electronics

MDPI

Article

Connecting Visual Data to Privacy: Predicting and Measuring Privacy Risks in Images

Hongpu Jiang [1,2], Jinxin Zuo [1,2,*] and Yueming Lu [1,2]

[1] School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing 100876, China; lbjhp@bupt.edu.cn (H.J.); ymlu@bupt.edu.cn (Y.L.)
[2] National Engineering Research Center of Disaster Backup and Recovery, Beijing University of Posts and Telecommunications, Beijing 100876, China
* Correspondence: zuojx@bupt.edu.cn

Abstract: More and more users openly share their information on online websites, with the resulting privacy issues being under scrutiny. Content such as a user's personal data and location information is often asked for before posting to enforce the user's privacy preferences; however, little attention has been paid to the lack of content (e.g., images) posted by the user. Even if privacy preferences are requested before images are published, publishers often remain unaware of the extent of privacy leakage associated with their data. To this end, we provide an image privacy metric scheme that incorporates users' privacy preferences, with the core idea of assisting users in making data publishing decisions. First, we propose privacy-specific spatial attention mechanisms that can effectively improve the prediction accuracy. Next, we integrate set pair analysis (SPA) theory and use the network output as the privacy value. Finally, we combine a user study to understand the privacy preferences of different users with respect to these attributes and combine it with principal component analysis to correct and enforce user privacy preferences. Our model is trained with the ability to predict privacy risk end-to-end, thus being able to guide the user in sharing data in open platforms. We use the image privacy dataset, VISPR, to predict privacy information better than other methods.

Keywords: privacy attribute prediction; set pair analysis; measuring privacy

Academic Editor: Dimitra I. Kaklamani

Received: 14 January 2025
Revised: 15 February 2025
Accepted: 16 February 2025
Published: 19 February 2025

Citation: Jiang, H.; Zuo, J.; Lu, Y. Connecting Visual Data to Privacy: Predicting and Measuring Privacy Risks in Images. *Electronics* **2025**, *14*, 811. https://doi.org/10.3390/electronics14040811

1. Introduction

In the fast-evolving field of data science, the safeguarding of personal privacy faces unprecedented challenges. Consider the case of a young professional who, while sharing a celebratory photo of a work promotion on Facebook, unknowingly reveals a sensitive document in the background with financial figures, or a family sharing a vacation video on YouTube, where their home address can be inferred from the street signs in the footage. These real-life scenarios illustrate the fact that user-generated content (UGC) on social media, including text, images, and videos, contains a vast amount of personal privacy information [1], as illustrated in Figure 1. Studies have shown that up to 66% of the content on these platforms includes personal privacy data [2]. The widespread collection and sharing of personal information, including sensitive biometric, financial, and geographical location data, poses a significant threat to personal privacy and security. This is largely due to advancements in mobile internet, cloud computing, and big data technologies. Although the immediacy of communication and the wide dissemination of social media have developed qualitatively, these platforms, while offering convenience, also come with privacy dilemmas for users. Research on privacy preferences and social networks has been

explored in past studies. Research by [3] has examined the types of personal information disclosed on social networking sites. Other studies focus on preserving privacy while using social networks [4], as well as exploring privacy settings [5].

However, traditional methods mainly rely on user profiles and privacy settings, which have significant limitations. They are relatively weak in managing the unstructured features of UGC, especially in accurately detecting and measuring privacy in images and video content. For example, they often struggle to analyze complex visual scenes and identify subtle privacy-related elements. Their effectiveness in handling visual content, such as accurately classifying multi-label privacy images, remains to be fully verified. Compared with traditional concepts of privacy, the scope of privacy in the internet environment is expanding, and both users and social media platforms are increasingly entangled in privacy issues. Ensuring that personal information is accurately perceived and measured is critically important. Although traditional mechanisms have focused on user profiles and privacy settings, they are relatively weak in managing the unstructured features of UGC, particularly images and video content. Their effectiveness in analyzing visual content remains to be fully verified. Research has shown that individuals tend to misinterpret privacy-sensitive details in images, further complicating the safeguarding of personal privacy [6].

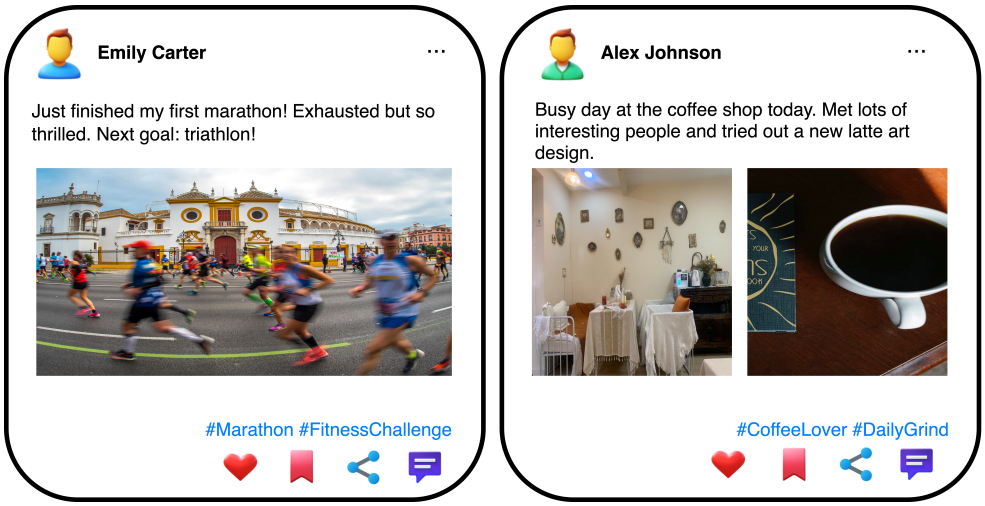

Figure 1. Examples of user-generated content.

Motivating scenarios where the proposed concept can play a role:

- Travel sharing: When users share travel photos on social media platforms, like Instagram or Facebook, they may unknowingly expose details such as the name of the hotel they stayed at, the license plate of a rental car, or even their face in front of a famous landmark, which could potentially be used for re-identification. Our system can detect these privacy-sensitive elements and alert users, allowing them to share their travel experiences while maintaining their privacy.
- Event photography: At events like concerts or conferences, people often take and share photos. These photos might contain the faces of other attendees, their badges with personal information, or even financial transaction receipts left in the venue. Our proposed concept can analyze these images in real-time as users attempt to share them, ensuring that no sensitive information is inadvertently shared.
- Home-based content sharing: With the rise of home-based businesses and online marketplaces, users may share photos of their homes or products. These images could reveal their home address, personal belongings with identifying marks, or unique

features of their living environment. Our technology can help users protect their privacy by identifying and flagging such details before sharing.

This study focuses on the application of deep learning technology to detect fine-grained privacy in visual content on social media platforms. Deep learning has demonstrated significant effectiveness in various machine learning domains [1,7–9]. However, limited research has been conducted on multi-label privacy image classification. The application of deep learning in visual privacy detection and measurement is explored, with improvements in prediction accuracy achieved by focusing on different regions of an image using a privacy-specific spatial attention mechanism, highlighting relevant privacy attributes. In addition, this study uses set pair analysis (SPA) theory, which evaluates the similarity, difference, and opposition between system elements. This provides a novel approach for analyzing privacy risks, effectively quantifying privacy risks, and offering guidance to users regarding the publication of sensitive content.

