BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

ADAPTIVE HYBRID FRAMEWORK FOR FACIAL PRIVACY PROTECTION AGAINST UNAUTHORIZED FACE RECOGNITION

Mohammadi Arif Pathan

Student Researcher
Department of Artificial Intelligence and Data Science
Ajeenkya DY Patil School of Engineering Pune, India

Prof. Supriya Survase

Project Guide
Department of Artificial Intelligence and Data Science
Ajeenkya DY Patil School of Engineering
Pune, India

Abstract

Facial recognition technology has become an important component of modern computer vision systems and is widely used for authentication, surveillance, access control, social media applications, and identity verification. However, the increasing use of automated facial recognition creates significant privacy risks when facial images are collected, processed, or shared without adequate user consent. Existing facial privacy protection approaches include adversarial perturbations, image transformations, semantic manipulation, personalised masks, adversarial makeup, and diffusion-based techniques. Although these methods can reduce the effectiveness of facial recognition, they may suffer from limitations related to image quality, transformation robustness, recognition model dependence, computational cost, and generalisation.

This paper presents a conceptual adaptive hybrid framework for protecting facial images against unauthorised face recognition. The proposed framework combines semantic perturbation, adaptive pixel-level noise, and adaptive image filtering. The approach first analyses the facial image and its important features, then applies semantic-level modifications followed by content-aware pixel perturbation and image filtering. The objective is to reduce the recognizability of the protected image while maintaining acceptable visual quality. The paper reviews existing facial privacy protection approaches, identifies their limitations, and discusses an integrated framework that can adapt protection according to image characteristics and recognition threats. The proposed framework is intended as a conceptual research direction for privacy-preserving computer vision and does not claim a completed implementation or experimental accuracy result.

Keywords: Adversarial Machine Learning, Adaptive Image Filtering, Artificial Intelligence, Computer Vision, Deep Learning, Diffusion Models, Facial Privacy Protection, Facial Recognition, Image Processing, Privacy-Preserving AI, Semantic Perturbation

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

1 Introduction

Facial recognition has developed rapidly with advances in artificial intelligence, deep learning, and computer vision. Modern recognition systems can extract facial features from an image and compare them with previously stored representations to identify or verify an individual. These systems are used in applications such as mobile authentication, access control, surveillance, digital identity systems, and social media platforms.

Although facial recognition provides useful functionality, its increasing deployment also creates important privacy concerns. Facial images are biometric information because they can be associated with the identity of a person. Unlike a password, a face cannot simply be replaced after it has been compromised. Therefore, unauthorized collection and processing of facial images can create long-term privacy risks.

The problem becomes more significant with the widespread availability of online image-sharing platforms. A person may upload an image for social, professional, or personal purposes without knowing whether the image may later be collected by an automated recognition system. In such situations, traditional security mechanisms such as passwords are not sufficient because the privacy risk exists within the visual information itself.

Facial privacy protection therefore aims to modify facial images before they are shared or published. Adversarial methods such as Fawkes and LowKey demonstrate that carefully designed modifications can interfere with facial recognition systems while attempting to preserve the visual appearance of the original image [5], [6]. Other approaches investigate adversarial makeup, personalized masks, semantic editing, and diffusion-based protection [?], [1]–[4].

However, existing approaches do not completely solve the problem. A privacy protection method may work against one recognition architecture but perform poorly against another. Similarly, image compression, resizing, cropping, format conversion, or other transformations can reduce the effectiveness of an applied perturbation.

This paper therefore investigates an adaptive hybrid approach in which multiple protection mechanisms are combined. The proposed framework uses semantic perturbation, adaptive pixel-level noise, and adaptive image filtering to improve the flexibility of facial privacy protection.

The paper is a conceptual research study and does not claim a completed real-world implementation, trained model, dataset evaluation, or numerical privacy-protection result. Instead, it reviews existing research and proposes an integrated framework that can be investigated experimentally in future work.

2 Research Questions and Objectives

The research is guided by the following questions:

1) How does modern facial recognition technology identify individuals from facial images?
2) What techniques are currently used to protect facial images from unauthorized recognition?
3) What are the major limitations of adversarial and image-based facial privacy protection methods?
4) How can semantic perturbation improve facial privacy protection?

