Privacy-Preserving Image Features via Adversarial Affine Subspace Embeddings
通過對抗仿射子空間嵌入保護隱私的圖像特徵

Mihai Dusmanu, Johannes L. Schönberger, Sudipta N. Sinha, Marc Pollefeys

1 Department of Computer Science, ETH Zürich
2 Microsoft

https://arxiv.org/abs/2006.06634
CVPR 2021 Best Paper Candidate

Abstract 摘要

Many computer vision systems require users to upload image features to the cloud for processing and storage.
許多計算機視覺系統需要用戶將圖像特徵上傳到雲端進行處理和存儲。

These features can be exploited to recover sensitive information about the scene or subjects, e.g., by reconstructing the appearance of the original image.
可以利用這些特徵來恢復有關場景或主題的敏感信息，例如，通過重建原始圖像的外觀。

To address this privacy concern, we propose a new privacy-preserving feature representation.
為了解決這個隱私問題，我們提出了一種新的隱私保護特徵表示。

The core idea of our work is to drop constraints from each feature descriptor by embedding it within an affine subspace containing the original feature as well as adversarial feature samples.
我們工作的核心思想是通過將每個特徵描述符嵌入到包含原始特徵和對抗性特徵樣本的仿射子空間中來刪除每個特徵描述符的約束。

Feature matching on the privacy preserving representation is enabled based on the notion of subspace-to-subspace distance.
基於子空間到子空間距離的概念啟用隱私保護表示上的特徵匹配。

We experimentally demonstrate the effectiveness of our method and its high practical relevance for the applications of visual localization and mapping as well as face authentication.
我們通過實驗證明了我們的方法的有效性及其與視覺定位和映射以及人臉認證應用的高度實際相關性。

Compared to the original features, our approach makes it significantly more difficult for an adversary to recover private information.

與原始特徵相比，我們的方法使對手恢復私人資訊得更加困難。

1. Introduction 前言

Image feature extraction and matching are two fundamental steps in many computer vision applications, such as 3D reconstruction [1, 2], image retrieval [1, 3], or face recognition [4].
圖像特徵提取和匹配是許多計算機視覺應用中的兩個基本步驟，例如 3D 重建 [1, 2]、圖像檢索 [1, 3] 或人臉識別 [4]。

Image features can be categorized into lowlevel [1, 5], mid-level [3, 6] or high-level [7, 8] depending on their information content and receptive field.
圖像特徵可以根據其資訊內容和感受分為低級 [1, 5]、中級 [3, 6] 或高級 [7, 8]。

Furthermore, features can be hand-crafted or learned using data-driven techniques.
此外，特徵可以手工製作或使用數據驅動技術學習。

However, they are almost always represented as vectors in high-dimensional feature spaces.
然而，它們幾乎總是在高維特徵空間中表示為向量。

Multiple feature vectors are then compared using appropriate distance metrics, which forms the basis of nearest neighbor search or other retrieval and recognition techniques.
然後使用適當的距離度量來比較多個特徵向量，這構成了最近鄰搜索或其他檢索和識別技術的基礎。

Recently, there has been rapid progress in feature inversion methods that reconstruct the image appearance from features extracted in the original image [9, 10, 11, 12] as shown in Figure 1.
最近，特徵反演方法取得了快速進展，該方法從原始圖像 [9, 10, 11, 12] 中提取的特徵重建圖像外觀，如圖 1 所示。

This raises serious privacy concerns, since images may contain sensitive information about the scene or subjects.
這引起了嚴重的隱私問題，因為圖像可能包含有關場景或主題的敏感信息。

Increased awareness of these privacy issues has spurred significant efforts to develop privacy-preserving machine learning systems.
對這些隱私問題的認識不斷提高，促使人們大力開發保護隱私的機器學習系統。

In recent years, researchers have proposed a large body of approaches to tackle the various aspects of the problem, including homomorphic cryptosystems [13], differential privacy [14], federated learning [15], and specific solutions for camera localization [16, 17].
近年來，研究人員提出了大量方法來解決該問題的各個方面，包括同態密碼系統 [13]、差分隱私 [14]、聯邦學習 [15] 以及相機定位的具體解決方案 [16、17]。]。

In this paper, we propose a new feature representation suitable for visual recognition and matching tasks that makes it significantly more difficult for an adversary to reconstruct the image contents.
在本研究之中，我們提出了一種適用於視覺識別和匹配任務的新特徵表示，使對手重建圖像內容變得更加困難。

Our approach has only marginal computational overhead, which makes it amenable to a wide range of practical scenarios.
我們的方法只有微不足道的計算開銷，這使其適用於廣泛的實際場景。

The core idea behind our method is to represent a descriptor point in Rn as an affine subspace of Rn passing through the original point. We refer to this process as lifting.
我們的方法背後的核心思想是將 Rn 中的描述符點表示為 Rn 穿過原始點的仿射子空間。 我們將此過程稱為提升。

The chosen dimension of the subspace determines a trade-off between accuracy, runtime, and the level of privacy of the feature representation.
選擇的子空間維度決定了特徵表示的準確性、運行時間和隱私級別之間的權衡。

To make inverting the representation difficult, we propose a strategy for constructing a lifted subspace containing additional adversarial feature points.
為了使表示反轉變得困難，我們提出了一種構建包含額外對抗性特徵點的提升子空間的策略。

We empirically demonstrate strong privacy preservation even for low-dimensional affine subspaces.
即使對於低維仿射子空間，我們也憑經驗證明了強大的隱私保護。

Pairwise feature comparison is a fundamental step required in many recognition tasks.
成對特徵比較是許多識別任務所需的基本步驟。

In our proposed framework, such comparisons are done directly on the lifted subspaces based on either point-to-subspace or subspace-to-subspace distance.
在我們提出的框架中，這種比較是直接在基於點到子空間或子空間到子空間距離的提升子空間上進行的。

The paper is organized as follows.
本文的結構如下。

First, we formally present the idea of lifting and the technique for matching lifted features.
首先，我們正式提出提升的思想和匹配提升特徵的技術。

Next, we analyze the performance of these features for two applications:

接下來，我們分析這些功能對於兩個應用程序的性能：

a) image matching for visual localization and mapping as well as
a) 用於視覺定位和映射的圖像匹配以及

b) face authentication.
b) 人臉認證。

Finally, we demonstrate that our proposed representation is resilient to potential privacy attacks.
最後，我們證明了我們提出的表示對潛在的隱私攻擊具有彈性。

The code of our method and the evaluation protocol will be released as open-source.
我們的方法和評估協議的代碼將作為開源發布。

Figure 1: Privacy-Preserving Image Features. Inversion of traditional local image features is a privacy concern in many applications.
圖 1：隱私保護圖像功能。 傳統局部圖像特徵的反轉是許多應用程序中的隱私問題。

Our proposed approach obfuscates the appearance of the original image by lifting the descriptors to affine subspaces.
我們提出的方法通過將描述符提升到仿射子空間來混淆原始圖像的外觀。

Distance between the privacy-preserving subspaces enables efficient matching of features.
隱私保護子空間之間的距離可以實現特徵的高效匹配。

The same concept can be applied to other domains such as face features for biometric authentication.
相同的概念可以應用於其他領域，例如用於生物特徵認證的面部特徵。

Image credit: laylamoran4battersea (Layla Moran).
圖片來源：laylamoran4battersea (Layla Moran)。

2. Related Work 相關工作

We first review image features used for applications considered in this paper.
我們首先回顧用於本文考慮的應用程序的圖像特徵。

We then discuss existing work about privacy attacks on image features and defense mechanisms.
然後，我們討論有關對圖像特徵和防禦機制的隱私攻擊的現有工作。

Feature Descriptors. 特徵描述符。

In the traditional local feature extraction paradigm, after keypoint detection and shape estimation, normalized image patches are extracted from images.
在傳統的局部特徵提取範式中，在關鍵點檢測和形狀估計之後，從圖像中提取歸一化的圖像塊。

Feature description takes a patch as input and outputs an n-dimensional vector.
特徵描述將一個補丁作為輸入並輸出一個 n 維向量。

Handcrafted local descriptors are based on direct pixel sampling [18] or a histogram of image gradients [1, 5].
手工製作的局部描述符基於直接像素採樣 [18] 或圖像梯度直方圖 [1, 5]。

Recent advances in deep learning have led to descriptors based on convolutional neural networks (CNNs).
深度學習的最新進展導致了基於卷積神經網絡 (CNN) 的描述符。

Such learnable descriptors are trained using triplet [19] or list-wise [20] losses and hard-negative mining techniques [21].
這種可學習的描述符使用三元組 [19] 或列表式 [20] 損失和硬負挖掘技術 [21] 進行訓練。

Local features have been successfully used for tasks such as large-scale 3D reconstruction from crowd-sourced images [2] and image retrieval [22, 23].
這種可學習的描述符使用三元組 [19] 或列表式 [20] 損失和硬負挖掘技術 [21] 進行訓練。

Face recognition methods start by face detection and alignment to obtain a canonical face image [24].
人臉識別方法從人臉檢測和對齊開始，以獲得規範的人臉圖像 [24]。

Subsequently, a well chosen low-dimensional subspace of pixelspace can provide good recognition performance [4].
隨後，一個精心選擇的像素空間的低維子空間可以提供良好的識別性能[4]。