An automated mechanism combining deep learning and SPA is proposed in this study to predict and measure privacy in visual content on social media platforms. This approach alleviates the burden on users to manually inspect historical UGC to prevent privacy leaks while providing a more precise assessment of users' privacy.

Our main contributions in this paper are outlined as follows:

- Privacy-Specific Spatial Attention: Privacy-specific spatial attention weights are generated for each privacy attribute label. In multi-label privacy classification, these attention weights enable the model to focus on image regions relevant to specific privacy attributes, thereby improving its ability to capture fine-grained privacy information. Furthermore, the performance of various backbone networks for this task is explored, with results showing an improvement in accuracy compared to previous works.
- Proposing an End-to-End Framework for Fine-Grained Privacy Detection and Quantification: We introduced an end-to-end framework that leverages deep learning's multi-label classification capabilities in conjunction with set pair analysis (SPA) theory. This framework demonstrates the ability to effectively capture various privacy components within images and, importantly, to estimate the overall privacy level of each image. The integration of these techniques represents a meaningful advancement in automating and quantifying visual privacy assessment.
- Experimental Validation on Real-World Datasets: To demonstrate the effectiveness of our proposed image privacy assessment framework, we designed and conducted rigorous experiments on **VISPR** [10] (**vis**ual **pr**ivacy). The results indicate that our approach performs well in the challenging task of fine-grained visual privacy detection, highlighting its practical value and potential applicability in scenarios where privacy preservation is a key consideration.

2. Related Work

2.1. Deep Privacy Prediction

The advent of deep learning has revolutionized the landscape of image privacy perception, enabling more sophisticated and accurate analyses. In 2016, Tran introduced the use of convolutional neural networks for image privacy perception, employing the AlexNet architecture and a facial emotion analysis model [11], marking a significant advancement in this area. This field of research has expanded to various domains, including facial recognition [12,13], license plate recognition [14], and social relationship analysis [15–17]. A particular phenomenon noted in these studies is the "privacy paradox", where users continue to share images despite being aware of the associated privacy risks. This highlights the complexity and challenges in predicting user behavior concerning privacy [18–20].

In this study, we enhance the accuracy of privacy attribute prediction by incorporating spatial privacy attention and introduce a novel challenge in computer vision designed to help users assess privacy risks before sharing images on social media, encompassing a broad range of personal information in a single framework.

2.2. Privacy Measurement

Research on privacy has mainly focused on privacy protection, with relatively less emphasis on privacy risk measurement, primarily centering on the privacy measurement of location or trajectories through metrics such as information entropy, the probability of inferring the actual location, and the duration of sustainable trajectories [21–26]. In the scenario of privacy-protected neural network inference, the accuracy and efficiency of privacy measurements are of crucial importance. Although existing technologies are dedicated to protecting data privacy, they have certain limitations. For example, Jun Feng [27] pointed out that the current state-of-the-art secure two-party neural network (2P-NN) inference technology incurs significant computational and communication overhead when dealing with ImageNet-scale deep neural networks. The Panther system they proposed offers new ideas for solving this problem. Differential privacy (DP) is a commonly used privacy-preserving technique that reduces data leakage risks by adding noise and has been widely applied in areas such as federated learning. However, DP has certain limitations in privacy measurement. First, DP focuses on data protection, while privacy measurement aims to assess privacy risks, making their objectives different. Second, DP relies on noise injection, which may affect the precise calculation of privacy risks. Lastly, DP is primarily designed for structured data, while privacy measurement often involves unstructured data, like images, requiring deeper feature extraction and semantic understanding. Therefore, while DP is an important privacy-preserving approach, its direct application in privacy measurement remains limited, and future work can explore its integration with privacy risk assessment. A privacy quantification model was proposed, analyzing users' privacy concerns using theories of commercial rights and data statistics [28]. A noise-based data perturbation technique was adopted to assess the degree of privacy by evaluating the closeness between the original data within the interval and the perturbed data [29]. A privacy index function was introduced to evaluate users' privacy exposure; however, it was limited to a single social network [30]. Privacy in published tweets can be quantified to a certain degree of uncertainty through the mention of private information like names and addresses [31]. The use of SPA and other mathematical models in privacy measurement represents innovative approaches to handling uncertainty in cross-domain data sharing [32,33], particularly noting SPA's applicability in dynamic, uncertain scenarios. Recent methodologies based on SPA theory addressed the challenge of measuring privacy when background knowledge is unclear, allowing for more dynamically adaptable privacy analysis [34]. The development of privacy measurement techniques, including those based on the theory of anonymous sets and event entropy, underscores the ongoing need for effective tools to detect and manage privacy leaks.

However, previous works on privacy analysis have mainly focused on coarse-grained privacy detection, limiting the depth of their implications. To address this, Song proposed a multi-task learning model guided by a classification of privacy-oriented features crafted for predicting personal aspects revealed in posts, introducing a comprehensive taxonomy for representing user privacy [35]. WQ Huang demonstrated the applicability of set pair analysis for privacy measurement under conditions of uncertain background knowledge, a foundation we build upon to further analyze privacy metrics in emerging social media platforms [36]. Although pioneer studies achieved significant milestones, they predominantly utilized shallow learning methods and a set of handcrafted privacy-oriented features.

The work most closely related to ours is [10], where they focus on developing privacy advisors and conducting user studies to personalize predictions based on users' privacy preferences. However, their approach heavily relies on users' privacy preferences when predicting privacy scores, leading to a lack of objectivity in the privacy risk scores and poor interpretability. In this paper, we focus on leveraging an attention mechanism to enhance the performance of fine-grained privacy detection and explore personal privacy quantification in conjunction with SPA.

In order to align with users' specific privacy preferences, we draw on the user research method in the work of Orekondy et al. We recruited 305 AMT workers to rate 67 privacy attributes (excluding the "safe" attribute) on a scale of 1–5 to evaluate the degree of privacy violation when details of the attributes are accidentally disclosed. This is used to collect users' preference data. Subsequently, we use principal component analysis (PCA) to transform these related ratings into uncorrelated principal components and determine the weights based on the absolute values of the coefficients of each attribute in the principal components. Attributes with large absolute coefficient values, such as "credit card", have higher weights in the calculation of comprehensive privacy risks. In this way, the determined weights are closely related to users' preferences, enhancing the interpretability of the weight assignment process.

2.3. Visual Privacy Datasets

Datasets are often a major constraint in the development of machine learning, and detecting multiple privacy attributes heavily relies on high-quality datasets. Privacy tasks often rely on images that reveal sensitive details, such as faces, names, or opinions. However, many existing datasets do not provide enough of these types of images to effectively study privacy-related issues. Although some datasets [37] include such data, they tend to be either too small or not representative of real-world social media images. The PIPA dataset [15,38], which consists of 37,107 Flickr images for people recognition in unconstrained settings, is the closest match. However, it does not cover other critical privacy aspects like license plates, political opinions, or official identification documents. The PE-ViD video dataset [39] focuses on person-centric bounding box annotations across 20 video sequences in controlled settings. Visual privacy datasets such as PicAlert [17] and YourAlert [40] offer user-classified privacy labels, while VISPR [10] provides a more extensive set of 22k images annotated with a variety of privacy-related labels. However, both the PIPA and PicAlert datasets are currently unavailable, which limits related research.

