacks, such as model memorization attacks [54] or membership inference [55], the model employs local differential privacy techniques. This involves adding noise to the characteristics of the facial distribution extracted from the original images, thereby protecting against the unauthorized disclosure of personal information. The perturbed data are stored on third-party servers. In order to address privacy concerns on end devices while maintaining high data utility for analytical tasks such as inversion attacks [56], reference [57] presents a novel scheme in the field of differential privacy. This approach employs an efficient face representation technique within the Bloom filter space. In order to overcome the issue of imbalanced data distribution within a dataset, reference [58] proposes a method that partitions the data into two levels based on the quantity of data points in each partitioned grid. This approach employs an adaptive partitioning strategy that meets the requirements of identity shielding. DP-GAN, proposed in reference [59], is a framework for generating privacy-preserving facial images that is specifically designed for semantic-rich data. The approach leverages a deep generative model and trains it in a differentially private manner using original data. However, despite its effectiveness in generating privacy-preserving facial images, this approach is not suitable for some specific tasks related to privacy protection. A new model named PPSGAN is proposed in reference [60] to address the issue related to privacy preservation. The model utilizes the self-attention mechanism to add noise to the features that are independent of privacy preservation. By doing this, the generated images can still match the original label. This approach aims to generate facial images that preserve privacy while maintaining the image's semantic meaning. While differential privacy does de-identify privacy from identification, privacy protection and data utility are coupled due to the need to add noise to protect privacy, i.e., increasing noise to enhance privacy protection will directly result in a decrease in data utility.

In general, while the aforementioned techniques have demonstrated notable advancements, given the particularity of the protection for specific tasks in machine learning datasets such as facial expression recognition, semantic segmentation, and so on, these generally designed methods are not applicable. Hence, the design of customized privacy-protection and utility-maintaining techniques for specific datasets holds significant value.

3. Preliminary

3.1. Affine Transformation

In order to obtain a different face that still preserves original features, our proposed method utilizes affine transformation [61,62], while maintaining both affine subspace dimensions and the ratios of the lengths of parallel line segments. As a result, units of parallel affine subspaces stay parallel after affine transformation. Affine transformation consists of translation, scaling, homothety, similarity, reflection, rotation, shear mapping, and combinations of them in any combination and sequence. Generally written in homogeneous coordinates, the affine transformation is as shown in Equation (1)

$$\begin{pmatrix} a_2 \\ a_2 \end{pmatrix} = A \cdot \begin{pmatrix} a_1 \\ a_1 \end{pmatrix} + B \tag{1}$$

where (a_1, b_1) represents the value of pixel intensity in an original image and (a_2, b_2) is the new coordinate after the transformation.

$$A = \begin{pmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{pmatrix}, B = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{2}$$

Matrix A depicted in Equation (2) represents the pure rotation

Now, given the old coordinates (x, y), the new coordinates (x', y') for the image are obtained considering the initial angle θ and the angle of rotation θ'. x and y are depicted in Equation (3)

$$x = rcos\theta, y = rsin\theta \tag{3}$$

As $x' = rcos(\theta + \theta')$ and $y' = rsin(\theta + \theta')$, x' and y' are presented in Equation (4)

$$x' = rcos\theta cos\theta' - rsin\theta sin\theta', y' = rsin\theta cos\theta' - rcos\theta sin\theta'. \tag{4}$$

So, the conversion between (x', y') and (x, y) can be described by Equation (5).

$$x' = xcos\theta' - ysin\theta', y' = ycos\theta' + xsin\theta' \tag{5}$$

3.2. Image Enhancement

As the pictures generated are not always visually good, it is necessary to enhance the image texture. To recover high-quality faces from the generated images, GFP-GAN [6] is adopted in our proposed model. The architecture of the whole model can be divided into two parts. One of the parts is a module for removing image degradation, which is known as U-net, and the second part is a prior in the form of a facial generative adversary network that has been trained in advance. Two networks are connected by a mapping of latent embeddings and a series of several layers called Channel-Split Spatial Feature Transform (CS-SFT) layers. For the original input image, U-net retrieves useful features, forming two features: F_{latent} and $F_{spatial}$. The formulation is as follows:

$$F_{latent}, F_{spacial} = U - Net(x) \tag{6}$$

After passing through multiple layers of the network, F_{latent} is mapped to intermediate latent codes W. The characteristics are recorded in the pre-trained StyleGAN2 model [63]

and those stored features are utilized to generate convolutional features, denoted by F_{GAN}. Equation (7) shows the above process.

$$W = MLP(F_{latent}), F_{GAN} = StyleGAN(W) \tag{7}$$

To enhance the preservation of accuracy, it employs a Spatial Feature Transform (SFT) [64] to modify the GAN features F_{GAN} by transforming the input spatial features $F_{spatial}$. More specifically, the model utilizes a set of convolutional layers to generate the parameters (m, n) for an affine transformation from the input features $F_{spatial}$, as shown in Equation (8)

$$m, n = Conv(F_{spatial}) \tag{8}$$

After that, the F_{GAN} feature is manipulated by changing its scale and position based on the parameters obtained in the previous step, as represented by Equation (9)

$$F_{output} = SFT(F_{GAN}|m, n) \tag{9}$$

It takes advantage of input features $F_{spatial}$ to modify certain GAN features in terms of space while keeping the remaining features unchanged. Finally, the restored face y is generated with the channel-split SFT layers implemented at every resolution level. The process is as shown in Equation (10).

$$y = CS - SFT(F_{GAN}|m, n) \tag{10}$$

4. Method

In this paper, we propose a novel model for preserving facial semantics while concealing identity. The overall process can be divided into multiple stages and appropriate methods are used at each stage. For the rest of this section, we provide a brief description of the FPSP model and its overall structure in Section 4.1. Section 4.2 elaborates on the specific details used and how they work together to achieve the research goals. In Section 4.3, we describe the method execution process, including the data processing and application of algorithms to generate utility-preserved and privacy-free images.

4.1. System Model

When it comes to retaining facial semantics, concealing the identity of individuals is complicated work. Previous approaches do not take the possible loss of the utility of images into consideration in the process of removing identical information. As a result, for specific tasks, the quality of service will be far from expected. To tackle the trade-off between privacy and semantics, we propose the Facial Privacy and Semantic Preservation (FPSP) model with a delicate architecture design and loss evaluating facial semantics. Figure 1 showcases the general architecture of the model.

Firstly, to all images in the dataset, we cropped faces from images and generated a 128-dimensional face feature embedding for each face through the network. Then, we classified images into 15 clusters according to the face embedding. In the next stage, in each cluster, we generated a surrogate face for every image and applied four protection methods to every face image to construct de-identified datasets. Following that, we input these datasets into a generative model guided by a semantic-related quality evaluation. In the training process, the semantics of generated images were evaluated to calculate the loss, which is a part of the loss function. In maintaining service quality, backpropagation is a significant part that helps update the output continuously.