More recently, CNN-based features have become the de facto choice for face descriptors.
最近，基於 CNN 的特徵已成為面部描述符的事實上的選擇。

These networks are trained using different classification losses [25, 26, 27].
這些網絡使用不同的分類損失進行訓練 [25, 26, 27]。

Feature Subspaces. 特徵子空間。

Wang et al. [28] also use a subspace representation for feature matching.
Wang et al. [28] 還使用子空間表示進行特徵匹配。

Different to their method, we consider affine instead of linear subspaces.
與他們的方法不同，我們考慮仿射而不是線性子空間。

Accordingly, our distance function is not based on principal angles but on the closest pair of points between the two subspaces.
因此，我們的距離函數不是基於主角度，而是基於兩個子空間之間最近的一對點。

Further, contrary to grouping descriptors of similar patches together to improve matching performance, we add adversarial descriptors to the subspaces to improve privacy.
此外，與將相似補丁的描述符分組在一起以提高匹配性能相反，我們向子空間添加對抗性描述符以提高隱私。

Feature Inversion and Compromising Privacy.
特徵反轉和隱私洩露。

Weinzaepfel et al. [9] proposed a method for reconstructing images from local image features using a database of patches with associated descriptors.
Weinzaepfel et al. [9] 提出了一種使用具有相關描述符的補丁數據庫從局部圖像特徵重建圖像的方法。

Dosovitsky and Brox [10, 29] extended on this work by using a CNN and perceptual losses, while Pittaluga et al. [30] showed that it was possible to recover detailed images from sparse 3D point clouds reconstructed using structure-from-motion.
Dosovitsky and Brox [10, 29] 通過使用 CNN 和感知損失擴展了這項工作，而 Pittaluga et al. [30] 表明可以從使用結構運動重建的稀疏 3D 點雲中恢復詳細圖像。

Similarly, Zhmoginov and Sandler [11] and Mai et al. [12] proposed methods for reconstructing face images from their descriptors.
同樣，Zhmoginov and Sandler [11] 以及 Mai et al. [12] 提出了從描述符重建人臉圖像的方法。

Moreover, they showed that the reconstructed images could even be used by an attacker to fool an authentication system.
此外，他們還表明，攻擊者甚至可以使用重建的圖像來欺騙身份驗證系統。

Privacy-Preserving Methods.
隱私保護方法。

Differential privacy [14] expands upon Dalenius [31] by formalizing the problem of querying a database without inadvertently releasing information distinguishing the individual entries in the database.
差異隱私 [14] 擴展了 Dalenius [31]，將查詢數據庫的問題形式化，而不會無意中發佈區分數據庫中各個條目的信息。

An extended overview can be found in [32].
可以在 [32] 中找到擴展概述。

Instead of protecting information leakage from a database, our scenario is quite different in that we are interested in protecting private information in the query as well as contributing new information to a database in a privacy-preserving manner.
我們的場景不是保護數據庫中的信息洩漏，而是完全不同的，因為我們有興趣保護查詢中的私人信息以及以保護隱私的方式向數據庫提供新信息。

McMahan et al. [33] introduced federated learning, a distributed client-server framework for training a model, where training data remains with the clients, thus offering better privacy guarantees.
McMahan et al. [33] 引入了聯邦學習，一種用於訓練模型的分佈式客戶端 - 服務器框架，其中訓練數據保留在客戶端，從而提供更好的隱私保證。

Kairouz et al. [15] reviews the topic and discusses open problems.
Kairouz et al. [15] 回顧了該主題並討論了未解決的問題。

In contrast, we address a different setting, where tasks require image features computed by clients to be shared with the server.
相比之下，我們解決了一個不同的設置，其中任務需要客戶端計算的圖像特徵與服務器共享。

In this context, our approach makes it difficult to recover private image information from the shared features.
在這種情況下，我們的方法很難從共享特徵中恢復私有圖像信息。

Existing works on local features process images encrypted using different homomorphic cryptosystems [34, 35] in the cloud.
現有的關於本地特徵的工作處理使用雲中不同同態密碼系統 [34, 35] 加密的圖像。

Jiang et al. [36] proposed an alternative by additively splitting the image into two ciphertext matrices using a private prime modulus.
Jiang et al. [36] 提出了一種替代方案，即使用私有素數模數將圖像附加地分成兩個密文矩陣。

These methods guarantee that the original images remain private, but they do not prevent information leakage by inverting the obtained local features.
這些方法保證原始圖像保持私密性，但它們並不能通過反轉獲得的局部特徵來防止信息洩漏。

One could also use `2 distance computation on encrypted feature vectors [37, 38, 39].
還可以對加密的特徵向量使用`2 距離計算 [37, 38, 39]。

However, recent works regarding homomorphic representation search [38, 39] remain computationally expensive, while our method only comes with marginal overhead.
然而，最近關於同態表示搜索 [38, 39] 的工作在計算上仍然很昂貴，而我們的方法只帶來了邊際開銷。

Furthermore, these cryptosystems provide security through encryption, where a breach of the secret keys is a privacy risk.
此外，這些密碼系統通過加密提供安全性，其中洩露密鑰會帶來隱私風險。

In contrast, our system does not have the same single point of failure and provides parameters to trade off accuracy, runtime, and privacy.
相比之下，我們的系統沒有相同的單點故障，並提供參數來權衡準確性、運行時間和隱私。

Speciale et al. [16, 17] proposed solutions tailored to image-based localization, where geometric information is concealed by lifting 2D or 3D points to randomly oriented lines passing through the original locations.
Speciale 等。 [16, 17] 提出了針對基於圖像的定位定制的解決方案，其中通過將 2D 或 3D 點提升到通過原始位置的隨機定向線來隱藏幾何信息。

Recent work extends on the same idea to solve the full structure-frommotion problem [40, 41].
最近的工作擴展了相同的想法，以解決完整的結構-從運動問題 [40, 41]。

We draw inspiration from their approach, but instead lift feature descriptors to higher dimensional affine subspaces to conceal appearance information.
我們從他們的方法中汲取靈感，而是將特徵描述符提升到更高維度的仿射子空間以隱藏外觀信息。

3. Method 方法

In this paper, we will represent features from a particular domain as vectors in Rn, where n is the dimensionality of the original feature space.
在本文中，我們將來自特定域的特徵表示為 Rn 中的向量，其中 n 是原始特徵空間的維數。

We denote span the linear span of a set of vectors vi ∈ R^n.
我們將 span(v1, ... , vm) 表示為一組向量 vi ∈ R^n 的線性。

span(v1, ... , vm) = $\{\sum_{i=1}^{m} \lambda_i v_i | \lambda_i \in \mathbb{R}\}$

An m-dimensional affine subspace A will be represented as the vector sum of a translation vector a0 and a linear subspace span(a1, …, am), giving A = a0 + span(a1, …, am).
m 維仿射子空間 A 將表示為平移向量 a0 和線性子空間 span(a1, ..., am) 的向量和，給出 A = a0 + span(a1, ..., am)。

The core idea of our method is to lift the original feature vector or descriptor d ∈ Rn to an m-dimensional affine subspace D ∈ Rn satisfying d ∈ D.
我们方法的核心思想是将原始特征向量或描述符 d ∈ Rn 提升到满足 d ∈ D 的 m 维仿射子空间 D ∈

Rn。

We denote the lifted affine subspace representation as private features.
我們將提升的仿射子空間表示表示為私有特徵。

There are two major requirements that we must address.
我們必須滿足兩個主要要求。

Firstly, we need to define a distance function that can be used to reliably and efficiently compare two features in this new representation.
首先，我們需要定義一個距離函數，該函數可用於可靠有效地比較這種新表示中的兩個特徵。

Secondly, we must construct the affine subspace in a way that effectively conceals the original feature vector d and makes it difficult for an attacker to carry out a successful privacy attack aiming to recover the vector d given the private representation D.
其次，我們必須以有效隱藏原始特徵向量 d 的方式構造仿射子空間，並使攻擊者難以進行成功的隱私攻擊，以在給定私有表示 D 的情況下恢復向量 d。

3.1. Distance Functions 距離函數

Most applications require feature descriptor comparison, which is accomplished using appropriate pairwise distance measures.
大多數應用程序需要特徵描述符比較，這是使用適當的成對距離度量來完成的。

In our analysis, we restrict ourselves to the Euclidean distance (denoted kk) as it is most commonly used in practice.
在我們的分析中，我們將自己限制在歐幾里得距離（表示為 kk），因為它在實踐中最常用。

To compute the distance between private features, we either use the point-to-subspace or subspace-to-subspace distance.
為了計算私有特徵之間的距離，我們使用點到子空間或子空間到子空間的距離。

Note that both distances are upper bound by the original point-to-point distance.
請注意，這兩個距離都是原始點到點距離的上限。

Point-to-Subspace Distance. - 點到子空間的距離

To compute the distance between a private descriptor d represented as an affine subspace D and a regular descriptor e, one can use the point-tosubspace distance defined as:
要計算表示為仿射子空間 D 的私有描述符 d 和常規描述符 e 之間的距離，可以使用定義為的點到子空間距離：

$$\mathrm{dist}(\mathcal{D}, e) = \min_{x \in \mathcal{D}} \|e - x\| = \|e - p_{\perp}^{\mathcal{D}}(e)\| \ , \qquad (1)$$

where pD⊥(e) denotes the orthogonal projection of e onto D.
其中 pD⊥(e) 表示 e 在 D 上的正交投影。