2.4. Use of AI Tools

In the preparation of this manuscript, the authors primarily drafted their content without significant reliance on generative AI (GenAI) tools. However, during the writing process, AI-assisted tools were used for specific purposes. The authors used ChatGPT (OpenAI) to refine language, check grammar, and improve the clarity of technical descriptions. All AI-generated suggestions were carefully reviewed and revised by the authors to ensure technical accuracy and adherence to scientific integrity. Additionally, Zotero (or another citation manager) was utilized for managing and formatting references, ensuring correct citation formats without altering the content of the references.

All scientific content, data analysis, and conclusions in this manuscript were developed and verified solely by the authors, who bear full responsibility for the submitted work.

3. Visual Privacy Prediction and Measurement

The complete process of privacy detection and measurement is illustrated in Figure 2. Our research approach and technical procedure are as follows: (i) Privacy-specific spatial

attention mechanism: First, to more accurately capture privacy information in images, we propose a privacy-specific spatial attention mechanism. Convolutional neural networks (CNNs) are a class of deep learning models that have shown remarkable performance in image processing tasks. They use convolutional layers to automatically learn hierarchical features from images, which are particularly effective in extracting visual patterns. Based on the foundation of CNNs, our proposed mechanism automatically identifies privacy-sensitive regions in the image and assigns higher attention weights to these areas. Through the spatial attention layer, the model can focus on regions related to privacy, enhancing its ability to capture privacy attributes. (ii) Privacy measurement model based on SPA: After obtaining the privacy attribute vector of the image, we establish a set theory model to describe the inclusion and intersection relationships among the privacy attributes of the image, and use set operations to perform privacy measurement, calculating the overall level of privacy contained in the given image. This model refers to the measurement methods in set pair analysis theory. (iii) Generate privacy recommendations: Based on the privacy level of the image, our system can determine whether the given image can be safely shared or needs to be modified to some extent.

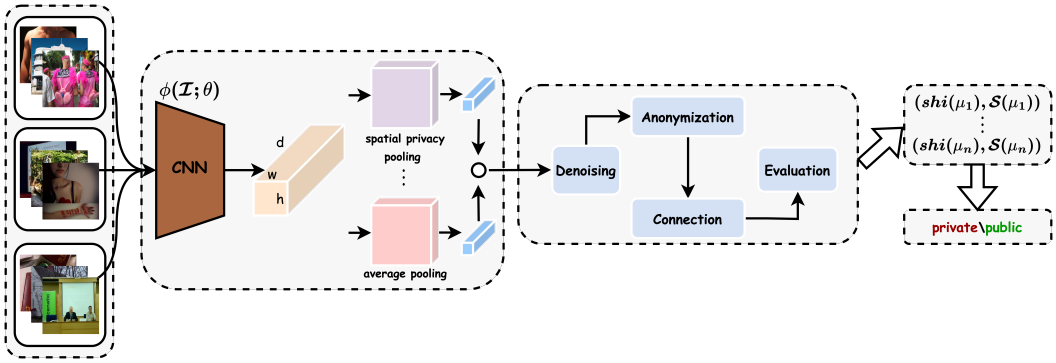

Figure 2. Employing automated deep learning and set pair analysis techniques to enhance privacy measurement practices on social media platforms.

Through this technical process, we aim to model and measure image privacy more precisely while ensuring computational efficiency. This plays a significant role in guiding users to share privacy-sensitive images and achieving controllable privacy measurements. During the experimental phase, a pre-trained multi-label classification model will be utilized to extract privacy components directly from users' visual privacy datasets.

3.1. Privacy Attribute Prediction

As shown in Table 1, we compare the performance of various models based on the mean average precision (mAP) metric, evaluated on the test set. The results highlight the superior effectiveness of our proposed method, ResNet-50+PSSA, which achieves an mAP of 46.88. This represents a significant improvement over all other methods tested.

When compared to traditional SVM-based methods (refer to machine learning approaches that utilize support vector machines (SVMs), which are supervised learning models used for classification and regression analysis), which yield mAP scores of 37.93 for CaffeNet and 39.88 for GoogleNet, our deep learning models show a clear advantage. For instance, ResNet-50, a widely used backbone, achieves an mAP of 40.50, while our ResNet-50+PSSA model surpasses it by 6.38 points. Furthermore, even when compared to other state-of-the-art models such as ResNet-101, which has an mAP of 40.50, and GoogleNet, which scores 41.20, our method demonstrates a clear performance boost. In addition, the VIT-B16-224 model has an mAP of 42.50, and the VIT-L16-224 model attains an mAP of 43.80. However, our ResNet-50+PSSA model still outperforms them with its mAP of 46.88, further highlighting the superiority of our proposed approach.

Table 1. Accuracy of our methods given by mean average precision and other metric scores, evaluated on test. The highest and second-highest accuracies under each setting are in **bold** and underlined, respectively.

Training	Backbone	*mAP* (%)	CP	CR	CF1	OP	OR	OF1
SVM	CaffeNet	37.93	0.65	0.62	0.63	0.55	0.52	0.53
	GoogleNet	39.88	0.68	0.65	0.66	0.58	0.55	0.56
	Resnet-50	40.50	0.70	0.67	0.68	0.60	0.57	0.58
End-to-End	ResNet-18	37.93	0.65	0.62	0.63	0.55	0.52	0.53
	CaffeNet	40.91	0.72	0.69	0.70	0.62	0.59	0.60
	GoogleNet	41.20	0.73	0.70	0.71	0.63	0.60	0.61
	ResNet-34	39.88	0.68	0.65	0.66	0.58	0.55	0.56
	ResNet-50 [10]	44.91	0.78	0.75	0.76	0.68	0.65	0.66
	ResNet-101	40.50	0.70	0.67	0.68	0.60	0.57	0.58
	VIT-B16-224	42.50	0.75	0.72	0.73	0.65	0.62	0.63
	VIT-L16-224	43.80	0.77	0.74	0.75	0.67	0.64	0.65
	ResNet-50+PSSA	**46.88**	**0.82**	**0.79**	**0.80**	**0.72**	**0.69**	**0.70**

It is also noteworthy that the ResNet-50 model presented in Orekondy [10] achieves an mAP of 44.91, but our enhancement with the PSSA mechanism leads to a further improvement, bringing the mAP up to 46.88, surpassing previous work and demonstrating the effectiveness of our approach.

Thus, the integration of the PSSA mechanism with ResNet-50 significantly improves the model's ability to detect and extract privacy attributes, establishing it as the most effective method in this evaluation.

3.1.1. Spatial Attention Mechanism for Privacy Attribute Prediction

In the task of multi-label privacy attribute extraction, a primary challenge is that privacy-sensitive information is distributed unevenly across the image. Certain regions might contain sensitive data, such as faces, IDs, or other personal information, while other areas might not include any relevant privacy data. To address this, we propose a privacy-specific spatial attention (PSSA) mechanism that allocates varying attention weights to different image regions, thereby enhancing the model's ability to focus on areas containing privacy-relevant information.

Given an image I, we first process it through a feature extractor (CNN backbone) ϕ to obtain a feature tensor $\mathbf{X} \in \mathbb{R}^{d \times h \times w}$, where d, h, and w represent the depth, height, and width of the feature map, respectively, as follows:

$$\mathbf{X} = \phi(I; \theta) \tag{1}$$

Here, θ is the parameter set of the CNN backbone. We use ResNet-50 [41] as the backbone with an input resolution of 224×224. This results in a feature tensor with the shape $2048 \times 7 \times 7$, which can be divided into 49 feature patches $\mathbf{x}_i \in \mathbb{R}^{2048}$ for $i = 1, 2, \ldots, 49$.