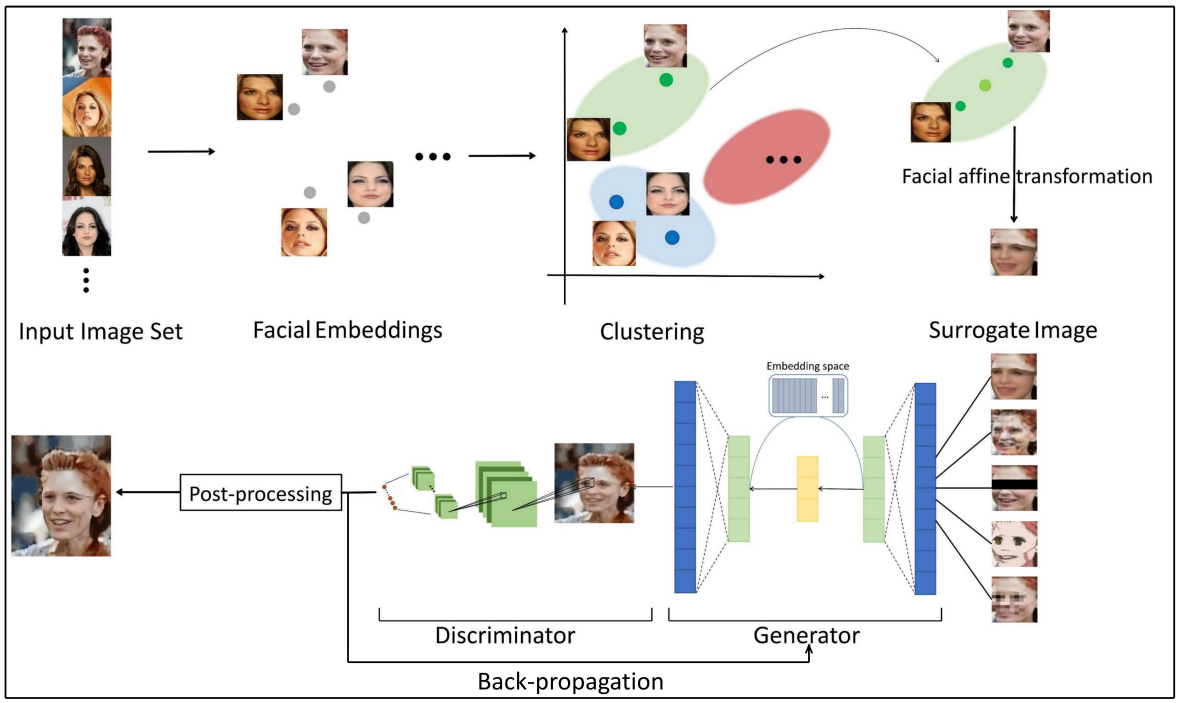

Figure 1. The overall structure of the Facial Privacy and Semantic Preservation (FPSP) model.

4.2. Architecture and Working

4.2.1. Face Extraction and Generation of Corresponding Vector

Taking a set of N different sample images, to reduce the influence of background information and effectively improve the reliability of facial features, the pre-trained MTCNN [65] model was used. It carries out bounding box regression, probability prediction of a real face, and localization of facial landmarks (such as mouths, eyes, and noses) at the same time by applying several networks in a cascade. By detecting the geometric structure of the image, the boundary rectangle of the detected face was returned. According to the detected five facial points, the input face image was cropped from the rectangle. Face alignment, transformation, and normalization [66], which depend on the position of the located key landmark, were performed on cropped faces. After that, faces were further resized to $128 \times 128 \times 3$ pixels, which is inspired by VQ-VAE [67], in which the chosen size speeds up training and sampling and captures the global structure. A set of N identified images $X = \{x_1, x_2, \ldots, x_n\}$ is produced. To extract 128-dimensional facial embedding from each face in X, FaceNet [68], which utilizes deep convolutional networks [69,70] (DNNs) to map face images to a compact Euclidean space, was applied, generating facial features $V = \{v_1, v_2, \ldots, v_n\}$.

4.2.2. Facial Clustering in Dataset

The principal goal of this stage is to divide V into 15 different categories such that the features in one group are very similar, while the difference among different groups is quite large. Firstly, we randomly selected 15 face features $V_\mu = \{v_{\mu_1}, v_{\mu_2}, \ldots, v_{\mu_n}\}$ as centroids. Then, to assign other face features to the closest cluster, we calculated the square errors J between every element in V and every centroid. Facial features in the same cluster are more homogeneous when the value of J is lower. Equation (11) defines the objective function J.

$$J = \sum_{i=1}^{N} \sum_{k=1}^{15} \omega_{ik} \left\| v_i - v_{\mu_k} \right\|^2 \tag{11}$$

where ω_{ik} is set to 0 when v_i is not in one cluster or 1 when v_i is in one cluster.

By controlling the value of ω_{ik}, we obtained the lowest square error based on the random centroids. Furthermore, the belonging of different clusters of face embedding is defined by ω_{ik}. The next step is to update the centroid in each cluster by computing the average value among all face embeddings in the cluster. To obtain the global lowest variation within clusters, we kept iterating the above steps until there was no change to the centroids. The whole faces were then split into 15 clusters.

4.2.3. Surrogate Face Generation

The third stage is to generate synthetic surrogate face images that achieve anonymity in the cluster, i.e., any surrogate face image cannot be linked to the input picture among images in the original cluster in an unambiguous manner on the premise of maintaining the facial features. Given the face image x_i and its corresponding centroid $x_\mu(x_i, x_\mu \in X)$, to obtain the surrogate face image f_{x_i}, we first utilized Dlib [71] to detect key facial feature landmarks on both x_i and x_μ, producing two key point sets S_m and S_n, correspondingly. For every key coordinate (e.g., (a_i, b_i)) in S_m and the corresponding pixel S_n (e.g., (a_j, b_j)), calculating the location of the new key point (e.g., (a_m, b_m)) by Equation (12), we obtained a set of points S_i, which are the key facial feature landmarks of the surrogate face image f_i.

$$a_m = (1-k)a_i + ka_j, b_m = (1-k)b_i + kb_j \tag{12}$$

where k is used to assign the proportional weight of the image coordinates and the corresponding centroid image coordinate. The value of k ranges from 0 to 1.

Then, we performed Delaunay Triangulation for all the sets S_m,S_n,S_i, producing a one-to-one correspondence between triangles in the images x_i, x_μ, and f_i. Selecting each triangle T from the image x_i and its corresponding triangle Y in the image f_i, we calculated the affine transform that converts the triangle T to Y. Similarly, we calculated the transformation matrix of the corresponding triangle W in the image x_μ to the triangle Y. Applying the transformations above to images x_i and x_μ correspondingly, we obtained warped images x_i' and x_μ'. Then, f_i was calculated by using the equation below with warped images.

$$f_i = (1-t)x_i' + tx_\mu' \tag{13}$$

where t is used to assign the proportional weight and it ranges from 0 to 1.