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

5) How can adaptive pixel-level noise be used according to facial image characteristics?
6) How can adaptive image filtering complement semantic and pixel-level protection?
7) How can a hybrid framework balance privacy, robustness, image quality, and computational cost?

The major objectives of this research are:

1) To study facial recognition technology and its privacy implications.
2) To understand existing facial privacy protection techniques.
3) To analyse the limitations of current privacy preserving methods.
4) To study semantic perturbation for facial privacy protection.
5) To investigate adaptive pixel-level noise for facial images.
6) To investigate adaptive image filtering based on image content.
7) To propose a conceptual hybrid framework combining multiple protection mechanisms.
8) To identify suitable evaluation parameters for privacy, robustness, image quality, and computational efficiency.

3 Facial Recognition and Privacy

Facial recognition systems generally perform a sequence of operations including face detection, preprocessing, feature extraction, and identity comparison. A facial image is first processed to identify the region containing a face. A recognition model then converts the facial region into a feature representation. This representation can subsequently be compared with stored representations to determine whether two facial images belong to the same person.

The accuracy and effectiveness of modern recognition systems have increased because of deep learning architectures. Convolutional neural networks and other representation learning models can learn complex patterns associated with facial identity.

The same capability that makes recognition useful also creates privacy concerns. When a facial image is publicly available, an unauthorised system may attempt to identify the person without direct interaction or explicit consent.

Facial privacy protection therefore aims to modify the image in such a way that:

- the protected image remains visually acceptable to a human viewer;
- automated facial recognition becomes less reliable;
- the protection remains effective after common image transformations;
- the computational cost remains practical; and
- the protection can generalize across different recognition models.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

4 Literature Review

Research in facial privacy protection has investigated several approaches. These approaches can broadly be classified into adversarial perturbation, semantic manipulation, personalised protection, adversarial appearance modification, and diffusion-based methods.

4.1 Fawkes

Fawkes investigated image cloaking to protect individuals against unauthorised facial recognition [6]. The basic idea is to modify an image in a way that remains difficult for humans to notice while altering the features used by recognition models.

Fawkes demonstrated that adversarial image modifications could be used as a privacy tool rather than only as an attack technique. However, protection effectiveness can depend on the target recognition models and the conditions under which the protected image is processed.

4.2 LowKey

Cherepanova et al. proposed LowKey as a privacy-preserving approach for facial images [5]. The method investigates stronger protection against face recognition systems by applying carefully optimized perturbations.

Such approaches demonstrate the potential of adversarial machine learning for facial privacy. However, robustness against transformations and generalization across recognition models remain important considerations.

4.3 StyleAdv

Le et al. investigated StyleAdv, an approach based on adversarial modification of visual characteristics [4]. The research illustrates how semantic and appearance level modifications can interfere with recognition while maintaining a natural visual appearance.

This direction is relevant to the proposed framework because semantic modification can complement conventional pixel-level perturbation.

4.4 DiffAM

Sun et al. proposed DiffAM, which investigates diffusion-based adversarial makeup for facial privacy protection [3]. Diffusion models provide a powerful mechanism for modifying visual content while preserving realistic appearance.

However, diffusion based generation may require greater computational resources compared with simpler image-processing methods.

4.5 CloakDiff

Lu et al. investigated CloakDiff, which applies diffusion-based methods to the problem of facial privacy protection [2]. Diffusion-based privacy techniques demonstrate the possibility of combining strong semantic image modification with adversarial protection.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

The computational cost and practical deployment requirements of diffusion models remain important engineering challenges.

4.6 DiffUE

Meerza et al. investigated DiffUE, representing a recent direction in diffusion-based privacy protection research [1]. Such research indicates increasing interest in generative models for protecting facial identity while maintaining visually plausible images.

The literature suggests that future privacy systems may need to combine multiple protection levels instead of relying on a single perturbation mechanism.

5 Comparative Analysis

Table I summarizes selected research directions related to facial privacy protection.