Subspace-to-Subspace Distance - 子空間到子空間的距離

To compute the distance between two private descriptors d, e represented as affine subspaces D, E of dimensions mD,mE , one can use the subspace-to-subspace distance defined as:
要計算兩個私有描述符之間的距離 d, e 表示為維度 mD,mE 的仿射子空間 D, E ，可以使用定義為的子空間到子空間距離：

$$\mathrm{dist}(\mathcal{D}, \mathcal{E}) = \min_{x \in \mathcal{D}, y \in \mathcal{E}} \|y - x\| \ . \qquad (2)$$

Let us denote a closest pair of points in the two subspaces as x*∈ D and y*∈ E, respectively.
讓我們將兩個子空間中最近的一對點分別表示為 x*∈D 和 y*∈E。

$$x^* = d_0 + \sum_{i=1}^{m_{\mathcal{D}}} \alpha_i d_i \ , y^* = e_0 + \sum_{i=1}^{m_{\mathcal{E}}} \beta_i e_i \ , \qquad (3)$$

Then, we have:
然後，我們有：

where α∈αRmD, β∈ RmE .
其中 α ∈αRmD, β∈RmE 。

In the following derivation, we assume that both subspaces have the same dimension m = mD = mE for simplicity.
在下面的推導中，為簡單起見，我們假設兩個子空間具有相同的維度 m = mD = mE。

A sufficient and necessary condition for dist(D, E) = ||y*- x*|| is that the line y* - x* is orthogonal to both D and E:
dist(D, E) = ||y*- x*|| 的充分必要條件 是直線 y* - x* 與 D 和 E 正交：

$$\begin{cases} (y^* - x^*)^T d_i = 0 \\ (y^* - x^*)^T e_i = 0 \end{cases} \ , \qquad (4)$$

which can be rewritten as:
可以改寫為：

$$\begin{cases} (e_0 - d_0)^T d_i = \sum_{j=1}^{m} \alpha_j d_i^T d_j + \sum_{j=1}^{m} \beta_j (-d_i^T e_j) \\ (e_0 - d_0)^T e_i = \sum_{j=1}^{m} \alpha_j e_i^T d_j + \sum_{j=1}^{m} \beta_j (-e_i^T e_j) \end{cases} . \tag{5}$$

This system can be formulated in a more compact form:
該系統可以用更緊湊的形式製定：

$$\begin{bmatrix} DD^T & -DE^T \\ ED^T & -EE^T \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} D \\ E \end{bmatrix} (e_0 - d_0) \ , \tag{6}$$

where

$$D = \left[d_1 \dots d_m\right]^T, E = \left[e_1 \dots e_m\right]^T \in M_{m \times n}(\mathbb{R}).$$

If the bases of the subspaces are orthonormal (DDT = EET = I), the system further simplifies to:
如果子空間的基是正交的（DDT = EET = I），系統進一步簡化為：

$$\begin{bmatrix} I & -DE^T \\ ED^T & -I \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} D \\ E \end{bmatrix} (e_0 - d_0) \ . \tag{7}$$

Thus, finding the subspace-to-subspace distance requires solving a linear system with 2m unknowns and equations.
因此，找到子空間到子空間的距離需要求解具有 2m 未知數和方程的線性系統。

$$\text{Let } \bar{M} = -DE^T.$$

Under the assumption that the matrix N = I - MMT is invertible, the block-matrix inversion formula can be used to rewrite Eq. 7 as follows:
在矩陣 N = I - MMT 可逆的假設下，塊矩陣求逆公式可用於重寫方程。 Eq. 7 如下：

$$\begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} N^{-1} & N^{-1}M \\ -M^T N^{-1} & -M^T N^{-1} M - I \end{bmatrix} \begin{bmatrix} D \\ E \end{bmatrix} (e_0 - d_0) \ . \tag{8}$$

The solutions to α, β can be substituted into Eq. 3 to obtain the subspace-to-subspace distance as ||x*- y*||.
α,β 的解可以代入方程。 3 获得子空间到子空间的距离为 ||x*- y*||。

Note that the problem can also be formulated using the dual representation of a subspace as the intersection of n-m hyperplanes.
请注意，该问题也可以使用子空间的对偶表示作为 n-m 个超平面的交集来表述。

We provide a derivation of the dual formulation and further discussion in the supplementary material.
我們在補充材料中提供了雙重公式的推導和進一步討論。

3.2. Affine Subspace Embedding 仿射子空間嵌入

Each subspace embedding is defined by a translation vector d0 and a basis {d1, … , dm}.
每個子空間嵌入由平移向量 d0 和基礎 {d1, ⋯ , dm} 定義。

The choice of these and the distribution of the original descriptors has significant impact on the effectiveness of our approach and the required dimensionality m of the subspace to achieve sufficient privacy preservation.
這些的選擇和原始描述符的分佈對我們方法的有效性和實現充分隱私保護所需的子空間維數 m 有重大影響。

For example, it is common practice to l2-normalize descriptors [1, 21, 26, 27].
例如，通常的做法是對描述符 [1, 21, 26, 27] 進行 l2 歸一化。

In such cases, lifting descriptors to affine lines (m = 1) is not secure.
在這種情況下，將描述符提升到仿射線 (m = 1) 是不安全的。

This is because a line intersects the unit hyper-sphere in at most 2 points.
這是因為一條線最多與單位超球面相交 2 個點。

It can be easy to detect which of the two intersections is statistically plausible and thereby exactly recover the original point.
可以很容易地檢測出兩個交點中的哪一個在統計上是合理的，從而準確地恢復原始點。

However, any value of m > 1 generally produces infinite intersection points and thus provides much more ambiguity which is desirable for privacy preservation.
然而，任何 m > 1 的值通常會產生無限的交點，從而提供更多的模糊性，這對於隱私保護是可取的。

We now describe different lifting strategies, which we later compare in our experimental evaluation.
我們現在描述不同的提升策略，我們稍後會在我們的實驗評估中進行比較。

Random Basis - 隨機基礎

One could sample random direction vectors for the linear subspace, i.e., di~ U([-1, 1])n referred to as random lifting.
可以為線性子空間採樣隨機方向向量，即 di~U([-1, 1])n，稱為隨機提升。

In our experiments, we found this approach to be vulnerable to relatively simple privacy attacks.
在我們的實驗中，我們發現這種方法容易受到相對簡單的隱私攻擊。

One could sample random direction vectors for the linear subspace, i.e., di ~U([-1; 1])^n referred to as random lifting.
可以為線性子空間採樣隨機方向向量，即 di ~U([-1; 1])^n 稱為隨機提升。

In our experiments, we found this approach to be vulnerable to relatively simple privacy attacks.
在我們的實驗中，我們發現這種方法容易受到相對簡單的隱私攻擊。

The original descriptor can be approximated by the nearest entry from a database of real-world descriptors according to the point-to-subspace distance.
根據點到子空間的距離，可以通過來自真實世界描述符數據庫的最近條目來近似原始描述符。

This is because random subspaces generally intersect the descriptor manifold once.
這是因為隨機子空間通常與描述符流形相交一次。

Adversarial Basis. 對抗性基礎。

To address this issue, one can ensure that the subspace passes through multiple regions of the descriptor manifold.
為了解決這個問題，可以確保子空間通過描述符流形的多個區域。

We propose to use a database of realworld descriptors W = {w1, … ,ws} as an approximation of the manifold and sample the basis vectors as di = wi -d, where wi~U(W).
我們建議使用真實世界描述符的數據庫 W = {w1, ⋯ ,ws} 作為流形的近似值，並將基向量採樣為 di = wi - d，其中 wi~U(W)。

We call this approach adversarial lifting, as it intentionally introduces plausible samples in the subspace to conceal the original descriptor.
我們稱這種方法為對抗性提升，因為它有意在子空間中引入似是而非的樣本來隱藏原始描述符。

Moreover, a defender can choose adversarial samples to hide specific private information, e.g., to hide the gender of a person, one can pick a feature vector from another gender, as shown in our experimental

evaluation.
此外，防御者可以選擇對抗樣本來隱藏特定的私人信息，例如，為了隱藏一個人的性別，可以從另一個性別中選擇一個特徵向量，如我們的實驗評估所示。

Adversarial sampling improves privacy but reduces descriptor matching accuracy, because the chance of accidental subspace intersections increases.
對抗性採樣提高了隱私，但降低了描述符匹配的準確性，因為意外子空間交叉的機會增加了。

To balance the accuracy and privacy trade-offs, we propose combining the adversarial and random lifting strategies, which we call hybrid lifting.
為了平衡準確性和隱私權衡，我們建議結合對抗性和隨機提升策略，我們稱之為混合提升。

In hybrid lifting, a subset of the basis vectors are selected randomly while the rest are chosen using adversarial sampling.
在混合提升中，隨機選擇基向量的一個子集，而使用對抗性採樣選擇其餘的基向量。

There are different ways to implement the adversarial and hybrid strategies depending on the task at hand.
根據手頭的任務，有不同的方法來實施對抗性和混合策略。

We describe a few such variants in the context of local features and face descriptors in Section 4.
我們在第 4 節的局部特徵和麵部描述符的上下文中描述了一些這樣的變體。

Translation Vector. 翻譯向量。

The origin can be set to any point in the subspace, except for the vector d itself, since it is precisely what we must conceal.
原點可以設置為子空間中的任何點，除了向量 d 本身，因為它正是我們必須隱藏的。

Thus, we sample a random point and project it to the subspace, as follows:
因此，我們採一個隨機點並將其投影到子空間，如下所示：

$$d_0 = p_{\perp}^{d+\mathrm{span}(d_1,\ldots,d_m)}(e) \text{ where } e \sim \mathcal{U}([-1,1])^n \quad . \quad (9)$$

Information Leakage. 信息洩露

It is important to carefully construct the subspace to avoid accidental leakage of information.
仔細構造子空間以避免信息的意外洩漏很重要。

For instance, in the adversarial formulation described above, all basis vectors (di for i > 0) point “away from" the initial descriptor d.