4.2.4. Learning Stage

Inspired by QM-VAE [1], a service-guided generative model that takes several privacy-removed facial images to generate one high-quality image, the framework of this part contains three different parts: the encoder, the decoder, and the embedding space e. Concatenated by the surrogate image and corresponding de-identified images generated by four traditional privacy protection methods (covering the eyes, blurring the faces, adding Laplace noise, and transforming the image into cartoon form), the input i is fed into the encoder, producing $z_e(i)$. Instead of directly transporting $z_e(i)$ to be finished like an ordinary autoencoder, $z_e(i)$ is transformed into the new embedding vector e_i by searching for the closest embedding. Equation (14) shows the one-hot defined form of posterior categorical distribution $q(z|i)$ probabilities.

$$q(z=k|i) = \begin{cases} 1 & if\ k = \underset{j}{argmin}\ ||z_e(i) - e_J||_2 \\ 0 & other \end{cases} \tag{14}$$

In the process of achieving the discretization process, we utilize a mapping described in Equation (15) to transform and identify the closest element of $z_e(i)$, which is subsequently passed to the decoder.

$$z_q(i) = e_k, \quad where\ k = \underset{j}{argmin}\ ||z_e(i) - e_J||_2 \tag{15}$$

Ultimately, $z_q(i)$ is passed through the decoder and becomes the output, $z_e(i)$. Our model utilizes gradients to facilitate the updating of values in Equation (15), which impacts the encoder's discretization and subsequently affects the final output. As the encoder and decoder share the same D-dimensional space, gradients also help adjust the encoder's output to decrease the reconstruction loss.

The complete loss function consists of three components. The first part L_q represents the specific machine learning task service quality loss. In this work, we focused on facial semantics (facial expression). In each iteration of training, the expression of generated images is recorded. To make sure the generated expression is the same as the original picture's label, the task-related quality loss, L_q, is determined using Equation (16). If the output's facial expression recognition result matches the label, it indicates that features related to facial semantics are maintained and L_q is set to 0. Alternatively, if they do not match, L_q is set to 1.

$$L_q = \begin{cases} 0 & if\ E(i) = E(z(i)) \\ 1 & if\ E(i) \neq E(z(i)) \end{cases} \tag{16}$$

In the next part, L_{m_1} quantifies the resemblance between the output images and the original images. The initial component of the function delineates the loss of image reconstruction between the original images and the corresponding identity-concealed images generated by the proposed model. The intention of the second component is to enhance the vector quantization embedding space by continuously updating the dictionary throughout the model training process. Its function is to slow down the update rate of the encoder, keeping the generated output close to the embedding vector.

$$L_{m_1} = logp(i|z_q(i)) + ||sg[z_e(i)] - e||_2^2 + \beta_{m_1}||z_e(i) - sg[e]||_2^2 \tag{17}$$

where sg represents the stop gradient, stopping the gradient flow from flowing through specific parts of the network, and is defined as a constant value in forward computing of the whole model. Changes in the parameter β_{m_1} ranging from 0.1 to 2.0 do not significantly affect the outcome.

Similar to L_{m_1}, the third part L_{m_2} shown in Equation (18) below calculates the degree of resemblance of faces contained in two images. The difference is that L_{m_2} is used to describe the reconstruction loss between output images and the corresponding surrogate facial images generated from the same cluster.

$$L_{m_2} = logp(i|f(i)) + ||sg[f(i)] - e||_2^2 + \beta_{m_2}||f(i) - sg[e]||_2^2 \tag{18}$$

where β_{m_2} ranges from 0.1 to 2.0.

The whole loss function L is defined below:

$$L = \alpha L_q + (1 - \alpha)((1 - \beta)L_{m_1} + \beta L_{m_2}) \tag{19}$$

where α is utilized to determine the relative weight of the semantic-related loss and the image fusion loss in the overall loss function and β is used to allocate a weight that corresponds to the relative importance of the original images and the surrogate images. The values of both α and β range from 0 to 1. In our experiment, we adjusted the values of both α and β to analyze the corresponding quality of service. The specific results are presented in Section 5.

4.3. Method Execution Process

Given the concept of concealing the privacy contained in the image and subsequently restoring it with preserved utility, we drew Figure 2 to illustrate the whole process of our approach.

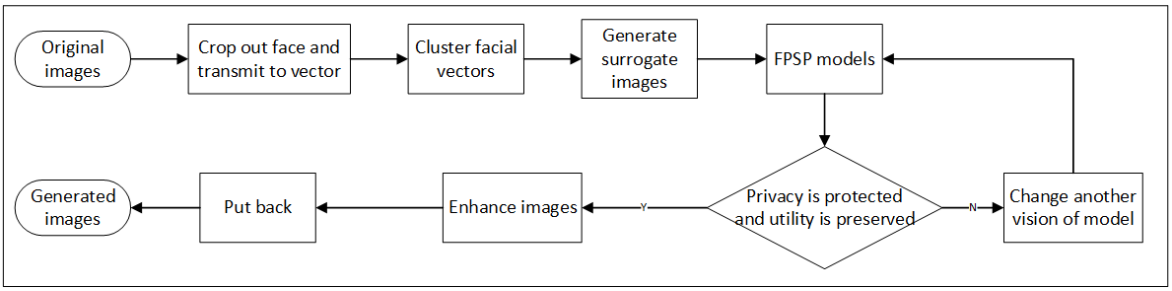

Figure 2. The workflow of the FPSP model.

In the beginning, we carry out the process of identification of facial regions on the original image in the dataset to crop out the face area. Then, the cropped face is transformed into facial embedding and classified into one cluster. Furthermore, we utilize its corresponding centroid to synthesize surrogate images. The surrogate face and four de-identified cropped faces modified by four face methods (covering the eyes, blurring the face, adding Laplace noise, and transforming into a cartoon face) are fed into the model to reconstruct the image that conceals identity while preserving utility.

The semantic-related estimator takes advantage of the difference in attributes within the serving scene before and after processing to quantify the service quality loss of images. It adjusts the manner in which the whole model is trained with the intention of the model generating facial images that fulfil the goal of privacy removal and utility preservation, supplying reliable and secure facial images that are suitable for use in the facial dataset that is related to specific machine-learning tasks.

5. Experimentation and Results

5.1. Experimental Setup

The proposed method for identity concealing in the dataset was executed on the CelebA [7] dataset. We chose part of the original dataset, forming a dataset containing 1000 female images, which were divided into six different categories based on the expression of the subject. The expression distribution of facial images in the dataset is shown in Figure 3. In the training process, the first 800 images were employed, leaving the remaining 200 for testing. The generated images of our proposed approach were then applied to three existing recognition algorithms, namely RCNN [9], DACL [10], and DAN [11], to evaluate the utility of generated dataset in the recognition of facial expressions. The validation results of these methods are presented in Section 5.4 of the paper.

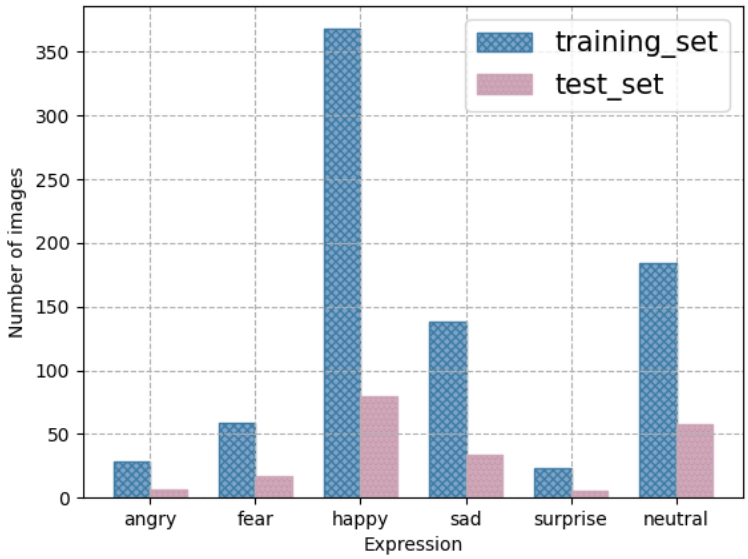

Figure 3. The expression distribution of the dataset.