To address the challenge of a spatially uneven privacy attribute distribution, we introduce the PSSA mechanism. The core idea is to apply a different attention weight to each region of the image based on its relevance to specific privacy attributes. We define privacy-specific attention scores α_j^i for the i-th privacy attribute at the j-th spatial location, which represents the importance of the i-th privacy attribute at location j:

$$\alpha_j^i = \frac{\exp\left(T \mathbf{x}_j^\top \mathbf{w}_i\right)}{\sum_{k=1}^{49} \exp\left(T \mathbf{x}_k^\top \mathbf{w}_i\right)} \tag{2}$$

where $\sum_{j=1}^{49} \alpha_j^i = 1$, and $T > 0$ is the temperature parameter controlling the sharpness of the attention distribution. The attention score α_j^i can be interpreted as the likelihood of the i-th privacy attribute being present at the spatial location j.

Next, we calculate the privacy-specific feature vector $\mathbf{a}^i$ for the i-th privacy attribute by taking a weighted sum of the feature tensor across all spatial locations:

$$\mathbf{a}^i = \sum_{k=1}^{49} \alpha_k^i \mathbf{x}_k \tag{3}$$

We also compute the global feature vector $\mathbf{g}$, which captures the overall features of the image, as follows:

$$\mathbf{g} = \frac{1}{49} \sum_{k=1}^{49} \mathbf{x}_k \tag{4}$$

By combining the global feature vector $\mathbf{g}$ and the privacy-specific feature vector $\mathbf{a}^i$, we obtain the final privacy-specific spatial attention (PSSA) feature $\mathbf{f}^i$ for the i-th privacy attribute:

$$\mathbf{f}^i = \mathbf{g} + \lambda \mathbf{a}^i \tag{5}$$

Here, λ is a hyperparameter that determines the relative contribution of the global and privacy-specific features. This combination enables the model to focus on both the global context of the image and the privacy-relevant details, thereby improving the ability to extract privacy information.

3.1.2. Explanation of the PSSA Module

We now provide a more detailed explanation of the PSSA module. First, we note that the process in Figure 3 is a simplified version of the PSSA mechanism. By substituting Equations (2)–(5) into Equation (6), we can derive the logit for the i-th privacy attribute as

$$y^i = \mathbf{w}_i^\top \mathbf{g} + \lambda \mathbf{w}_i^\top \sum_{k=1}^{49} \alpha_k^i \mathbf{x}_k \tag{6}$$

The first term $\frac{1}{49} \sum_{k=1}^{49} \mathbf{x}_k^\top \mathbf{w}_i$ represents the global logit, while the second term reflects the weighted residual information, where the attention scores α_k^i adjust the influence of each spatial location based on its relevance to the i-th privacy attribute.

As $T \to \infty$, the softmax output $\alpha_k^i = \frac{\exp\left(T \mathbf{x}_k^\top \mathbf{w}_i\right)}{\sum_{l=1}^{49} \exp\left(T \mathbf{x}_l^\top \mathbf{w}_i\right)}$ converges to a Dirac delta function, concentrating the attention on the most prominent spatial region. In this scenario, the logit for the i-th privacy attribute becomes

$$y^i = \mathbf{w}_i^\top \mathbf{g} + \lambda \max\left(\mathbf{x}_1^\top \mathbf{w}_i, \ldots, \mathbf{x}_k^\top \mathbf{w}_i\right) \tag{7}$$

This corresponds to the $\lambda \times Y_{\max}$ term in Figure 3, suggesting that the PSSA module functions similarly to max pooling, where attention is focused on the most significant region of the image.

3.1.3. Multi-Head Attention Extension

To improve the model further, especially in scenarios where privacy attributes are distributed unevenly across different regions of the image, we introduce a multi-head attention extension. This extension uses multiple attention branches, each with a different temperature T, enabling the model to capture varying levels of attention across different parts of the image. We denote the number of attention heads as H, where each head corresponds to a distinct temperature.

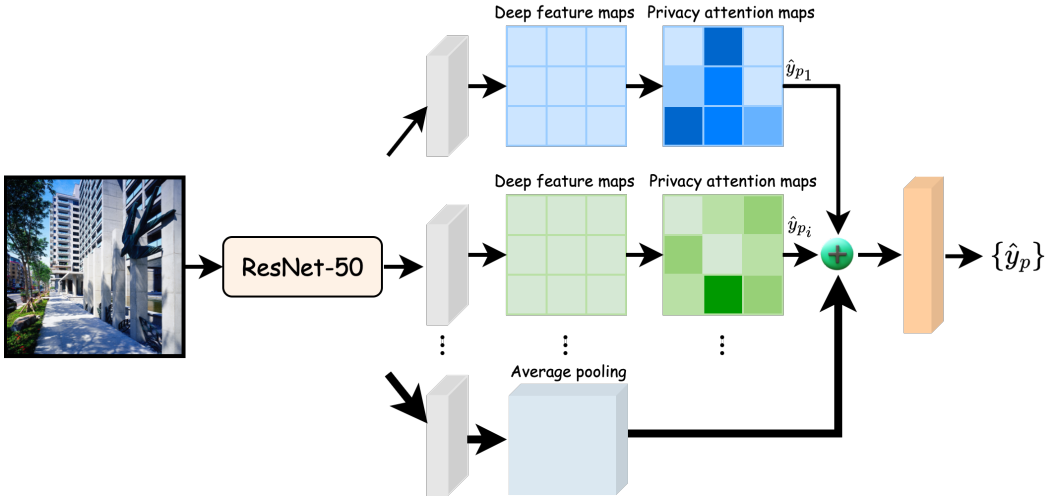

Figure 3. The structure of privacy-specific spatial attention.

For example, when $H = 2$, we set $T_1 = 1$ and $T_2 = \infty$ (i.e., max pooling); when $H = 4$, the temperatures are $T_{1:3} = 1, 2, 4$ and $T_4 = \infty$; and so on. This multi-head attention approach allows the model to capture different levels of attention across the image, improving the representation of privacy attributes.

By leveraging multi-head attention, the model becomes more robust, as it can better focus on different levels of granularity, improving the overall extraction of privacy-related features and making the model more adaptable to a variety of privacy data distributions.

Table 2 presents the impact of different numbers of attention heads (H) and temperature (T) settings on the mean average precision (mAP) of the model in the privacy-specific spatial attention (PSSA) mechanism. The experiments cover scenarios ranging from single-head attention to four-head attention. For single-head attention, the temperature settings include the default temperature, $T = 1$, and $T \to \infty$ (equivalent to max pooling). In the case of multi-head attention, different temperature combinations are set, such as $T_1 = 1$, $T_2 \to \infty$ for two-head attention and $T_1 = 1$, $T_2 = 2$, $T_3 = 4$, $T_4 \to \infty$ for four-head attention. The results indicate that as the number of attention heads increases and the temperature settings are optimized, the mAP of the model gradually improves. The highest mAP of 46.88% is achieved with four-head attention and multiple temperature settings, suggesting that the combination of multi-head attention and different temperature settings can effectively enhance the model's performance.

The structure of the privacy prediction network is shown in Figure 3.