例如，在上述對抗性公式中，所有基向量 (di for i > 0) 都指向“遠離”初始描述符 d。

An attacker could target parts of the descriptor manifold where these directions are feasible.
攻擊者可以針對這些方向可行的描述符流形部分。

More precisely, one could look for real-world descriptors d~ such that d~+di also intersects the descriptor manifold.
更準確地說，人們可以尋找現實世界的描述符 d~，使得 d~+di 也與描述符流形相交。

To mitigate this, given an initial subspace D, we generate a random basis as:
為了緩解這種情況，給定一個初始子空間 D，我們生成一個隨機基為：

4. Experimental Evaluation 實驗評估

In this section, we evaluate our method on two applications.
在本節中，我們將在兩個應用程序上評估我們的方法。

First, we experiment with local features on the task of image matching for visual localization and mapping.
首先，我們在視覺定位和映射的圖像匹配任務中嘗試使用局部特徵。

Second, we apply our method to global image features for face verification.
其次，我們將我們的方法應用於全局圖像特徵以進行面部驗證。

We report results in these two settings and assess the trade-offs between the degree of privacy preservation achieved, the accuracy of the target task and the computational complexity.
我們報告了這兩種設置的結果，並評估了實現的隱私保護程度、目標任務的準確性和計算複雜性之間的權衡。

As we cannot provide any theoretical guarantees on privacy preservation, we implement plausible privacy attacks and empirically demonstrate that our approach is robust against them.
由於我們無法為隱私保護提供任何理論保證，因此我們實施了看似合理的隱私攻擊，並憑經驗證明我們的方法對它們具有魯棒性。

Dist.	Time (ms)	Subspace dimension 2	4	8
s-to-s	GPU	2.02 ± 0.14	6.02 ± 0.14	N/A
	CPU	107.87 ± 0.95	195.50 ± 2.02	540.98 ± 25.18
p-to-s	GPU	2.02 ± 0.14	2.10 ± 0.30	4.17 ± 0.38
	CPU	25.25 ± 1.14	37.71 ± 0.55	63.24 ± 1.08

Table 1: **Runtime.** We report the average runtime over 100 runs of the distance matrix computation for an image pair, when varying the lifting dimension. Each image has 1000 128-dimensional floating point features. We consider both the subspace-to-subspace (s-to-s) and the point-to-subspace (p-to-s) distance. For the former, we implemented specialized CUDA solvers for lifting dimensions 2 and 4. Hardware: NVIDIA RTX 2080Ti, Intel Core i9-9900K.

Table 1: Runtime. We report the average runtime over 100 runs of the distance matrix computation for an image pair, when varying the lifting dimension.
表 1：運行時。 當改變提升維度時，我們報告了圖像對距離矩陣計算超過 100 次的平均運行時間。

Each image has 1000 128-dimensional floating point features.
每個圖像有 1000 個 128 維浮點特徵。

We consider both the subspace-to-subspace (s-to-s) and the point-to-subspace (p-to-s) distance.
我們同時考慮子空間到子空間（s-to-s）和點到子空間（p-to-s）的距離。

For the former, we implemented specialized CUDA solvers for lifting dimensions 2 and 4.
對於前者，我們實施了專門的 CUDA 求解器來提升維度 2 和 4。

Hardware: NVIDIA RTX 2080Ti, Intel Core i9-9900K.

4.1. Runtime 運行

Previous approaches to privacy-preserving descriptors take advantage of homomorphic encryption.
以前的隱私保護描述符方法利用了同態加密。

While these methods guarantee an exact distance computation, they are severely limited in terms of practical applicability, especially in real-time scenarios.
雖然這些方法保證了精確的距離計算，但它們在實際適用性方面受到嚴重限制，尤其是在實時場景中。

A recent work about encrypted representation search [39] reports that computing the distances between a single 128-dimensional query vector and a database with 1000 entries takes around 1 second (c.f . Figure 3 [39]).
最近關於加密表示搜索 [39] 的一項工作報告稱，計算單個 128 維查詢向量和具有 1000 個條目的數據庫之間的距離大約需要 1 秒（參見圖 3 [39]）。

Thus, obtaining the full distance matrix for an image pair with 1000 features each would take around 16 minutes.
因此，獲得每個具有 1000 個特徵的圖像對的全距離矩陣大約需要 16 分鐘。

In comparison, our method only induces minimal computational overhead, as shown in Table 1.
相比之下，我們的方法只引起最小的計算開銷，如表 1 所示。

For completeness, the runtime for computing the point-to-point distance matrix in the same setting is 1.01 +-0.10ms on GPU and 1.05+-0.46ms on CPU, respectively.
為完整起見，在相同設置下計算點對點距離矩陣的運行時間在 GPU 上分別為 1.01 +-0.10ms 和在 CPU 上為 1.05+-0.46ms。

4.2. Local Feature Descriptors 本地特徵描述符

In order to demonstrate the robustness and generalizability of our approach, we perform experiments using the arguably most popular hand-crafted local feature (SIFT [1]) as well as a recent state-of-the-art learned descriptor (Hard-Net [21]).
為了證明我們的方法的魯棒性和通用性，我們使用可以說是最受歡迎的手工製作的局部特徵（SIFT [1]）以及最近最先進的學習描述符（Hard-Net [21]）。

Both descriptors are by default l2-normalized.
默認情況下，兩個描述符都是“l2-歸一化的”。

We evaluate the private descriptors on the tasks of image matching, structure-from-motion and visual localization.
我們在圖像匹配、運動結構和視覺定位任務上評估私有描述符。

Subspace Selection. 子空間選擇。

The adversarial lifting database is obtained by clustering 10 million local features from 60,000 images of the Places365 dataset [43] into s = 256,000 clusters using spherical k-means [44].
對抗提升數據庫是通過使用球形 k 均值 [44] 將 Places365 數據集 [43] 的 60,000 張圖像中的 1000 萬個局部特徵聚類到 s = 256,000 個集群中獲得的。

In the context of 3D computer vision tasks, it is usually desirable to have many thousands of features per image [45].
在 3D 計算機視覺任務的上下文中，通常希望每個圖像具有數千個特徵 [45]。

Let us consider the case of lifting to descriptor planes (m = 2) using uniform random sampling from the database of 256,000 centroids.
讓我們考慮使用來自 256,000 個質心的數據庫的均勻隨機採樣提升到描述符平面 (m = 2) 的情況。

Given an image pair with 8,000 descriptors per image, for each feature in the second image, there is a 1=16 chance of sampling a feature already selected in the first one.
給定一個圖像對，每個圖像有 8,000 個描述符，對於第二張圖像中的每個特徵，有 1=16 的機會對第一個圖像中已經選擇的特徵進行採樣。

Such a collision causes subspace intersections and thus leads to wrong feature matches.
這種碰撞會導致子空間交叉，從而導致錯誤的特徵匹配。

This is further exacerbated by typical match filtering strategies (e.g., mutual check, ratio-test [1]).
典型的匹配過濾策略（例如，相互檢查、比率測試 [1]）進一步加劇了這種情況。

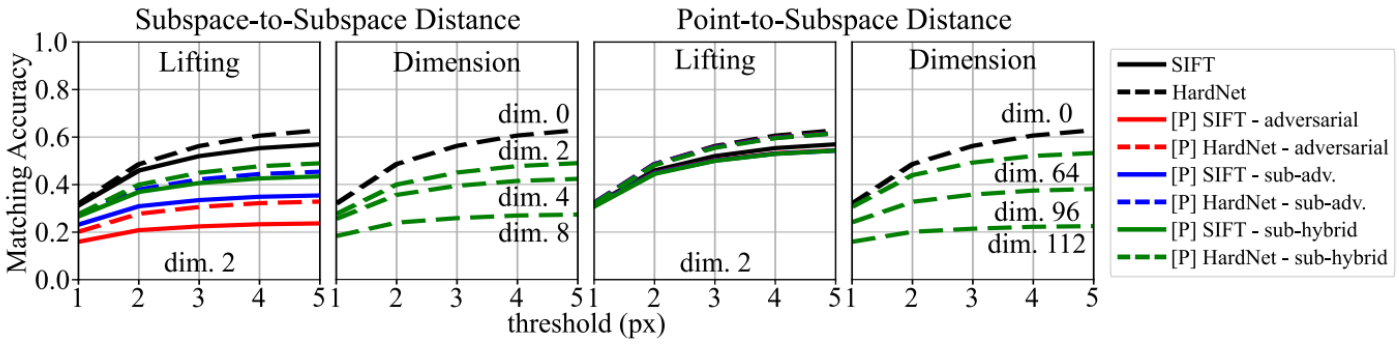

Figure 2: **Matching evaluation.** We plot the mean matching accuracy at different thresholds on the HPatches sequences [42]. Methods using our private representation are prefixed by [P]. We report results with different lifting methods and dimensions. HardNet outperforms SIFT on this benchmark and the ordering is respected after lifting as well.