The FPSP model was implemented based on the TensorFlow and Keras frameworks and trained on an RTX 3080 GPU, significantly reducing the training time compared to traditional training. According to paper [67], the β_{m_1} and β_{m_2} parameters were set to 0.25, as when the values are in the range of 0.1 to 2.0, the results are not significantly different. By using the ADAM optimizer with a learning [72] rate of 1×10^{-3}, the model can effectively update the weights of the neural network during training, improving its accuracy. Other parameters included in the model are listed in Table 1.

Table 1. Model parameters and settings.

Parameter Name	Brief Description	Range of Value
k	Allocate the relative weight to be given to the image coordinates and their corresponding centroid image coordinate	[0, 1]
t	Control the proportional weight of the image region and the corresponding centroid image region to synthesize a surrogate image	[0, 1]
β_{m_1}	Commitment loss in $Ł_{m_1}$, ensuring that the encoder is dedicated to the embedding	[0.1, 2]
β_{m_2}	Commitment loss in $Ł_{m_2}$, making sure the encoder commits to the embedding	[0.1, 2]
α	Determine the appropriate balance between the weight assigned to the loss of service and the loss of image fusion	[0, 1]
β	Control the proportional weight of original images and surrogate images	[0, 1]

5.2. Privacy Preservation Evaluation

We established a dataset consisting of original, surrogate, and de-identified images for training the model. By including a surrogate and four de-identified images in the dataset, we aimed to improve the model's ability to conceal facial privacy. Considering the protection performance of the model, in this training process, we set α in Equation (19) to 0.

During training, we experimented with different values of β to minimize the loss, after which we checked the maintenance quality loss to choose the most suitable percentage of β. The graph depicted in Figure 4 illustrates the change in the loss rate with different values of β.

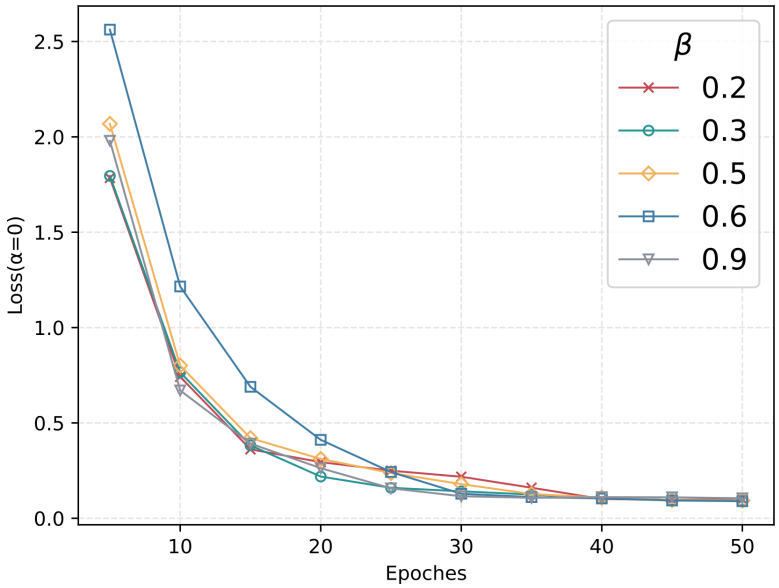

Figure 4. The changing trend of the loss rate when $\alpha = 0$.

As Figure 4 depicts, when β is about 0.2, the effort of the model is much better, and after 50 rounds, the loss rate is reduced to 0.09 which is less than other proportions. This represented a significant reduction of at least 2.4% compared to the initial loss rate.

To verify the effect of privacy preservation of our method, we conducted re-identification experiments that evaluate the risk of successfully identifying the subject contained in the input dataset utilizing de-identified facial images in the generated dataset. Drawing on the k-anonymity theory, we consider that when the generated image I is more similar to other original images in the same cluster compared to its original image, I meets the criteria.

Specifically, if the cosine similarity between I and its corresponding original image is larger than the smallest value of similarity calculated between I and other authentic images in the same cluster, the privacy of the image is protected. Let $N_{protected}$ be the number of privacy-protected images generated and N_{total} represent the total number of images in the dataset. The privacy protection rate, denoted as "RP", is calculated according to Equation (20). As the proportion of the surrogate image increases, generated images have a much greater chance of not being identified from the original cluster. We measured the percentage rate for privacy-preserved images, respectively.

$$RP = \frac{N_{protected}}{N_{total}} \tag{20}$$

As indicated in Figure 5, overall, there was a significant upward trend, which indicates that the effect of the model on image privacy protection improves. The highest percentage is up to 84% when the proportion of surrogate images is 80%. It can be concluded that the identity of images is obviously concealed after our method, achieving privacy protection.

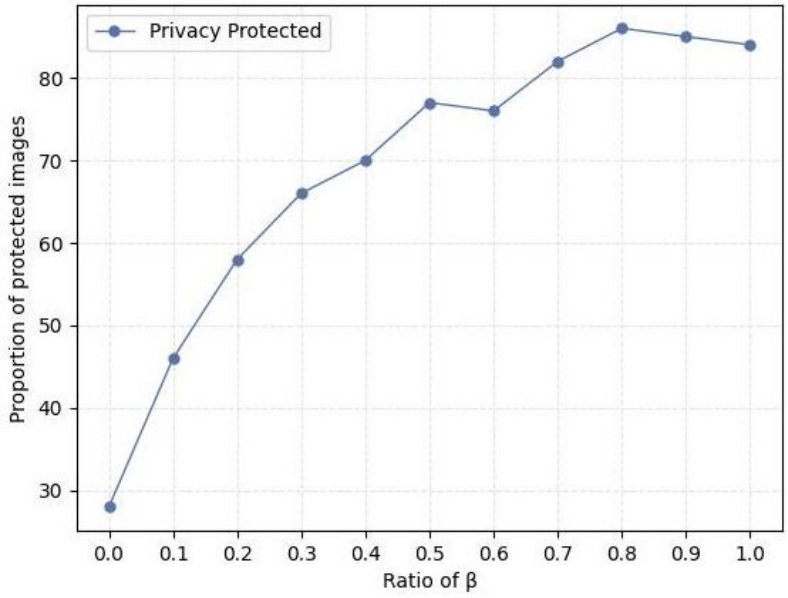

Figure 5. The percentage of privacy-protected images.

To preserve the privacy of facial images in the dataset, we utilized models with various values of β. Initially, we employed the model with a relatively small value of β for protection. If the protection is successful, the resulting images are directly incorporated into the output set. In cases where privacy has not been adequately preserved, the β value is increased to prompt the model to regenerate the output and reassess the outcome. The ratios of the cumulative number of protected images to the total number of images at different values of β are depicted in Table 2.