- Feature Extraction: Images are processed through a pre-trained CNN (e.g., ResNet50) to extract deep feature maps.
- Spatial Attention Generation: For each privacy attribute label, a spatial attention map is calculated using the feature map. This is accomplished by applying a lightweight convolutional network over the feature map, which learns to identify image regions associated with specific privacy attributes.
- Combining Privacy-Specific and Global Features: Each attention map is merged with global features (obtained via applying global average pooling on the feature map), through a weighted sum (where weights are the learned attention scores) to produce enhanced privacy-specific features.
- Model Output: The enhanced features are fed into the classification layer, typically one or several fully connected layers, to predict the presence of each privacy attribute in the image.

Table 2. Ablation study of temperature and attention heads in privacy-specific spatial attention mechanism using ResNet-50.

Number of Attention Heads (H)	Temperatures (T)	Mechanism	mAP (%)
1	Default	Single-head attention with default temperature	46.33
1	$T = 1$	Single-head attention with temperature $T = 1$	46.45
1	$T \to \infty$	Single-head attention equivalent to max pooling	46.50
2	$T_1 = 1, T_2 = \infty$	Two-head attention with $T_1 = 1$ and $T_2 \to \infty$	46.80
4	$T_1 = 1, T_2 = 2,$ $T_3 = 4, T_4 = \infty$	Four-head attention with multiple temperature settings	46.88

The automatic privacy extraction algorithm, as shown in Algorithm 1, presents a method for constructing a visual privacy dataset. This method takes a collection of original images and a list of privacy attributes as input, and through the processes of privacy region detection, annotation, and exclusion of non-private images, outputs a visual privacy dataset suitable for training.

Specifically, the algorithm first initializes the visual privacy dataset. Then, it iterates over the collection of original images, performing region detection for each attribute defined in the privacy attribute list to determine whether the image contains regions corresponding to those privacy attributes. If an image contains privacy attributes, the region is annotated and saved as a sample of the visual privacy dataset; if an image does not contain privacy attributes, it is saved as a non-privacy sample, and the generated non-privacy images are outputted to enrich the dataset. After processing all original images, the result is a visual privacy dataset that includes both annotated privacy images and generated non-privacy images.

The innovation of this method lies in its use of both privacy region detection and annotation as well as non-privacy content generation to construct the dataset. Privacy region detection allows for the annotation of sensitive information within images, while non-privacy content generation enriches the sample data, aiding the model trained on these dataset in distinguishing between privacy and non-privacy content. By detecting, annotating, and generating image regions, this method can effectively construct a visual privacy dataset, providing input for subsequent privacy measurement.

3.2. Measuring Visual Privacy with SPA

3.2.1. Basic Concept of SPA

During the data-sharing phase of image privacy, reducing the features that distinguish individuals within the shared dataset can reduce the distinguishability between individuals, thereby increasing the anonymity of user data. Considering the distinguishability metric and the advantage of SPA theory in addressing uncertainty issues, we use SPA theory from a database perspective to measure the privacy quantity of image data in the sharing phase.

Algorithm 1 Visual privacy prediction with privacy-specific spatial attention (PSSA)

Input: $U = \{u_1, u_2, \ldots, u_n\}$: user set
$O = \{o_1, o_2, \ldots, o_n\}$: original image set owned by users
$L = \{PA_0, PA_1, \ldots, PA_9\}$: list of privacy attributes, where PA_0 represents non-private
T: temperature parameter for spatial attention mechanism
λ: weight parameter for combining global and local features
Output: Visual privacy datasets S_p (private) and S_s (safe)

Initialize $S_p \leftarrow \emptyset$ and $S_s \leftarrow \emptyset$
for $u_i \in U$ **do**
  for each image $o_j \in o_i$ **do**
    is_private $\leftarrow$ False
    Extract feature tensor $\mathbf{X} = \phi(o_j; \theta)$ using deep learning model ϕ
    Compute attention scores for each privacy attribute using PSSA mechanism
    for each $PA_k \in L$ **except** PA_0 **do**
      Compute spatial attention $s_j^k = \frac{\exp(T\mathbf{x}_j^\top \mathbf{w}_k)}{\sum_{k=1}^{49} \exp(T\mathbf{x}_k^\top \mathbf{w}_k)}$
      Aggregate privacy features: $\mathbf{a}^k = \sum_{k=1}^{49} s_k^k \mathbf{x}_k$
      Combine global and local features: $\mathbf{f}^k = \mathbf{g} + \lambda \mathbf{a}^k$
      if Attribute PA_k identified in image o_j **then**
        is_private $\leftarrow$ True
        Break
      end if
    end for
    if *is_private* **then**
      Add o_j to S_p
    else
      Add o_j to S_s
    end if
  end for
end for
$S \leftarrow (S_p, S_s)$
return S

Set pair analysis, first introduced by Zhao in 1989, primarily addresses issues of uncertainty by abstracting them into two related sets, A and B, and characterizing the relationship between the sets from three perspectives: identity, discrepancy, and contrary. This approach assesses the development trends of the research issue. In the field of privacy measurement, this theoretical framework exhibits unique advantages by quantifying the connection between personal data and potential privacy risks, thus enhancing the understanding of the nature of privacy.

- The "identity" indicator represents the degree of consistency in common attributes between two sets, used in privacy measurement to evaluate the similarity between a personal dataset and a known safe dataset. This aids in identifying the effectiveness of data anonymization and is an important indicator for assessing the success of personal information anonymization and de-identification processes.
- The "discrepancy" indicator reveals the uncertain association between two sets, meaning that, in terms of privacy measurement, it measures the uncertainty between a personal dataset and potential privacy risks. This indicator is crucial for assessing the risk of data being unauthorizedly accessed or leaked, helping to build safer data processing.
- The "contrary" indicator assesses the negative correlation between two sets, highlighting the adversarial relationship between data utility and privacy management measures in privacy evaluation. By examining the balance between the value derived

from data use and privacy measures, this indicator is crucial to ensuring that data retains its usefulness and accuracy while being managed for privacy.

These three types of indicators in SPA applications in privacy measurement not only help to deeply explore the inherent connection between certainty and uncertainty of personal datasets and privacy risks but also effectively evaluate the effectiveness of privacy measures and the safety of data processing behaviors. By calculating the consistency, difference, and opposition indicators of "identity", "discrepancy", and "contrary" between the personal dataset and reference sets (such as safe datasets, and risk datasets), SPA provides a quantitative, precise, and intuitive tool, significantly enhancing the theoretical basis and practical methods of privacy measurement. The application of this method is not only suitable for privacy risk assessment of large-scale datasets but also provides an accurate basis for formulating and evaluating the appropriateness of data processing strategies, offering a powerful analytical tool for evaluating personal privacy and data security.

3.2.2. Connection Number

The concept of a connection number is an essential part of set pair analysis (SPA). Its general expression is as follows:

$$\mu = \frac{S}{N} + \frac{F}{N}i + \frac{P}{N}j \tag{8}$$

where N represents the total number of features in all object sets, S is the number of identical features in two object sets, P is the number of different features, and F is the number of features that are neither identical nor different. This relationship can be expressed as $N = S + P + F$. Let $a = S/N$, $b = F/N$, and $c = P/N$; thus, Equation (8) can be simplified to the following form of a connection number:

$$\mu = a + bi + cj \tag{9}$$

According to Equation (9), we know $a + b + c = 1$, and a, b, $c \in [0, 1]$, $i \in [-1, 1]$, $j = -1$. Here, a represents the part of identity, with a coefficient of 1, belonging to the positive level; b represents the part of discrepancy, with a coefficient having a range of $[-1, 1]$, lying in the uncertain level; and c represents the part of contrariness, with a coefficient of -1, entirely at the negative level. According to Equation (9), the multivariate connection number can be expressed as follows:

$$\mu = a + b_1 i_1 + b_2 i_2 + \ldots + b_n i_n + cj \tag{10}$$

where $a + b_1 + b_2 + \ldots + b_n + c = 1$, and $a, b_1, b_2 \ldots b_n, c \in [0, 1], i_p \in \left[-1 + \frac{2(p-1)}{n}, -1 + \frac{2p}{n}\right]$, $(p = 1, 2, \ldots, n), j = -1$. The notation $b_1 + b_2 + \ldots + b_n$ in Equation (10) represents the extension of $'b_i'$.