Table 1: Comparison of Selected Facial Privacy Protection Studies

Year	Study	Approach	Main Idea	Strength	Limitation
2026	DiffUE	Diffusion-based	Uses diffusion-based image manipulation for privacy	Semantic modification	Computational cost
2026	CloakDiff	Diffusion-based	Applies diffusion techniques for facial identity protection	Natural image modification	Generation complexity
2024	DiffAM	Diffusion + adversarial makeup	Modifies facial appearance to reduce recognition	Realistic appearance modification	Model dependence
2024	StyleAdv	Style-based adversarial method	Uses appearance-level modifications against recognition	Semantic visual modification	Robustness concerns
2021	LowKey	Adversarial perturbation	Optimizes perturbations against recognition systems	Adversarial protection	Model dependence
2020	Fawkes	Image cloaking	Applies adversarial modifications	Privacy oriented approach	Recognition dependence

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

6 Proposed Adaptive Hybrid Framework

The proposed framework combines semantic perturbation, adaptive pixel-level noise, and adaptive image filtering.

The central idea is that no single protection technique should be assumed to be sufficient for every facial image and every recognition system. Instead, the protection intensity can be adjusted according to image characteristics and the desired privacy level.

The proposed processing pipeline is:

Input Facial Image → Face Detection and Preprocessing → Facial Feature Analysis → Semantic Perturbation → Adaptive Pixel-Level Noise → Adaptive Image Filtering → Protected Facial Image → Evaluation

6.1 Input Facial Image

The system accepts a facial image as its input. The image may originate from a personal photograph, social media image, profile photograph, or another source where facial privacy is required.

6.2 Face Detection and Preprocessing

The facial region is identified and separated from irrelevant image content. Preprocessing may include resizing, normalization, alignment, and removal of unnecessary background information.

6.3 Facial Feature Analysis

Important facial characteristics are analysed to determine areas that may strongly contribute to identity recognition.

The analysis can be used to determine the appropriate protection intensity instead of applying the same modification to every image.

6.4 Semantic Perturbation

Semantic perturbation modifies meaningful visual characteristics of the face. The objective is to change identity-related information while retaining a natural overall appearance.

The protected semantic representation can be represented as:

$$I_s = f(I,S) \quad (1)$$

where I is the original image, S represents semantic modification parameters, and f represents the semantic transformation function.

6.5 Adaptive Pixel-Level Noise

Pixel-level perturbation can then be applied to further interfere with recognition features.

A simple representation is:

$$I_p = I + N \quad (2)$$

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

where I represents the input image and N represents an adversarial noise component.

In the proposed framework, the noise is adaptive rather than fixed. Its magnitude and spatial distribution can depend on image characteristics.

After semantic modification, the protected image can be represented as:

$$I_p = I_s + N \quad (3)$$

where I_s represents the semantically modified image.

6.6 Adaptive Image Filtering

An additional filtering stage can be applied according to the characteristics of the protected image.

The filtering function can be represented as:

$$I_f = g(I_p, F) \quad (4)$$

where I_p is the perturbed image, F represents adaptive filtering parameters, and g represents the filtering function.

The purpose of this stage is to balance privacy protection and visual quality.

6.7 Protected Facial Image

The final output is an image designed to preserve acceptable visual appearance while reducing the effectiveness of unauthorized facial recognition.

The system should not assume that privacy is achieved simply because a single recognition model fails. Evaluation should consider multiple recognition architectures and common image transformations.

7 System Architecture

The conceptual architecture of the proposed framework consists of five major layers.

7.1 Input Layer

The input layer receives facial images from the user or an image-sharing application.

7.2 Analysis Layer

The analysis layer performs face detection, preprocessing, and facial feature analysis.

7.3 Protection Layer

The protection layer contains the three major mechanisms:

1) semantic perturbation;
2) adaptive pixel-level noise; and
3) adaptive image filtering.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

7.4 Output Layer

The output layer produces the protected facial image.

7.5 Evaluation Layer

The evaluation layer measures privacy protection, robustness, image quality, and computational requirements.

8 Evaluation Parameters

A future experimental implementation of the proposed framework should evaluate the following parameters.

8.1 Privacy Protection

Privacy protection can be evaluated by measuring the recognition performance before and after applying the protection framework.

A successful protection method should reduce unauthorised identification or verification performance while maintaining acceptable visual quality.

8.2 Robustness

The protected image should be tested after common transformations such as:

- JPEG compression;
- resizing;
- cropping;
- brightness changes;
- image format conversion;
- mild image filtering.

Robustness is important because an attacker may process the image before feeding it into a recognition system.

8.3 Image Quality

Privacy protection should not make the image unusable for normal human viewing.