As shown in Table 2, in general, the privacy of the majority of images is effectively protected when employing a relatively small value of β. In the last few images, as there are no facial features of the original face involved in fusion during the synthetic process

(i.e., they are completely fake images), the probability of being recognized as any face in the cluster is equal. Thus, they were not included in the statistics.

Table 2. The ratio of the cumulative number of privacy-protected images in different models.

	0.2	**0.4**	**0.5**	**0.6**	**0.8**	**1**
Ratio of protected images	58.8%	79.2%	87.4%	92.0%	96.2%	98.5%

In Figure 6, the top row presents the original pictures, the middle row or rows exhibit the model-generated images, and the last row showcases the images that have been identified as belonging to the same person as the corresponding original picture. Notably, the six images on the left side of the figure demonstrate successful privacy protection when the parameter β is set to 0.2. In contrast, the three images on the right side present cases where the identity is not concealed in the second row when β takes the same value, while the third row reveals the successful concealment of identity when the value of β increases, i.e., the proportion of surrogate images is increased.

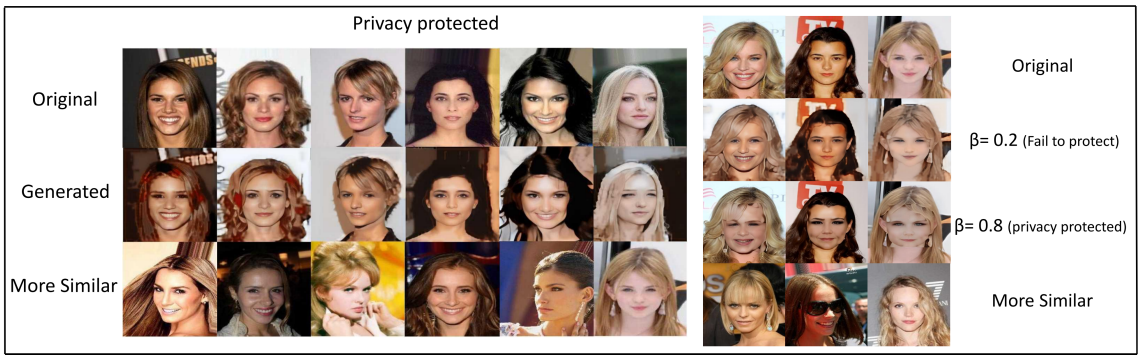

Figure 6. Demonstration of de-identified images.

5.3. Images' Utility Maintenance

Aiming to improve the quality of generated images of the model, i.e., to obtain the minimum loss rate to image quality, we adjusted the key parameter α, which allocates the proportion of the quality loss, while keeping β fixed at 0.2, and checked the model effect of quality maintenance to each value of α, respectively. The downward trend of quality loss is presented in Figure 7.

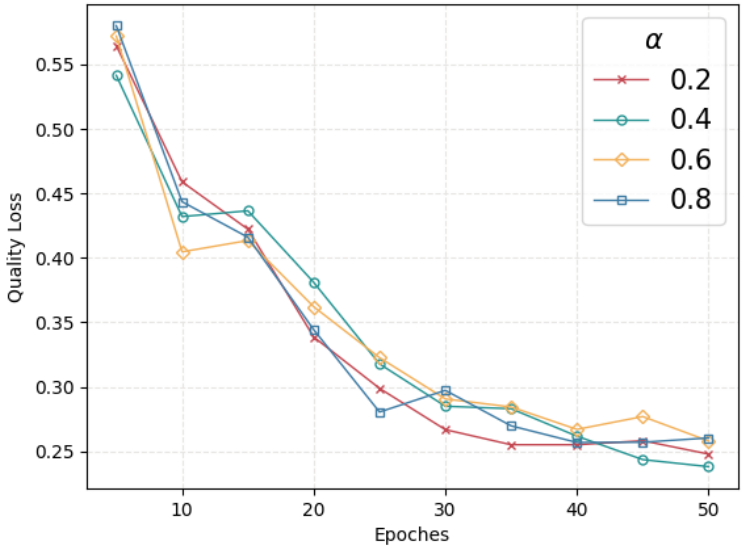

Figure 7. The changing trend of expression loss rate.

As the figure above illustrates, with a value of 0.4 of α, the model yielded the best output. After 50 epochs, the loss was reduced to approximately 0.23, which was the lowest among other types of α percent. This achieved a semantic retention rate of 77%, demonstrating the efficacy of the proposed approach in preserving the semantics of images while de-identifying. In general, using an α value of 0.4 for the output is more beneficial in terms of training iterations as compared to using a simple image fusion approach. This suggests that within the overall loss function, semantic loss plays a crucial role in maintaining the quality of the generated images.

Our method contains expression loss, which plays an important role in preserving semantics, outperforming other face privacy protection models that do not take the quality of facial semantics (facial expression) into consideration such as AMT-GAN [8]. While AMT-GAN had a loss rate of 0.37, the proposed model only lost 0.23, demonstrating a better performance overall. The results are presented in Figure 8, further supporting the effectiveness of the FPSP model.

Figure 8. Representation of output images generated by the FPSP model and AMT-GAN.

We performed the following analyses under an α value of 0.4. Let $N_{unchanged}$ be the number of processed images with unchanged expression labels and $N_{category}$ be the total number of images in the dataset. The loss rate of six expressions that underwent processing using different methods is computed using Equation (21). The results of this analysis are presented in Table 3, demonstrating their respective impacts on the loss rate of the expressions.

$$Loss\ rate = \frac{N_{unchanged}}{N_{category}} \tag{21}$$

Table 3. Facial expression loss rate under different methods.

Loss Rate	Angry	Fear	Happy	Sad	Surprise	Neutral
Our Model	0.235	0.342	0.136	0.273	0.429	0.231
Blindfold	0.441	0.737	0.029	0.669	0.893	0.814
Cartoon	0.794	0.513	0.489	0.360	0.785	0.459
Laplace	0.911	0.895	0.665	0.756	0.964	0.702
Mosaic	0.294	0.302	0.299	0.361	0.321	0.169
AMTGAN	0.625	0.764	0.120	0.635	0.740	0.536

Upon analysis of the results presented in Table 3, it becomes apparent that the proposed model was able to maintain the service quality of each expression to a certain degree, highlighting that our model is a practical approach that can be applied to the dataset for protecting privacy while preserving utility. The slightly elevated rate of the fear group and surprise group was due to the limited size of the sample.

We computed the disparity between the values of the facial expression recognition outcomes for the original images and those processed by six different methods. Subsequently, we generated an error bar plot to represent the results, which is presented in Figure 9. As

depicted in the diagram, our model demonstrates remarkable proficiency in maintaining efficacy across a wide range of facial expression recognition results, showing a higher level of consistency and an exceptional ability to preserve efficacy when faced with a diverse set of outcomes.