The presence of dynamic variables $b_i(b_1 i_1, b_2 i_2, \ldots, b_n i_n)$ in the expression of the connection number is advantageous as adjusting the values of dynamic variables can adapt to the uncertainty of the problem. Thus, if X is a non-empty set, then $\Lambda = \{\langle x, a_\Lambda(x), b_\Lambda(x), c_\Lambda(x)\rangle \mid x \in X\}$ represents an SPA set, where $a_\Lambda(x)$, $b_\Lambda(x)$, and $c_\Lambda(x)$, respectively, represent the degree of support (identity), uncertainty (discrepancy), and opposition (contrariness) of element x in X, denoted as the degree of connection:

$$\begin{aligned} \mu_\Lambda(x) =& a_\Lambda(x) + b_\Lambda(x)i + c_\Lambda(x)j \\ & a_\Lambda(x) : x \to [0,1] \\ & b_\Lambda(x) : x \to [0,1] \\ & c_\Lambda(x) : x \to [0,1] \end{aligned} \tag{11}$$

which satisfies the normalization condition $a_\Lambda(x) + b_\Lambda(x) + c_\Lambda(x) = 1$. Here, $i \in [-1,1]$ is called the coefficient of uncertainty; j is the coefficient of opposition, usually taken as $j = -1$. Set pair analysis reflects the laws of change and the internal relationship between the fuzzy, random, certain, and uncertain aspects of things, making the SPA method theoretically and practically valuable for visual privacy measurement. In the previous section, we discussed extracting privacy attributes from images. In this section, we explore user-specific visual privacy feedback. The goal is to calculate a privacy risk score for each image, representing the privacy leakage(refers to the unauthorized access, use, or disclosure of sensitive personal information, posing a significant risk to individual privacy and potentially leading to financial loss, reputational damage, or other negative consequences) risk for a specific user. As shown in Algorithm 2, we convert the privacy attributes contained in the user-shared dataset into privacy scores. Within the framework of SPA theory, "identity", "discrepancy", and "contrary" constitute the three basic dimensions describing the relationship between two sets or objects. This theoretical approach is not only applicable to a wide range of scientific and engineering problems but also demonstrates its unique potential in the domain of visual privacy measurement. This paper explores how to use these three dimensions to quantify the relationship between individual image privacy features and the overall dataset privacy features, further assessing the potential risk of privacy leakage. Below is an in-depth analysis of the application of "identity", "discrepancy", and "contrary" in visual privacy measurement:

The application dimension of "identity": The "identity" dimension represents the consistency or similarity in specific attributes or features between two sets. In the context of SPA theory, this dimension is used to quantify the degree of match in shared features between two sets. Applied to visual privacy measurement, this dimension can assess the degree of match between privacy features in individual images and known privacy risk features. For example, if an image contains elements highly matched with a privacy-sensitive database (such as facial features or license plate numbers), the image can be deemed to have a higher privacy risk score due to its significant consistency in the "identity" dimension.

The application dimension of "discrepancy": The "discrepancy" dimension depicts the differences or inconsistencies in certain attributes or features between two sets, quantifying the uncertain association or partial match level between two sets. In the field of visual privacy measurement, this dimension helps identify those images that may contain privacy risks but do not completely match with known privacy-sensitive features. Such images, although containing privacy-related features, only partially match with features in the database, possibly representing a lower or uncertain risk of privacy leakage.

The application dimension of "contrary": The "contrary" dimension reflects the complete mismatch or opposition in specific attributes or features between two sets. In SPA theory, this dimension is used to describe the absolute inconsistency in key attributes between two sets. Applying this concept to visual privacy measurement allows for the identification of images that do not match any features in the privacy-sensitive feature database. If an image does not match any features in the database, it scores higher in the "contrary" dimension, indicating that the image likely does not contain privacy risks. By conducting a comprehensive quantitative analysis of the "identity", "discrepancy",

and "contrary" dimensions for each image in the dataset, this study can effectively assess the potential privacy risk level of each image, thereby facilitating the implementation of more precise and dynamic privacy measures. Figure 4 is used to briefly illustrate the identity/discrepancy/contrary relationships.

Algorithm 2 Visual privacy measurement with PCA and weighting

Input: Extracted visual privacy dataset S, privacy attribute set $A = \{A_1, A_2, \ldots, A_{67}\}$, desired cumulative variance ratio γ (e.g., $\gamma = 0.9$)
Output: P: privacy measurement for n users

eigenvalues $\leftarrow$ ComputeEigenvalues(A)
total_variance $\leftarrow \sum_{i=1}^{|\text{eigenvalues}|} \text{eigenvalues}_i$
Initialize an empty list variance_ratios
for $i = 1$ to $|$eigenvalues$|$ **do**
  ratio $\leftarrow \frac{\text{eigenvalues}_i}{\text{total\_variance}}$
  Append ratio to variance_ratios
end for
cumulative_variance $\leftarrow 0$
index $\leftarrow 0$
while cumulative_variance $< \gamma$ **do**
  cumulative_variance $\leftarrow$ cumulative_variance + variance_ratios[index]
  index $\leftarrow$ index $+ 1$
end while
$\tau \leftarrow$ variance_ratios[index $- 1$]
Initialize an empty list selected_attributes
for $i = 1$ to $|A|$ **do**
  if variance_ratios$[i] \geq \tau$ **then**
    Append $A[i]$ to selected_attributes
  end if
end for
total_selected_variance $\leftarrow \sum_{i=1}^{|\text{selected\_attributes}|} \text{variance\_ratios}[i]$
Initialize an empty list w
for $i = 1$ to $|$selected_attributes$|$ **do**
  $w_i \leftarrow \frac{\text{variance\_ratios}[i]}{\text{total\_selected\_variance}}$
  Append w_i to w
end for
Initialization of n empty subsets
while not at end of S **do**
  for each element $s_i \in S_p$ **do**
    $ide \leftarrow I(s_i)$
    $dis \leftarrow D(s_i)$
    $con \leftarrow C(s_i)$
    $p_i \leftarrow AGG(ide, dis, con)$
    $p_i' \leftarrow \sum_{i=1}^{|\text{selected\_attributes}|} w_i \cdot p_i$
  end for
  $Shi_i \leftarrow SYN(p_i')$
  $S_i \leftarrow SCO(p_i')\ i$
  $P_i \leftarrow \{Shi_i, p_i'\}$
end while
$P \leftarrow P_1 + P_2 + \ldots + P_n$
return P

In this study, we introduce a novel visual privacy measurement algorithm integrating deep learning with SPA theory. We develop a method to quantify image privacy risks by assessing privacy features within images and the dataset. We initialize n subsets based on dataset size and complexity, use deep learning models to extract privacy-sensitive features, assign samples to subsets based on similarity, and calculate identity (*ide*), discrepancy (*dis*),

and contrary (*con*) within subsets. These indicators are combined into a comprehensive privacy score (*Sub_score*) for each subset, which is then summarized to yield the final privacy measurement score P. The number of subsets n is carefully chosen for optimal analysis granularity and computational efficiency. Simultaneously, the weights of identity, discrepancy, and contrary are adjusted based on their importance contribution to privacy risk assessment, ensuring a reasonable reflection of different privacy features in the overall privacy measurement. Additionally, privacy feature extraction models that perform excellently in privacy-related tasks such as facial recognition and license plate recognition are chosen to ensure the accuracy and efficiency of the privacy feature extraction process. Through this method, we aim to provide an efficient and accurate quantitative assessment tool for visual privacy.