Figure 2: Matching evaluation. We plot the mean matching accuracy at different thresholds on the HPatches sequences [42].
圖 2：匹配評估。 我們在 HPatches 序列 [42] 上繪製了不同閾值下的平均匹配精度。

Methods using our private representation are prefixed by [P].

使用我們私有表示的方法以 [P] 為前綴。

We report results with different lifting methods and dimensions.
我們報告具有不同提升方法和尺寸的結果。

HardNet outperforms SIFT on this benchmark and the ordering is respected after lifting as well.
HardNet 在此基準測試中的表現優於 SIFT，並且在提升後也遵循排序。

To reduce the number of wrong matches, we randomly split the database into S pairwise disjoint sub-databases W1, ⋯,WS satisfying W = ∪S i=1Wi, card(Wi) = s/S.
為了減少錯誤匹配的數量，我們將數據庫隨機拆分為 S 個成對不相交的子數據庫 W1, ⋯,WS，滿足 W = ∪S i=1Wi, card(Wi) = s/S。

For an image I, we then first randomly select a sub-database W ~ U({W1, …,WS}).
對於圖像 I，我們首先隨機選擇一個子數據庫 W ~ U({W1, ⋯,WS})。

Next, the basis vectors are generated using only elements of W, i.e., vi = wi - d, where wi ~ U(W).
接下來，僅使用 W 的元素生成基向量，即 vi = wi - d，其中 wi ~ U(W)。

If two images select different sub-databases in this sub-adversarial lifting strategy, the probability of random collision is 0.
如果在這個子對抗提升策略中兩幅圖像選擇不同的子數據庫，則隨機碰撞的概率為 0。

For images using the same sub-database, the number of collisions is very high.
對於使用相同子庫的圖像，碰撞次數非常高。

Overall, with this strategy, instead of degrading the matching performance for all image pairs, we achieve good matching performance in 15/16 cases for S = 16.
總體而言，使用這種策略，我們不會降低所有圖像對的匹配性能，而是在 S = 16 的 15/16 種情況下實現了良好的匹配性能。

In addition, we also evaluate a sub-hybrid lifting strategy, where half of the basis vectors are random and the other half uses a sub-database.
此外，我們還評估了一種亞混合提升策略，其中一半的基向量是隨機的，另一半使用子數據庫。

Image Matching. 圖像匹配。

We compare raw descriptors with their private counterparts on the image sequences from the HPatches dataset [42].
我們在 HPatches 數據集 [42] 的圖像序列上將原始描述符與其私有對應物進行比較。

This dataset consists of 116 scenes with 6 images each: 57 of them exhibit illumination changes, while the other 59 show significant viewpoint changes.
該數據集由 116 個場景組成，每個場景有 6 張圖像：其中 57 個顯示光照變化，而其他 59 個顯示出顯著的視點變化。

For each scene, we match the first image against the other 5 yielding 580 image pairs in total.
對於每個場景，我們將第一個圖像與其他 5 個圖像進行匹配，總共產生 580 個圖像對。

For evaluation, we follow protocol introduced by Dusmanu et al. [46] which reports the mean matching accuracy of a mutual nearest neighbors matcher for different values of the threshold up to which a match is considered correct.
對於評估，我們遵循 Dusmanu et al. [46] 介紹的協議。 它報告了相互最近鄰匹配器對於閾值的不同值的平均匹配精度，直到匹配被認為是正確的。

Figure 2 shows results for both distances with different lifting methods and dimensions.
圖 2 顯示了使用不同提升方法和尺寸的兩種距離的結果。

Random lifting is not plotted as it performs identical with the raw descriptors.
隨機提升未繪製，因為它與原始描述符執行相同。

As mentioned above, adversarial lifting performs poorly for local features due to subspace collisions.
如上所述，由於子空間碰撞，對抗性提升對局部特徵的表現不佳。

This is, in part, addressed by the use of sub-databases and further improved by sub-hybrid lifting.
這部分是通過使用子數據庫解決的，並通過子混合提升進一步改進。

The point-to-subspace distance only preserves the privacy of one image and is useful for cloud- and client-based visual localization systems, equivalent to Speciale et al. [16, 17].
點到子空間的距離只保留一張圖像的隱私，對於基於雲和客戶端的視覺定位系統很有用，相當於 Speciale et al. [16, 17]。

This approach is able to achieve good matching performance even for very high lifting dimensions.
即使對於非常高的提升尺寸，這種方法也能夠實現良好的匹配性能。

Structure-from-Motion. 來自運動的結構。

Next, we integrate the best performing private representation from above (sub-hybrid lifting) into an end-to-end 3D reconstruction pipeline [47] and evaluate it on the crowd-sourced 3D reconstruction benchmark of Schönberger et al. [45].
接下來，我們將上面性能最好的私有表示（亞混合提升）集成到端到端 3D 重建管道 [47] 中，並在 Schönberger 等人的眾包 3D 重建基准上對其進行評估。 [45]。

For each image, we retrieve the top 50 most similar images using NetVLAD [48] and only match against these.
對於每張圖像，我們使用 NetVLAD [48] 檢索前 50 個最相似的圖像，並僅與這些圖像匹配。

Next, we run geometric verification (with a minimum inlier ratio of 0:1) followed by sparse reconstruction using COLMAP [47, 49] and finally report the reconstruction statistics in Table 2.
接下來，我們運行幾何驗證（最小內點比為 0:1），然後使用 COLMAP [47, 49] 進行稀疏重建，最後在表 2 中報告重建統計數據。

For this evaluation, we preserve the privacy of all input images.
對於此評估，我們保留所有輸入圖像的隱私。

As already observed in our image matching evaluation, the private features come with accuracy trade-offs.
正如在我們的圖像匹配評估中已經觀察到的那樣，私有特徵伴隨著準確性的權衡。

As we increase the dimensionality of the subspace, the reconstruction completeness degrades accordingly.
隨著我們增加子空間的維數，重建完整性相應地降低。

Despite the fewer number of registered images, the 3D models remain relatively accurate and clearly distinguishable.
儘管註冊圖像數量較少，但 3D 模型仍然相對準確且清晰可辨。

The generally lower track length for private features is caused by missing matches leading to longer feature tracks being split into multiple smaller ones.
私有要素的軌跡長度通常較短是由於缺少匹配導致較長的要素軌跡被拆分為多個較小的軌跡。

Visual Localization. 視覺定位

We also consider the case of localizing to an already built map on the challenging Aachen Day-Night long-term visual localization dataset [50].
我們還考慮了在具有挑戰性的亞琛晝夜長期視覺定位數據集 [50] 上定位到已經構建的地圖的情況。

This is equivalent to the scenario tackled by Speciale et al. [17], where the goal is to protect the privacy of users of an imagebased localization service, such as Google Visual Positioning System [51] or Microsoft Azure Spatial Anchors [52].
這相當於 Speciale et al. [17]解決的場景，其目標是保護基於圖像的本地化服務用戶的隱私，例如 Google Visual Positioning System [51] 或 Microsoft Azure Spatial Anchors [52]。

We first triangulate the database model from the given camera poses and intrinsics using DoG keypoints with raw SIFT and HardNet descriptors, respectively.
我們首先分別使用帶有原始 SIFT 和 HardNet 描述符的 DoG 關鍵點從給定的相機姿勢和內在函數對數

據庫模型進行三角剖分。

For each query image (824 day-time and 98 night-time), we retrieve the top 50 database images using NetVLAD [48].
對於每個查詢圖像（824 個白天和 98 個夜間），我們使用 NetVLAD [48] 檢索前 50 個數據庫圖像。

We preserve the privacy of all query images with sub-hybrid lifting and use point-to-subspace distance for matching.
我們通過亞混合提升保護所有查詢圖像的隱私，並使用點到子空間距離進行匹配。

Finally, we use the COLMAP [47] image registrator with fixed intrinsics to obtain poses that are submitted to the long-term visual localization benchmark [53].
最後，我們使用具有固定內在函數的 COLMAP [47] 圖像註冊器來獲取提交給長期視覺定位基準 [53] 的姿勢。