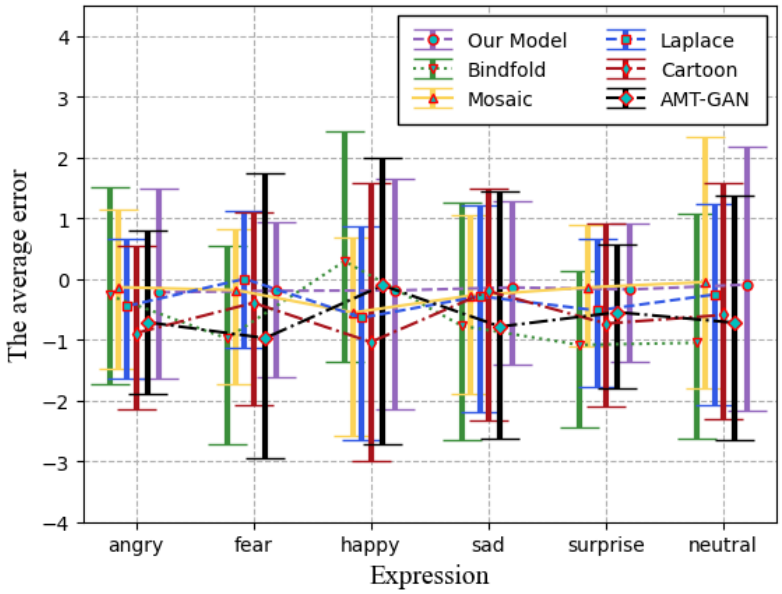

Figure 9. The error bar diagrams of output images processed by different methods.

5.4. Evaluation of the De-Identified Facial Image Dataset for Machine Learning

In the problem of the utility of generated processed dataset, our method aims to retain information related to quality, maintain the similarity between the original image and the processed image as much as possible, and erase only privacy-related information, which can ultimately benefit the effectiveness of the image.

As an illustration of the effectiveness of our method, we used facial expression recognition as a service and employed three different de-identification methods: direct output images without post-processing (DOP), images with post-processing (GFP) using the GFP-GAN technique [6] for facial enhancement, and images with relabels (RL) to validate the usability of the proposed model. To evaluate the performance of our method, we utilized three state-of-the-art facial expression algorithms, RTCNN [9], DACL [10], and DAN [11], and selected the best quality-preserving model that was trained for 50 rounds with $\alpha = 0.4$ for analysis.

For the dataset processed using DOP and GFP, we employed the aforementioned facial expression recognition algorithms to detect and compare the results with the original dataset labels. If they matched, the semantics of the face were considered preserved. The number of images with matching labels was compared to the total number of images to obtain a value for utility preservation. For RL, we first performed facial expression recognition on the processed dataset to obtain new labels. We then used the new labels and processed images as inputs to retrain the facial expression recognition algorithm, followed by testing the testing set for facial expression recognition. The number of images with matching labels of the testing set was then calculated and compared to obtain the value. Let $N_{matched}$ be the number of generated images with matching labels and N_{total} represent the total number of images in the dataset. The preserved utility PU is calculated according to Equation (22). The results are shown in Table 4.

$$PU = \frac{N_{matched}}{N_{total}} \tag{22}$$

Table 4. Usability of the processed dataset under different de-identification methods.

	DOP	DOP + RL	DOP + GFP	DOP + GFP + RL	Original
RTCNN	0.43	0.40	0.48	0.45	0.53
DACL	0.57	0.58	0.59	0.59	0.59
DAN	0.40	0.59	0.47	0.49	0.61

The data presented in Table 4 reveal that the accuracy rates for most of the above methods utilizing the generated dataset are similar to the accuracy rate achieved on the original dataset. This indicates that our method is successful in protecting the utility of images with a concealed identity.

Figure 10 displays a demonstration of our model on street pictures captured by unmanned aerial vehicles that include lots of faces of pedestrians. This showcases the complete process of maintaining utility while prioritizing privacy protection, as implemented in our model, providing a clear representation of our model's capability of utility maintenance and privacy protection.

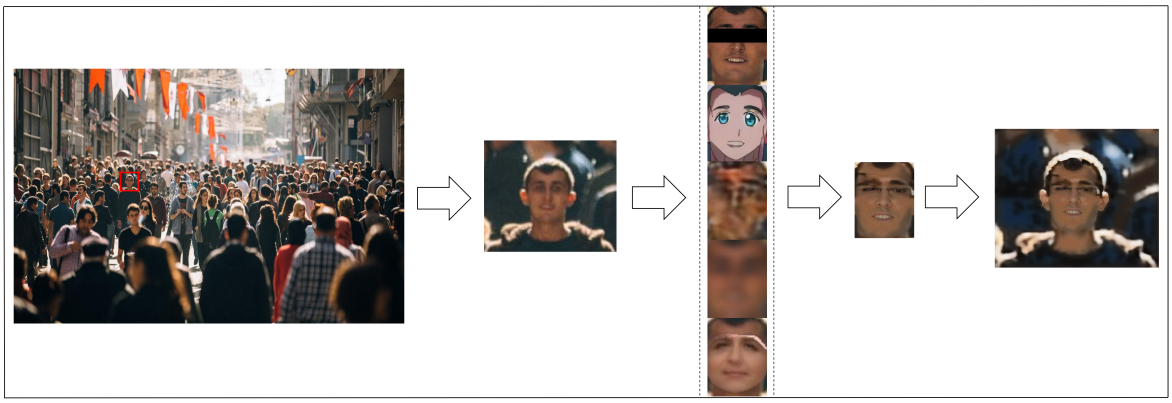

Figure 10. A demo of the FPSP model.

6. Discussion

As the experimental data show above, the model we propose is capable of obscuring identities in the dataset and preserving the semantic information lost during protection, which is achieved by the utilization of appropriate approaches used at every stage. Additionally, the paper introduces the concept of de-identification in machine learning datasets, which includes preserving task-related features. The experiments validate that our model can quantitatively process a dataset that is available for subsequent machine learning. Our approach is the first to provide a general solution for identity concealing in facial datasets used in modern machine learning, offering great flexibility and efficiency.

We provide the results of the quantitative experiments that assess the efficacy of the model in concealing identity information in the dataset used for machine learning tasks. However, as the proportion of surrogate faces increases, the quality of the generated images is somewhat degraded, leading to the perception of falseness when visually inspected. In future research, we aim to address this limitation by incorporating advanced network architectures and collecting a richer and more diverse dataset of surrogate faces, reducing the sense of falseness and providing more accurate and high-quality results in visual.

7. Conclusions

In this paper, we introduce the concept of de-identification of datasets for machine learning while preserving service-related features, and present a novel framework named the FPSP model that aims to simultaneously maintain privacy and semantic fidelity. To achieve this goal, we leverage surrogate images generated by the centroid of the facial cluster in the input and adjust the semantic-related loss in the loss function. The efficiency

of the model in preserving facial semantics has been demonstrated to be as high as 77%. Extensive experiments show that when applied to facial expression recognition, the performance of the identity concealed dataset processed by our model is comparable to the original recognition rate. Our future work will explore the application of our approach to more complex and less constrained machine learning datasets, as well as how to apply our model to more intricate scenarios while achieving relatively good preservation of utility and protection of privacy.

Author Contributions: Conceptualization, B.S.; methodology, Y.Y., Z.N., Y.Q. and B.S.; software, Y.Y. and Z.N.; validation, Y.Y. and Y.Q.; formal analysis, Z.N.; investigation, Y.Q.; resources, B.S.; data curation, Y.Y., Z.N. and Y.Q.; writing—original draft preparation, Y.Y. and Y.Q.; writing—review and editing, Y.Y., Z.N, Y.Q, B.S., X.Z. and Y.T.; visualization, Y.Y.; supervision, B.S.; project administration, B.S. All authors have read and agreed to the published version of the manuscript.