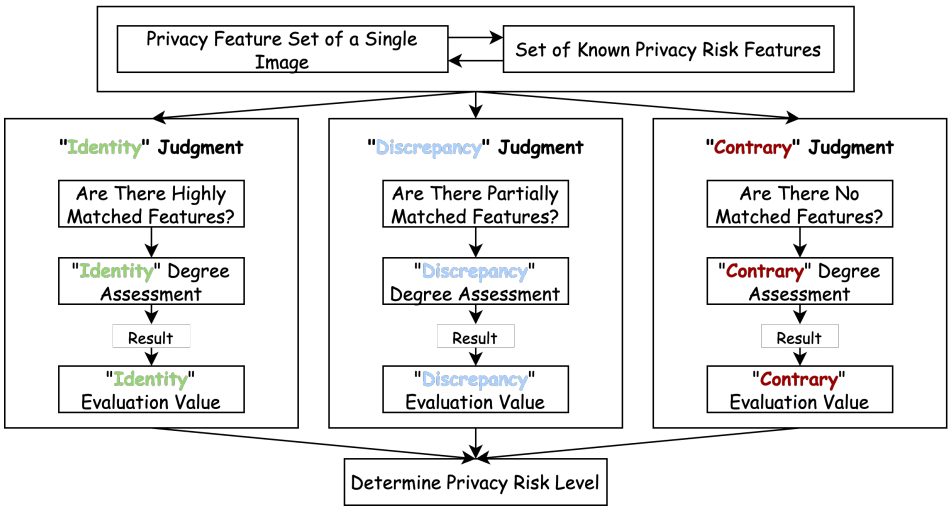

Figure 4. Intuitive diagrams of identity/discrepancy/contrary relationships in privacy risk assessment based on set pair analysis.

We incorporate a user study [10] using PCA to select privacy attributes from 67 options, identifying and weighting those above a threshold. This ensures that relevant features are in the model. We choose top-performing privacy feature extraction models for facial and license plate recognition to guarantee accuracy and efficiency. Our goal is to offer an effective quantitative tool for visual privacy assessment.

3.2.3. Evaluation Metrics

To assess the method's performance in the privacy extraction task, we calculate the average precision (AP) for each attribute, which is the area under the precision-recall curve, and we use the class mean average precision (C-MAP) as well. Furthermore, we incorporate the recall rate of the prediction task into the evaluation metrics. The experiment uses the **VISPR** image dataset, which contains a wide range of privacy information labels and undergoes scaling and normalization preprocessing. The model training employs the cross-entropy loss function, and evaluation metrics include accuracy, recall, F1-Score, and the introduced AP. The AP metric is particularly suitable for imbalanced datasets, calculating the AP value for each privacy attribute by computing the area under the precision-recall curve, thereby obtaining the mAP (mean average precision) and providing a quantification of the model's overall performance in identifying different privacy information in images [42–44]. Here, are the calculations for CP, CR, CF1, OP, OR, and OF1 metrics as derived from the document:

Per-Category Metrics

- Per-Category Precision (CP):

$$\mathrm{CP} = \frac{1}{C}\sum_{i=1}^{C}\frac{TP_i}{TP_i + FP_i} \tag{12}$$

where C is the number of categories, TP_i is the number of true positives for category i. FP_i is the number of false positives for category i.

- Per-Category Recall (CR):

$$\mathrm{CR} = \frac{1}{C}\sum_{i=1}^{C}\frac{TP_i}{TP_i + FN_i} \tag{13}$$

FN_i is the number of false negatives for category i.

- Per-Category F1-Score (CF1):

$$\mathrm{CF1} = \frac{1}{C}\sum_{i=1}^{C}\frac{2 \cdot \mathrm{Precision}_i \cdot \mathrm{Recall}_i}{\mathrm{Precision}_i + \mathrm{Recall}_i} \tag{14}$$

where $\mathrm{Precision}_i$ and Recall_i are the precision and recall for category i, respectively.

Overall Metrics

- Overall Precision (OP):

$$\mathrm{OP} = \frac{\sum_{i=1}^{C} TP_i}{\sum_{i=1}^{C}(TP_i + FP_i)} \tag{15}$$

- Overall Recall (OR):

$$\mathrm{OR} = \frac{\sum_{i=1}^{C} TP_i}{\sum_{i=1}^{C}(TP_i + FN_i)} \tag{16}$$

- Overall F1-Score (OF1):

$$\mathrm{OF1} = \frac{2 \cdot \mathrm{OP} \cdot \mathrm{OR}}{\mathrm{OP} + \mathrm{OR}} \tag{17}$$

These formulas help in evaluating the performance of multi-label classification models by considering both individual category metrics and overall metrics across all categories.

3.3. Complexity Analysis of the Proposed Approach

3.3.1. Time Complexity

- Privacy-Specific Spatial Attention Mechanism (PSSA): In the PSSA mechanism, the image is first processed through a feature extractor (such as ResNet-50). The time complexity of the forward propagation of ResNet-50 is $O(N)$, where N is the number of images. For each privacy attribute, the time complexity of calculating the attention scores is $O(h \times w \times d)$, where h, w, and d are the height, width, and depth of the feature map, respectively. When calculating the privacy-specific feature vector and the final PSSA feature, operations on the feature tensor are also involved, and their time complexity is also $O(h \times w \times d)$. Assuming there are M privacy attributes, the total time complexity of the PSSA mechanism is $O(N \times M \times (h \times w \times d))$.
- Multi-head Attention Extension: When the multi-head attention extension is introduced, since multiple attention branches are used, each with a different temperature T. If there are H attention heads, when calculating the attention scores, feature vectors, and other operations for each head, the time complexity becomes $O(N \times M \times H \times (h \times w \times d))$, which is higher than that of the single-head attention mechanism.

- Privacy Measurement Model based on SPA: In the privacy measurement model based on SPA, when calculating the connection number, each sample's features need to be processed. Assuming the number of samples is S and the number of features for each sample is n, the time complexity of calculating the connection number is $O(S \times n)$. When performing aggregation and evaluation operations, traversing and calculating the sample data are involved, and its time complexity is also $O(S \times n)$. Therefore, the total time complexity of the privacy measurement model based on SPA is $O(S \times n)$.

Overall, the time complexity of the entire method is mainly determined by the PSSA mechanism and the SPA-based measurement model. In practical applications, factors such as the number of images N, the number of privacy attributes M, the number of samples S, the number of features n, and the dimensions h, w, d of the feature map will affect the overall time complexity. In large-scale datasets and complex model settings, the time complexity may increase significantly, and it is necessary to consider optimizing the algorithm and using hardware acceleration to improve efficiency.