Dataset	Method	Reg. images	Sparse points	Track length	Reproj. error
Madrid Metropolis 1344 images	SIFT	400	28,862	7.01	0.72
	[P] SIFT - dim. 2	302	17,232	6.37	0.59
	[P] SIFT - dim. 4	227	11,461	5.54	0.56
	HardNet	459	42,180	7.25	0.89
	[P] HardNet - dim. 2	367	28,367	6.49	0.68
	[P] HardNet - dim. 4	268	15,562	6.32	0.58
Gendarmen-markt 1463 images	SIFT	896	74,348	6.37	0.84
	[P] SIFT - dim. 2	783	64,554	5.44	0.71
	[P] SIFT - dim. 4	458	33,291	5.23	0.60
	HardNet	999	112,245	6.68	0.96
	[P] HardNet - dim. 2	864	89,865	5.98	0.80
	[P] HardNet - dim. 4	751	63,862	5.50	0.69
Tower of London 1576 images	SIFT	635	64,490	7.78	0.70
	[P] SIFT - dim. 2	525	55,439	6.58	0.61
	[P] SIFT - dim. 4	439	37,819	6.10	0.56
	HardNet	749	89,818	7.85	0.81
	[P] HardNet - dim. 2	557	69,161	7.19	0.68
	[P] HardNet - dim. 4	498	49,570	6.69	0.61

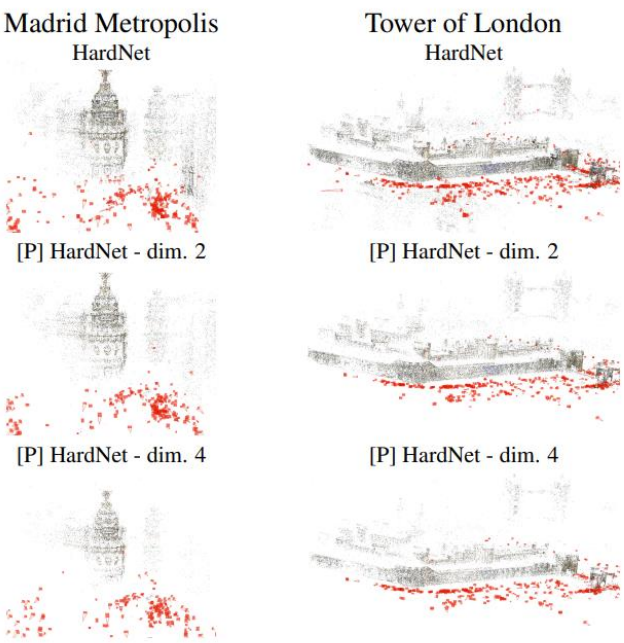

Table 2: **Local Feature Evaluation Benchmark.** We report reconstruction statistics such as the number of registered images and sparse points and the average track length and reprojection error on internet photo collections of landmarks [45]. Methods prefixed by [P] use sub-hybrid lifting for all features of input images. On the right side, we visualize the final sparse models.

Table 2: Local Feature Evaluation Benchmark.
表 2：局部特徵評估基準。

We report reconstruction statistics such as the number of registered images and sparse points and the average track length and reprojection error on internet photo collections of landmarks [45].
我們報告了重建統計數據，例如註冊圖像和稀疏點的數量以及地標的互聯網照片集的平均軌跡長度和重投影誤差 [45]。

Methods prefixed by [P] use sub-hybrid lifting for all features of input images. On the right side, we visualize the final sparse models.
以 [P] 為前綴的方法對輸入圖像的所有特徵使用亞混合提升。 在右側，我們將最終的稀疏模型可視化。

Query	Method	Thresholds 0.25m, 2°	0.5m, 5°	5.0m, 10°
Day (824)	SIFT	82.9%	89.6%	92.2%
	[P] SIFT - dim. 2	79.5%	87.0%	91.1%
	[P] SIFT - dim. 4	79.6%	86.5%	91.1%
	[P] SIFT - dim. 16	76.7%	84.0%	87.4%
	HardNet	86.3%	92.5%	95.6%
	[P] HardNet - dim. 2	84.3%	89.8%	94.3%
	[P] HardNet - dim. 4	83.5%	90.2%	93.6%
	[P] HardNet - dim. 16	82.0%	88.3%	92.2%
Night (98)	SIFT	41.8%	48.0%	55.1%
	[P] SIFT - dim. 2	32.7%	36.7%	42.9%
	[P] SIFT - dim. 4	32.7%	38.8%	43.9%
	[P] SIFT - dim. 16	25.5%	31.6%	34.7%
	HardNet	60.2%	67.3%	73.5%
	[P] HardNet - dim. 2	49.0%	53.1%	58.2%
	[P] HardNet - dim. 4	40.8%	44.9%	49.0%
	[P] HardNet - dim. 16	32.7%	37.8%	43.9%

Table 3: **Aachen Day-Night Localization Challenge.** We report the percentage of localized query images for both day and night scenarios under different camera pose accuracy threshold on the Aachen Day-Night dataset [50]. For the private methods (prefixed by [P]), we use sub-hybrid lifting for query images and point-to-subspace distance for matching.

Table 3: Aachen Day-Night Localization Challenge.
表 3：亞琛晝夜定位挑戰。

We report the percentage of localized query images for both day and night scenarios under different camera pose accuracy threshold on the Aachen Day-Night dataset [50].
我們報告了在亞琛晝夜數據集 [50] 上不同相機姿勢精度閾值下白天和夜間場景的本地化查詢圖像的百分比。

For the private methods (prefixed by [P]), we use sub-hybrid lifting for query images and point-to-subspace distance for matching.
對於私有方法（以 [P] 為前綴），我們使用亞混合提升查詢圖像和點到子空間距離進行匹配。

Following the standard evaluation protocol, we report the percentage of localized query images for different real-world thresholds in Table 3.
遵循標準評估協議，我們在表 3 中報告了不同現實世界閾值的本地化查詢圖像的百分比。

On the day queries, we are able to achieve competitive performance even when lifting to 16 dimensional subspaces.
當天查詢，即使提升到 16 維子空間，我們也能夠實現有競爭力的性能。

As previously, the accuracy gradually decreases when increasing the lifting dimension.
如前所述，隨著提升維度的增加，精度逐漸降低。

Furthermore, even on the extremely hard night-to-day matching queries where pose estimation has very low inlier ratios, we are still able to localize a reasonable number of queries.
此外，即使在姿勢估計具有非常低的內點比率的極其困難的夜間匹配查詢中，我們仍然能夠定位合理數量的查詢。

Privacy Attack. 隱私攻擊。

To analyze attacks on the proposed private descriptor representation, we provide the adversary with multiple tools.
為了分析對提議的私有描述符表示的攻擊，我們為對手提供了多種工具。

We assume that they have access to a database V of 128,000 real-world descriptors built using the same procedure as the lifting database (described above).
我們假設他們可以訪問使用與提升數據庫（如上所述）相同的過程構建的 128,000 個真實世界描述符的數據庫 V。

Further, the attacker has unrestricted access to the lifting algorithm and is able to use it on-demand.
此外，攻擊者可以不受限制地訪問提升算法，並且能夠按需使用它。

Finally, they have access to extensive training data (the MegaDepth [54] dataset) as well as the architecture and loss from Pittaluga et al. [30] allowing them to train new feature inversion networks.
最後，他們可以訪問大量的訓練數據（MegaDepth [54] 數據集）以及 Pittaluga et al. [30]的架構和損失，允許他們訓練新的特徵反演網絡。

First, we consider a nearest neighbor attack (NNA) where each subspace is approximated by its closest correspondence from a database of real-world descriptors.
首先，我們考慮最近鄰攻擊 (NNA; nearest neighbor attack)，其中每個子空間都通過其與現實世界描述符數據庫中的最接近的對應關係來近似。

Formally, for each private representation D associated to a descriptor d, the database V is used to retrieve the closest element to the subspace d~ = arg minv∈V dist(D, v).
形式上，對於與描述符 d 相關聯的每個私有表示 D，數據庫 V 用於檢索最接近子空間 d~ = arg minv∈V dist(D, v) 的元素。

Next, the approximated descriptors d~ can be fed to a regular feature inversion network to reconstruct the appearance of the original image.
接下來，可以將近似描述符 d~ 饋送到常規特徵反轉網絡以重建原始圖像的外觀。

Second, we consider a direct inversion attack (DIA) where the affine subspaces are provided as input to a CNN.
其次，我們考慮直接反轉攻擊 (DIA)，其中仿射子空間作為 CNN 的輸入提供。

To this end, we train multiple feature inversion networks from Difference-of-Gaussians (DoG) keypoints and private descriptors lifted to 2, 4, and 6 dimensions, respectively.
為此，我們分別從高斯差分 (DoG) 關鍵點和私有描述符提升到 2、4 和 6 維來訓練多個特徵反演網絡。

Note that the architectures proposed in previous works [30, 10] are very compute and memory intensive – training them on higher dimensional subspaces would be a challenge in itself.
請注意，以前的工作 [30, 10] 中提出的架構是非常計算和內存密集型的—在更高維的子空間上訓練它們本身就是一個挑戰。

Attack	Lifting	Dim.	MAE (↓)	SSIM (↑)	PSNR (↑)
	raw	0	0.105	0.755	17.937
NNA	random	2	0.112	0.738	17.448
NNA	sub-hybrid	2	0.206	0.530	12.288
DIA	sub-hybrid	2	0.176	0.648	13.447
DIA	sub-hybrid	4	0.179	0.594	13.531
DIA	sub-hybrid	6	0.194	0.559	12.823