Funding: The authors extend their appreciation to National Key Research and Development Program of China (International Technology Cooperation Project No. 2021YFE014400) and the National Science Foundation of China (No. 42175194) for funding this work.

Data Availability Statement: Not applicable.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Qiu, Y.; Niu, Z.; Song, B.; Ma, T.; Al-Dhelaan, A.; Al-Dhelaan, M. A Novel Generative Model for Face Privacy Protection in Video Surveillance with Utility Maintenance. *Appl. Sci.* **2022**, *12*, 6962. [CrossRef]
2. Maximov, M.; Elezi, I.; Leal-Taixé, L. Ciagan: Conditional identity anonymization generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5447–5456.
3. Cao, J.; Liu, B.; Wen, Y.; Xie, R.; Song, L. Personalized and invertible face de-identification by disentangled identity information manipulation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 3334–3342.
4. Sun, Q.; Ma, L.; Oh, S.J.; Van Gool, L.; Schiele, B.; Fritz, M. Natural and effective obfuscation by head inpainting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5050–5059.
5. Chibelushi, C.C.; Bourel, F. Facial expression recognition: A brief tutorial overview. In *CVonline: On-Line Compendium of Computer Vision*; Southern Methodist University: Dallas, TX, USA, 2003; Volume 9.
6. Wang, X.; Li, Y.; Zhang, H.; Shan, Y. Towards real-world blind face restoration with generative facial prior. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 9168–9178.
7. Liu, Z.; Luo, P.; Wang, X.; Tang, X. Large-scale celebfaces attributes (celeba) dataset. *Retrieved August* **2018**, *15*, 11.
8. Hu, S.; Liu, X.; Zhang, Y.; Li, M.; Zhang, L.Y.; Jin, H.; Wu, L. Protecting facial privacy: Generating adversarial identity masks via style-robust makeup transfer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 15014–15023.
9. Arriaga, O.; Valdenegro-Toro, M.; Plöger, P. Real-time convolutional neural networks for emotion and gender classification. *arXiv* **2017**, arXiv:1710.07557.
10. Farzaneh, A.H.; Qi, X. Facial expression recognition in the wild via deep attentive center loss. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2021; pp. 2402–2411.
11. Wen, Z.; Lin, W.; Wang, T.; Xu, G. Distract your attention: Multi-head cross attention network for facial expression recognition. *arXiv* **2021**, arXiv:2109.07270.
12. Pan, Z.; Yu, W.; Lei, J.; Ling, N.; Kwong, S. TSAN: Synthesized view quality enhancement via two-stream attention network for 3D-HEVC. *IEEE Trans. Circuits Syst. Video Technol.* **2021**, *32*, 345–358. [CrossRef]
13. Peng, B.; Lei, J.; Fu, H.; Jia, Y.; Zhang, Z.; Li, Y. Deep video action clustering via spatio-temporal feature learning. *Neurocomputing* **2021**, *456*, 519–527. [CrossRef]
14. Lei, J.; Li, X.; Peng, B.; Fang, L.; Ling, N.; Huang, Q. Deep spatial-spectral subspace clustering for hyperspectral image. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *31*, 2686–2697. [CrossRef]
15. Crowley, J.L.; Coutaz, J.; Bérard, F. Perceptual user interfaces: Things that see. *Commun. ACM* **2000**, *43*, 54-ff. [CrossRef]
16. Hudson, S.E.; Smith, I. Techniques for addressing fundamental privacy and disruption tradeoffs in awareness support systems. In Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work, Boston, MA, USA, 16–20 November 1996; pp. 248–257.
17. Neustaedter, C.G. *Balancing Privacy and Awareness of Home Media Spaces*; University of Calgary: Calgary, AB, Canada, 2003.
18. Ribaric, S.; Ariyaeeinia, A.; Pavesic, N. De-identification for privacy protection in multimedia content: A survey. *Signal Process. Image Commun.* **2016**, *47*, 131–151. [CrossRef]

19. Boyle, M.; Edwards, C.; Greenberg, S. The effects of filtered video on awareness and privacy. In Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work, Philadelphia, PA, USA, 2–6 December 2000; pp. 1–10.
20. Neustaedter, C.; Greenberg, S.; Boyle, M. Blur filtration fails to preserve privacy for home-based video conferencing. *ACM Trans. Comput.-Hum. Interact. (TOCHI)* **2006**, *13*, 1–36. [CrossRef]
21. Gross, R.; Airoldi, E.; Malin, B.; Sweeney, L. Integrating utility into face de-identification. In Proceedings of the Privacy Enhancing Technologies: 5th International Workshop, PET 2005, Cavtat, Croatia, 30 May–1 June 2005; Revised Selected Papers 5; Springer: Berlin/Heidelberg, Germany, 2006; pp. 227–242.
22. Gross, R.; Sweeney, L.; De la Torre, F.; Baker, S. Model-based face de-identification. In Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06), New York, NY, USA, 17–22 June 2006; p. 161.
23. Liu, J.; Yin, S.; Li, H.; Teng, L. A Density-based Clustering Method for K-anonymity Privacy Protection. *J. Inf. Hiding Multim. Signal Process.* **2017**, *8*, 12–18. [CrossRef]
24. Gross, R.; Sweeney, L.; De La Torre, F.; Baker, S. Semi-supervised learning of multi-factor models for face de-identification. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.
25. Samarzija, B.; Ribaric, S. An approach to the de-identification of faces in different poses. In Proceedings of the 2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 26–30 May 2014; pp. 1246–1251.
26. Sweeney, L. k-anonymity: A model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* **2002**, *10*, 557–570. [CrossRef]
27. Newton, E.M.; Sweeney, L.; Malin, B. Preserving privacy by de-identifying face images. *IEEE Trans. Knowl. Data Eng.* **2005**, *17*, 232–243. [CrossRef]
28. Burkhardt, H.; Neumann, B. Computer Vision—ECCV'98. In Proceedings of the 5th European Conference on Computer Vision, Freiburg, Germany, 2–6 June 1998; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1998; Volume 1.
29. Sun, Z.; Meng, L.; Ariyaeeinia, A. Distinguishable de-identified faces. In Proceedings of the 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia, 4–8 May 2015; Volume 4, pp. 1–6.
30. Meng, L.; Sun, Z.; Ariyaeeinia, A.; Bennett, K.L. Retaining expressions on de-identified faces. In Proceedings of the 2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 26–30 May 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 1252–1257.
31. Phillips, P.J.; Wechsler, H.; Huang, J.; Rauss, P.J. The FERET database and evaluation procedure for face-recognition algorithms. *Image Vis. Comput.* **1998**, *16*, 295–306. [CrossRef]
32. Meden, B.; Emeršič, Ž.; Štruc, V.; Peer, P. k-Same-Net: k-Anonymity with generative deep neural networks for face deidentification. *Entropy* **2018**, *20*, 60. [CrossRef]
33. Chuanlu, L.; Yicheng, W.; Hehua, C.; Shuliang, W. Utility Preserved Facial Image De-identification Using Appearance Subspace Decomposition. *Chin. J. Electron.* **2021**, *30*, 413–418. [CrossRef]
34. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [CrossRef]
35. Cai, Z.; Xiong, Z.; Xu, H.; Wang, P.; Li, W.; Pan, Y. Generative adversarial networks: A survey toward private and secure applications. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–38. [CrossRef]
36. Han, C.; Xue, R. Differentially private GANs by adding noise to Discriminator's loss. *Comput. Secur.* **2021**, *107*, 102322. [CrossRef]
37. Yang, R.; Ma, X.; Bai, X.; Su, X. Differential privacy images protection based on generative adversarial network. In Proceedings of the 2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), Guangzhou, China, 29 December 2020–1 January 2021; pp. 1688–1695.
38. Hukkelås, H.; Mester, R.; Lindseth, F. Deepprivacy: A generative adversarial network for face anonymization. In Proceedings of the Advances in Visual Computing: 14th International Symposium on Visual Computing, ISVC 2019, Lake Tahoe, NV, USA, 7–9 October 2019; Proceedings, Part I 14; Springer: Berlin/Heidelberg, Germany, 2019; pp. 565–578.
39. Qi, G.J. Loss-sensitive generative adversarial networks on lipschitz densities. *Int. J. Comput. Vis.* **2020**, *128*, 1118–1140. [CrossRef]
40. Chen, J.; Konrad, J.; Ishwar, P. Vgan-based image representation learning for privacy-preserving facial expression recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 1570–1579.
41. Louizos, C.; Swersky, K.; Li, Y.; Welling, M.; Zemel, R. The variational fair autoencoder. *arXiv* **2015**, arXiv:1511.00830.
42. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.
43. Liu, Y.; Peng, J.; James, J.; Wu, Y. PPGAN: Privacy-preserving generative adversarial network. In Proceedings of the 2019 IEEE 25Th international conference on parallel and distributed systems (ICPADS), Tianjin, China, 4–6 December 2019; pp. 985–989.
44. Li, Y.; Lu, Q.; Tao, Q.; Zhao, X.; Yu, Y. SF-GAN: Face de-identification method without losing facial attribute information. *IEEE Signal Process. Lett.* **2021**, *28*, 1345–1349. [CrossRef]
45. Nguyen, H.; Zhuang, D.; Wu, P.Y.; Chang, M. Autogan-based dimension reduction for privacy preservation. *Neurocomputing* **2020**, *384*, 94–103. [CrossRef]