Possible evaluation metrics include:

- Peak Signal-to-Noise Ratio (PSNR);
- Structural Similarity Index (SSIM);
- perceptual similarity; and
- human visual assessment.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

8.4 Computational Cost

The processing time, memory consumption, and hardware requirements should also be measured.

This parameter is especially important for diffusion-based techniques because generative image processing can be computationally expensive.

9 Applications

The proposed framework may be useful in several privacy-sensitive applications.

9.1 Social Media Privacy

Users can protect facial photographs before uploading them to social media platforms.

9.2 Personal Photograph Protection

Private photographs can be modified before being stored or shared online.

9.3 Privacy-Aware Image Sharing

The framework can support image sharing systems where users want to maintain control over biometric information.

9.4 Online Profile Protection

Profile photographs can be protected against unauthorized automated identification.

9.5 Privacy-Preserving Computer Vision

The framework can support applications where images must remain visually useful while reducing unnecessary exposure of identity information.

10 Limitations

The proposed framework is conceptual and several limitations must be considered.

First, privacy protection may depend on the recognition architecture against which the image is evaluated. A method that reduces recognition performance for one model may not provide the same protection against another.

Second, common image transformations can weaken adversarial modifications. Compression, resizing, cropping, and filtering may remove or alter useful perturbations.

Third, stronger protection may reduce visual quality. Therefore, privacy and image quality must be treated as competing objectives.

Fourth, diffusion-based protection may require significant computational resources and may not be suitable for low-power or real-time environments without optimization.

Finally, no privacy-protection technique should be interpreted as providing an absolute guarantee against every possible recognition system.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

11 Future Scope

Future research can extend the proposed framework in several directions.

1) Adaptive protection can be developed for different facial recognition architectures.
2) Transformation-aware protection can be investigated to improve robustness after compression, resizing, and cropping.
3) Lightweight diffusion models can be studied to reduce computational requirements.
4) Semantic perturbation and pixel-level adversarial noise can be jointly optimized.
5) Adaptive filtering can be improved using image-content analysis.
6) Real-time privacy protection can be investigated for mobile devices and social media applications.
7) Experiments can be conducted using multiple public facial image datasets.
8) Cross-model and black-box evaluation can be used to investigate generalisation.
9) Human visual-quality studies can be performed to determine whether protected images remain acceptable to users.

Conclusion

Facial recognition provides significant benefits in authentication, surveillance, security, and computer vision applications, but its widespread use also creates important privacy challenges.

Facial images contain biometric information that can potentially be collected and processed without the knowledge or consent of the individual.

Existing research has investigated several privacy protection strategies, including adversarial image cloaking, personalized perturbations, semantic modification, adversarial makeup, and diffusion-based image generation. These approaches demonstrate that facial images can be modified to interfere with automated recognition while attempting to maintain visual quality.

However, challenges remain in robustness, model dependence, computational cost, image quality, and generalization. A single protection mechanism may not be sufficient under changing recognition models and image transformations.

This paper therefore proposes a conceptual adaptive hybrid framework that combines semantic perturbation, adaptive pixel-level noise, and adaptive image filtering. The framework is intended to provide multiple levels of protection while maintaining acceptable image quality. Future experimental work should evaluate the framework using multiple recognition models, datasets, image transformations, privacy metrics, image-quality measures, and computational constraints.

The proposed approach represents a research direction toward privacy-aware computer vision systems in which individuals can have greater control over the use of their facial biometric information.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

References

[1] M. Meerza, et al., "DiffUE: Diffusion-Based Approaches for Facial Privacy Protection," 2026.

[2] L. Lu, et al., "CloakDiff: Diffusion-Based Facial Privacy Protection," 2026.

[3] Y. Sun, et al., "DiffAM: Diffusion-Based Adversarial Makeup for Facial Privacy Protection," 2024.

[4] N. Le, et al., "StyleAdv: Adversarial Style Manipulation for Facial Privacy Protection," 2024.

[5] A. Cherepanova, S. Abdalla, S. G. et al., "LowKey: Leveraging Adversarial Attacks to Protect Against Facial Recognition," 2021.

[6] J. Shan, J. Wu, and others, "Fawkes: Protecting Personal Privacy against Unauthorized Facial Recognition," 2020.