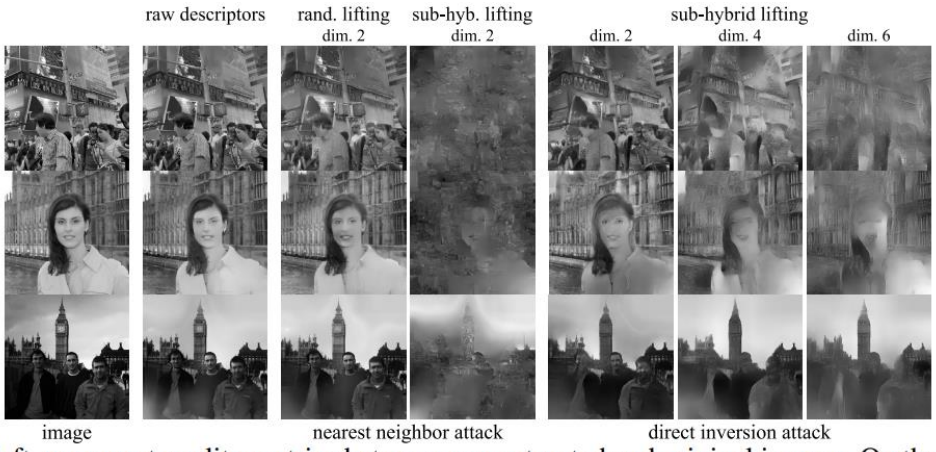

Figure 3: **Image reconstruction.** On the left, we report quality metrics between reconstructed and original images. On the right, we show several qualitative examples: first original image, then reconstructions from the raw descriptors and using the proposed privacy attacks on different lifting methods and dimensions. Image credit (top to bottom): *pagedooley* (Kevin Dooley), *laylamoran4battersea* (Layla Moran), *martinalvarez* (Martin Alvarez Espinar).

Figure 3: Image reconstruction.
圖 3：圖像重建。

On the left, we report quality metrics between reconstructed and original images.
在左側，我們報告了重建圖像和原始圖像之間的品質指標。

On the right, we show several qualitative examples: first original image, then reconstructions from the raw descriptors and using the proposed privacy attacks on different lifting methods and dimensions.
在右側，我們展示了幾個定性示例：首先是原始圖像，然後是從原始描述符重建，並在不同的提升方法和維度上使用提議的隱私攻擊。

Image credit (top to bottom): pagedooley (Kevin Dooley), laylamoran4battersea (Layla Moran), martinalvarez (Martin Alvarez Espinar).
圖片來源（從上到下）：pagedooley (Kevin Dooley)、laylamoran4battersea (Layla Moran)、martinalvarez (Martin Alvarez Espinar)。

We run the proposed privacy attacks on 10 images1 using HardNet descriptors and present the results in Figure 3.
我們使用 HardNet 描述符對 10 個圖像進行了提議的隱私攻擊，並在圖 3 中顯示了結果。

On the left, we quantitatively report image reconstruction quality metrics such as mean absolute error (MAE), structural similarity index measure (SSIM), and peak signal-to-noise ratio(PSNR); on the right, we show qualitative image reconstructions.
在左側，我們定量報告了圖像重建質量指標，例如平均絕對誤差 (MAE)、結構相似性指數度量 (SSIM) 和峰值信噪比 (PSNR)； 在右側，我們展示了定性圖像重建。

Please refer to the supplementary material for more examples.
更多示例請參考補充材料。

Using the raw descriptors, one can reconstruct the original image with very high fidelity (note the readability of text in the first example).
使用原始描述符，可以以非常高的保真度重建原始圖像（注意第一個示例中文本的可讀性）。

The nearest neighbor attack is successful on private features using random lifting, but not when using sub-hybrid lifting due to the adversarial samples.
最近鄰攻擊在使用隨機提升的私有特徵上是成功的，但由於對抗樣本，在使用亞混合提升時則不然。

For all reconstructions, the general outline of the buildings is recovered mainly due to the lack of features in the sky (e.g., third example).
對於所有重建，建築物的總體輪廓主要是由於天空中缺乏特徵而恢復的（例如，第三個示例）。

The direct inversion attack is able to reconstruct some parts of the original image, but the quality is significantly deteriorated.
直接反轉攻擊能夠重建原始圖像的某些部分，但質量明顯下降。

Furthermore, distinguishing details such as faces or text are heavily perturbed and become non-existent for higher lifting dimensions.
直接反轉攻擊能夠重建原始圖像的某些部分，但質量明顯下降。

4.3. Face Descriptors 人臉描述符

For this evaluation, we use a state-of-the-art deep face descriptor – the best performing ArcFace [27] model

with a ResNet-101 [55] backbone trained on MS-Celeb-1M [56].
對於這次評估，我們使用了最先進的深度人臉描述符—性能最好的 ArcFace [27] 模型，帶有在 MS-Celeb-1M [56] 上訓練的 ResNet-101 [55] 主幹。

Face Verification. 人臉驗證。

We report face verification accuracy on multiple datasets: LFW [57], CFP [58] (both frontal-frontal denoted FF and frontal-profile denoted FP), and AgeDB-30 [59].
我們報告了多個數據集上的人臉驗證準確性：LFW [57]、CFP [58]（正面-正面表示 FF 和正面輪廓表示 FP）和 AgeDB-30 [59]。

We follow the regular evaluation protocol, notably 10-fold cross validation where, for each fold, the training split is used to determine a distance threshold that separates between same / different identity and the accuracy is computed on the validation split.
我們遵循常規評估協議，特別是 10 倍交叉驗證，其中，對於每個折疊，訓練拆分用於確定區分相同/不同身份的距離閾值，並且在驗證拆分上計算準確性。

Finally, the mean classification accuracy over the 10 folds is reported in Figure 4.
最後，圖 4 報告了 10 倍的平均分類準確率。

We evaluate two scenarios: point-to-subspace (p-to-s) matching, where one of the images is represented using the original descriptor and the other one is lifted to a subspace, and subspace-to-subspace (s-to-s) matching, where both descriptors are private.
我們評估了兩種場景：點到子空間（p-to-s）匹配，其中一個圖像使用原始描述符表示，另一個被提升到子空間，以及子空間到子空間（s-to -s) 匹配，其中兩個描述符都是私有的。

As expected, the point-to-subspace matching performs better across the board.
正如預期的那樣，點到子空間匹配的性能更好。

For the subspaceto-subspace distance, the performance on the simple datasets (LFW and CFP-FF) only drops by a few percents.
對於子空間到子空間的距離，簡單數據集（LFW 和 CFP-FF）上的性能僅下降了幾個百分點。

For more complex datasets (frontal-profile matching in CFP-FP, large age differences in AgeDB-30), the performance drop is more significant.
對於更複雜的數據集（CFP-FP 中的正面輪廓匹配，AgeDB-30 中的大年齡差異），性能下降更為顯著。

Nevertheless, the simpler datasets are still very representative of common authentication systems (Microsoft Windows Hello [60], Apple Face ID [61]), making our approach highly relevant for such scenarios.
儘管如此，更簡單的數據集仍然非常能代表常見的身份驗證系統（Microsoft Windows Hello [60]、Apple Face ID [61]），使我們的方法與此類場景高度相關。

Privacy Attack. 隱私攻擊。

The privacy attack we are concerned with involves inferring distinguishing properties (gender, race) from only the ArcFace [27] descriptors.
我們關注的隱私攻擊涉及僅從 ArcFace [27] 描述符推斷可區分的屬性（性別、種族）。

For this purpose, we used FairFace [62], a face dataset consisting of 97,698 images with balanced gender (2 classes) and race (7 classes) annotations.
為此，我們使用了 FairFace [62]，這是一個由 97,698 張圖像組成的人臉數據集，具有平衡的性別（2 類）和種族（7 類）註釋。

We randomly selected 10,000 training images for the database needed by our lifting method.
我們為我們的提升方法所需的數據庫隨機選擇了 10,000 張訓練圖像。

The remaining 76,744 training images were used for the attack.
剩餘的 76,744 張訓練圖像用於攻擊。

The validation set of 10,954 images is used for evaluation.
10,954 張圖像的驗證集用於評估。

We attack an ArcFace descriptor using a K-nearest neighbors(K-NN) classifier [44] to predict the gender and race of the person.
我們使用 K-最近鄰（K-NN）分類器 [44] 攻擊 ArcFace 描述符來預測人的性別和種族。

We do this both on the original feature as well as the lifted feature for K = 10.
我們在原始特徵和 K = 10 的提升特徵上都這樣做。

We also implemented a variant of our hybrid lifting method (denoted hybrid+) that exploits the gender / race of each person.
我們還實施了一種混合提升方法的變體（表示為混合+），該方法利用了每個人的性別/種族。

In this variant, each feature is lifted by sampling database entries with a different gender / of a different race to obtain a balanced subspace, which better conceals these attributes.
在這個變體中，每個特徵都通過對不同性別/不同種族的數據庫條目進行採樣來提升，以獲得平衡的子空間，這更好地隱藏了這些屬性。

The results are reported in Figure 5.
結果如圖 5 所示。

The black vertical lines denote the approximate performance of a random classifier.