46. Lin, J.; Li, Y.; Yang, G. FPGAN: Face de-identification method with generative adversarial networks for social robots. *Neural Netw.* **2021**, *133*, 132–147. [CrossRef]
47. Dwork, C. Differential privacy: A survey of results. In Proceedings of the Theory and Applications of Models of Computation: 5th International Conference, TAMC 2008, Xi'an, China, 25–29 April 2008; Proceedings 5; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1–19.
48. Dwork, C.; Roth, A. The algorithmic foundations of differential privacy. *Found. Trends® Theor. Comput. Sci.* **2014**, *9*, 211–407. [CrossRef]
49. Yu, J.; Xue, H.; Liu, B.; Wang, Y.; Zhu, S.; Ding, M. Gan-based differential private image privacy protection framework for the internet of multimedia things. *Sensors* **2020**, *21*, 58. [CrossRef]
50. Zhou, G.; Qin, S.; Zhou, H.; Cheng, D. A differential privacy noise dynamic allocation algorithm for big multimedia data. *Multimed. Tools Appl.* **2019**, *78*, 3747–3765. [CrossRef]
51. Friedman, J.H.; Bentley, J.L.; Finkel, R.A. An algorithm for finding best matches in logarithmic expected time. *ACM Trans. Math. Softw. (TOMS)* **1977**, *3*, 209–226. [CrossRef]
52. Zhao, C.; Zhao, S.; Zhao, M.; Chen, Z.; Gao, C.Z.; Li, H.; Tan, Y.A. Secure multi-party computation: Theory, practice and applications. *Inf. Sci.* **2019**, *476*, 357–372. [CrossRef]
53. Chamikara, M.A.P.; Bertok, P.; Khalil, I.; Liu, D.; Camtepe, S. Privacy preserving face recognition utilizing differential privacy. *Comput. Secur.* **2020**, *97*, 101951. [CrossRef]
54. Nasr, M.; Shokri, R.; Houmansadr, A. Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning. In Proceedings of the 2019 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 19–23 May 2019; pp. 739–753.
55. Shokri, R.; Stronati, M.; Song, C.; Shmatikov, V. Membership inference attacks against machine learning models. In Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP), San Jose, CA, USA, 22–26 May 2017; pp. 3–18.
56. Fredrikson, M.; Jha, S.; Ristenpart, T. Model inversion attacks that exploit confidence information and basic countermeasures. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, Denver, CO, USA, 12–16 October 2015; pp. 1322–1333.
57. Xue, W.; Hu, W.; Gauranvaram, P.; Seneviratne, A.; Jha, S. An efficient privacy-preserving IoT system for face recognition. In Proceedings of the 2020 Workshop on Emerging Technologies for Security in IoT (ETSecIoT), Sydney, NSW, Australia, 21 April 2020; pp. 7–11.
58. Qardaji, W.; Yang, W.; Li, N. Differentially private grids for geospatial data. In Proceedings of the 2013 IEEE 29th International Conference on Data Engineering (ICDE), Brisbane, QLD, Australia, 8–12 April 2013; pp. 757–768.
59. Zhang, X.; Ji, S.; Wang, T. Differentially private releasing via deep generative model (technical report). *arXiv* **2018**, arXiv:1801.01594.
60. Kim, T.; Yang, J. Selective feature anonymization for privacy-preserving image data publishing. *Electronics* **2020**, *9*, 874. [CrossRef]
61. Noble, B.; Daniel, J.W. *Applied Linear Algebra*; Prentice-Hall Englewood Cliffs: Englewood Cliffs, NJ, USA, 1977; Volume 477.
62. Chai, X.; Shan, S.; Gao, W. Pose normalization for robust face recognition based on statistical affine transformation. In Proceedings of the Fourth International Conference on Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint, Singapore, 15–18 December 2003; Volume 3, pp. 1413–1417.
63. Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; Aila, T. Analyzing and improving the image quality of stylegan. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 8110–8119.
64. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Change Loy, C. Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018.
65. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Process. Lett.* **2016**, *23*, 1499–1503. [CrossRef]
66. Haghighat, M.; Abdel-Mottaleb, M.; Alhalabi, W. Fully automatic face normalization and single sample face recognition in unconstrained environments. *Expert Syst. Appl.* **2016**, *47*, 23–34. [CrossRef]
67. Van Den Oord, A.; Vinyals, O.; Kavukcuoglu, K. Neural discrete representation learning. In *Advances in Neural Information Processing Systems*; NIPS: Long Beach, CA, USA, 2017; Volume 30.
68. Schroff, F.; Kalenichenko, D.; Philbin, J. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 815–823.
69. LeCun, Y.; Boser, B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* **1989**, *1*, 541–551. [CrossRef]
70. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [CrossRef]

71. King, D.E. Dlib-ml: A machine learning toolkit. *J. Mach. Learn. Res.* **2009**, *10*, 1755–1758.
72. Bock, S.; Weiß, M. A proof of local convergence for the Adam optimizer. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–8.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.