黑色垂直線表示隨機分類器的近似性能。

Similar to image matching, pure random lifting is again not effective at concealing the private attributes.
與圖像匹配類似，純隨機提昇在隱藏私有屬性方面同樣無效。

Adversarial lifting has the best results in terms of privacy, but its face verification accuracy is also the worst.
對抗性提昇在隱私方面的效果最好，但其人臉驗證精度也是最差的。

Hybrid lifting offers a trade-off between random lifting (high performance) and adversarial lifting (good for privacy).
混合提升提供了隨機提升（高性能）和對抗提升（有利於隱私）之間的權衡。

Finally, the hybrid+version is most effective at concealing the gender.
最後，混合+版本在隱藏性別方面最有效。

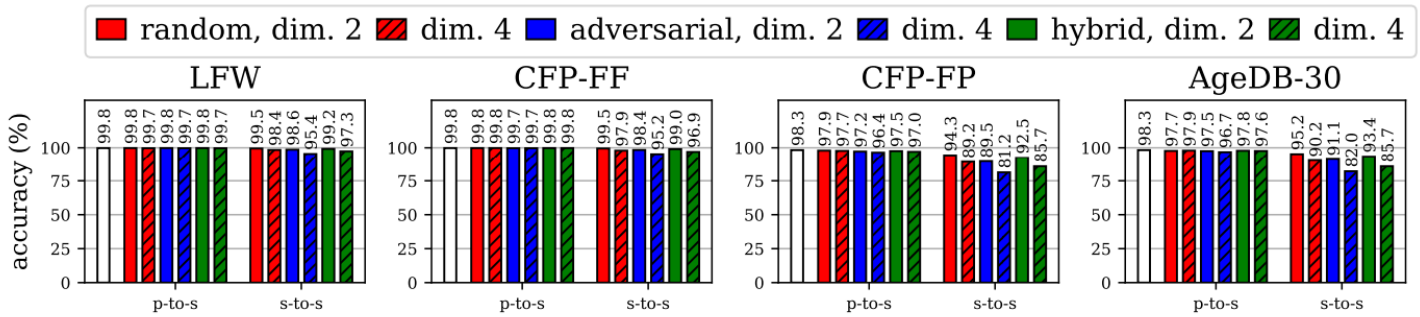

Figure 4: **Face verification.** We show the accuracy on different face verification datasets. The white bar represents the reference accuracy of raw ArcFace descriptors. The point-to-subspace distance (*p-to-s*), performs within at most 2% of the original descriptors. For the subspace-to-subspace distance (*s-to-s*), the performance drop is more significant in the difficult scenarios (CFP-FP and AgeDB-30), but frontal authentication (LFW and CFP-FF) is still very accurate (95% at worst).

Figure 4: Face verification. We show the accuracy on different face verification datasets.
圖 4：人臉驗證。 我們展示了不同人臉驗證數據集的準確性。

The white bar represents the reference accuracy of raw ArcFace descriptors.
白條代表原始 ArcFace 描述符的參考精度。

The point-to-subspace distance (p-to-s), performs within at most 2% of the original descriptors.
點到子空間的距離 (p-to-s) 最多在原始描述符的 2% 內執行。

For the subspace-to-subspace distance (s-to-s), the performance drop is more significant in the difficult scenarios (CFP-FP and AgeDB-30), but frontal authentication (LFW and CFP-FF) is still very accurate (95% at worst).
對於子空間到子空間的距離（s-to-s），在困難場景（CFP-FP 和 AgeDB-30）下性能下降更顯著，但正面認證（LFW 和 CFP-FF）仍然非常準確 （最壞情況下為 95%）。

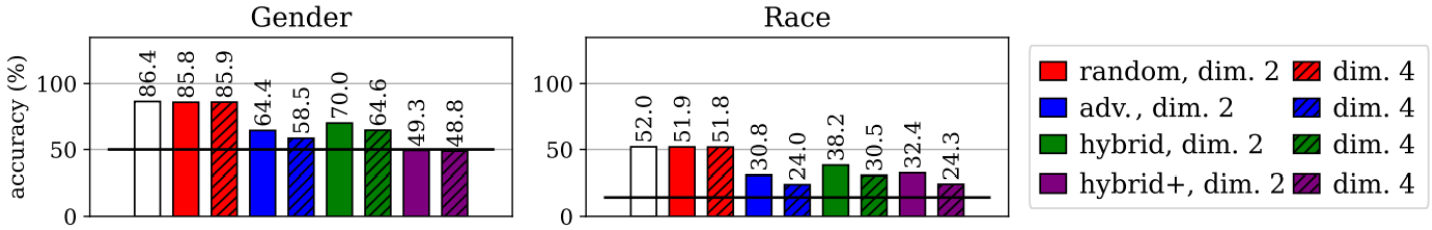

Figure 5: **FairFace.** We report the accuracy of a K-NN classifier trained to predict the gender and race of a subject from their ArcFace descriptor. The black line represents the approximate accuracy of a random classifier. The white bars represent the accuracy on raw ArcFace descriptors. Private representations using a database for lifting successfully conceal information.

Figure 5: FairFace. We report the accuracy of a K-NN classifier trained to predict the gender and race of a subject from their ArcFace descriptor.

圖 5：FairFace。 我們報告了經過訓練的 K-NN 分類器的準確性，該分類器用於從他們的 ArcFace 描述符預測對象的性別和種族。

The black line represents the approximate accuracy of a random classifier.

黑線代表隨機分類器的近似精度。

The white bars represent the accuracy on raw ArcFace descriptors.

白條代表原始 ArcFace 描述符的準確性。

Private representations using a database for lifting successfully conceal information.

使用數據庫的私人表示成功地隱藏了信息。

5. Limitations and Future Work 局限性和未來的工作

Speciale et al. [17] showed that solving the target task of camera localization reveals the concealed location of some features in the query image.

Speciale et al. [17] 表明解決相機定位的目標任務揭示了查詢圖像中某些特徵的隱藏位置。

Similarly, in our 3D reconstruction task, the pair of closest points on two matched affine subspaces provides a way to estimate the concealed feature descriptors.

類似地，在我們的 3D 重建任務中，兩個匹配的仿射子空間上的最近點對提供了一種估計隱藏特徵描述符的方法。

This implies that features associated with 3D points triangulated from multiple views are likely to be revealed.

這意味著可能會顯示與從多個視圖進行三角測量的 3D 點相關的特徵。

By inverting the estimated descriptors, an adversary might be able to approximately reconstruct the appearance of the stationary part of the scene.

通過反轉估計的描述符，對手可能能夠近似地重建場景靜止部分的外觀。

However, this is not a serious limitation, as feature descriptors extracted from image regions depicting people or other transient objects will generally not be matched in multiple overlapping images and therefore their

appearance is unlikely to be revealed.
然而，這並不是一個嚴重的限制，因為從描繪人或其他瞬態物體的圖像區域提取的特徵描述符通常不會在多個重疊圖像中匹配，因此它們的外觀不太可能被揭示。

For face verification, it is possible to infer the face descriptor after repeated authentications of a person if a history of the private descriptors is stored.
對於面部驗證，如果存儲了私人描述符的歷史記錄，則可以在一個人的重複身份驗證後推斷出面部描述符。

One potential mitigation is to generate near parallel subspaces for a particular individual, although it is unclear how this approach behaves with respect to the manifold of face descriptors.
一種潛在的緩解措施是為特定個體生成近乎平行的子空間，儘管尚不清楚這種方法在面部描述符的流形方面如何表現。

A potential option would be adding a trusted third-party in the system that receives private descriptors from both client and server and computes the distances without storing any data.
一個潛在的選擇是在系統中添加一個受信任的第三方，從客戶端和服務器接收私有描述符併計算距離而不存儲任何數據。

Apart from addressing these limitations, other directions for future work include training descriptors more suitable for lifting and implementing scalable matching inspired by prior work on subspace representations [63] to enable large-scale applications such as place recognition.
除了解決這些限制之外，未來工作的其他方向包括訓練更適合提升和實現可擴展匹配的描述符，這些描述受先前關於子空間表示的工作的啟發 [63]，以實現大規模應用，例如地點識別。

6. Conclusion 討論

We have proposed a novel privacy-preserving feature representation by embedding feature descriptors into affine subspaces containing adversarial samples.
我們通過將特徵描述符嵌入到包含對抗樣本的仿射子空間中，提出了一種新穎的隱私保護特徵表示。

To find similar features, nearest neighbor computation is enabled through point-to-subspace or subspace-to-subspace distance.
為了找到相似的特徵，最近鄰計算是通過點到子空間或子空間到子空間的距離啟用的。

We experimentally demonstrate the high practical relevance of our approach for crowd-sourced visual localization and mapping as well as face authentication, while rendering it difficult to recover sensitive information.
我們通過實驗證明了我們的方法在眾包視覺定位和映射以及面部認證方面的高度實際相關性，同時使其難以恢復敏感信息。

Acknowledgements. 致謝。

This work was supported by the Microsoft Mixed Reality & AI Zürich Lab PhD scholarship.
這項工作得到了微軟混合現實與人工智能蘇黎世實驗室博士獎學金的支持。