3.3.2. Space Complexity

- Privacy-Specific Spatial Attention Mechanism (PSSA): In the PSSA mechanism, it is necessary to store the parameters of the feature extractor, attention scores, privacy-specific feature vectors, and global feature vectors. The storage space for the parameters of the feature extractor (such as ResNet-50) is $O(P)$, where P is the number of model parameters. The storage space for the attention scores and feature vectors is related to the size of the feature map. Assuming the size of the feature map is $h \times w \times d$, the space complexity for storing these data are $O(h \times w \times d)$. For M privacy attributes, the total space complexity is $O(P + M \times (h \times w \times d))$.
- Multi-head Attention Extension: The multi-head attention extension will increase the storage space requirements. Since there are H attention heads, each head needs to store the corresponding attention scores and feature vectors. Therefore, the space complexity becomes $O(P + M \times H \times (h \times w \times d))$, and as the number of attention heads H increases, the space complexity will increase linearly.
- Privacy Measurement Model based on SPA: The privacy measurement model based on SPA needs to store connection numbers, subset data, and related intermediate calculation results. The storage of connection numbers is related to the number of samples S and the number of features n, and the space complexity is $O(S \times n)$. The storage of subset data and intermediate calculation results will also increase a certain amount of space overhead. Assuming it is $O(S \times m)$, where m is the number of features of the intermediate data, then the total space complexity of the privacy measurement model based on SPA is $O(S \times (n + m))$.

Overall, the space complexity of the entire method is affected by many factors, such as model parameters, feature map size, number of privacy attributes, number of attention heads, number of samples, and number of features. In practical applications, it is necessary to adjust the model parameters and data processing methods reasonably to avoid excessive space requirements. For example, model compression techniques and data dimensionality reduction methods can be used to reduce the storage space occupied.

4. Experiments and Discussion

4.1. Experimental Settings

In our experiments, we use the VISPR dataset [10], which consists of 22,000 images annotated with various privacy related labels. It contains various privacy attributes, such as faces, number plates, personal documents and other sensitive information, which is essential for evaluating privacy attribute extraction models.

To mimic real-world situations, we apply our proposed spatial attention mechanism to the VISPR dataset to detect and annotate privacy-sensitive regions in different image categories. We evaluate the effectiveness of our approach by comparing the accuracy of privacy attribute identification and region localization with existing baseline models. We also perform ablation studies to analyze the impact of different attention weighting temperatures and multi-head attention on model performance.

The VISPR dataset was constructed from publicly available images from the OpenImages dataset and Flickr. To meet ethical standards, we followed the guidelines for handling personally identifiable information (PII) in the EU Data Protection Directive 95/46/EC and the US Privacy Act of 1974. A user study was conducted with 305 Amazon Mechanical Turk (AMT) participants to assess privacy preferences. Although explicit IRB approval was not required for publicly available data, informed consent was obtained from participants prior to the survey.

To anonymize the data, all VISPR images were manually annotated with 68 privacy attributes. All images with explicit identifiers (such as full names, and passport numbers) were reviewed to avoid privacy violations. Although the dataset preserves visual privacy attributes for research purposes, researchers using VISPR are advised to use additional privacy-preserving techniques such as image blurring or redaction where necessary.

4.1.1. Training Details

The experimental setup involves training the model using a standard split of the VISPR dataset, with 80% of the images used for training and 20% for testing. We use a ResNet-50 backbone with the input image resolution set to 448×448 and implement the PSSA mechanism as part of the model's feature extraction process. The model is trained for 30 epochs using the Adam optimizer with a learning rate of 0.001, a momentum of 0.9, and a weight decay of 0.0001. The performance is evaluated based on standard metrics.

4.1.2. Privacy Prediction on VISPR

Due to the use of a self-constructed visual privacy dataset, the outcomes of the privacy perception task were not optimal. However, the model performed better on attributes with strong scene relevance and human-related properties, while it faced difficulties with attributes involving subtle differences and relationship inference. Quantitative results are displayed in Table 3, and Figure 5 presents evaluation metrics such as the average precision scores and recall rates for some privacy attributes. The following observations were made: (i) The privacy extraction model exhibited good performance with attributes that have high relevance to the scene, such as fingerprints. It also performed well in identifying human-related attributes, like facial information. (ii) Subtle differences led to confusion, for example, when predicting whether a license plate was complete or whether a subject was nude. (iii) Failures were observed due to details in the images, such as bank cards being recognized as square cards. (iv) Another limitation was the inability to recognize relationship-based attributes, which require reasoning based on the interaction of multiple visual cues within the image, not just their presence.

Table 3. Original privacy information from users.

Index	PA_0	PA_1	PA_2	PA_3	PA_4	PA_5	PA_6	PA_7	PA_8	PA_9
1	0.6635	0	0	0	0	0	0	0	0	0
2	0	0.6177	0	0.6421	0	0.5543	0.6363	0	0.5851	0
3	0	0	0	0.6124	0	0	0	0	0.6319	0
4	0	0	0	0.7143	0	0	0	0	0	0
5	0	0	0	0.6348	0	0.5680	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...

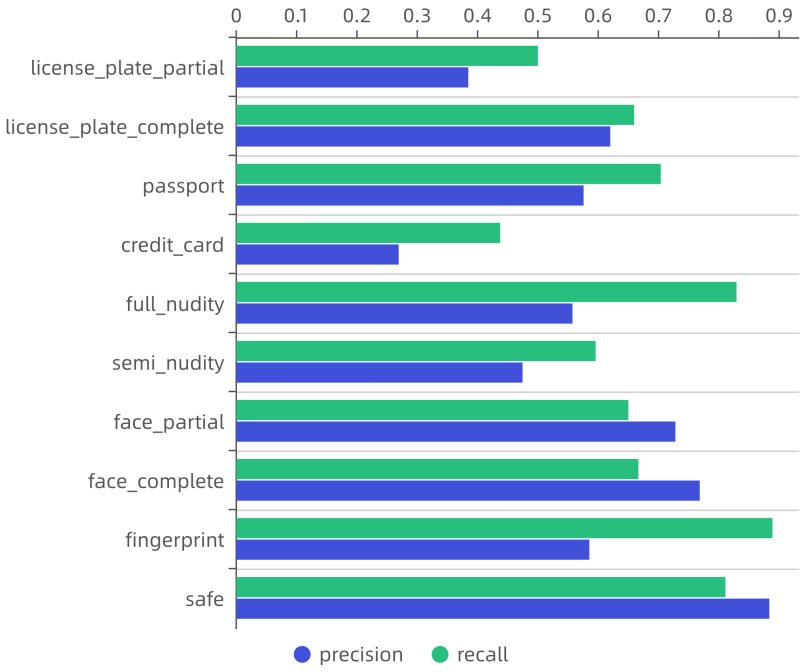

Figure 5. Average precision and recall scores.

4.1.3. Evaluation

Here, we list a portion of the original visual privacy data from the **VISPR**, which serves as the input for privacy quantification, as shown in Table 1.

Defining and quantifying visual privacy is no easy task. In our research, we use the predictive scores of each privacy attribute from a multi-label classification model as the extraction results for corresponding privacy components for two main reasons: (i) If we were to use binary outcomes (0 and 1) to represent the presence of a privacy component, it would diminish the variability in the predictive outcomes for the same class and reduce the usability of the data, offering no basis for subsequent data processing and analysis. (ii) Utilizing predictive scores as the extraction results of privacy components provides users with a measure akin to confidence levels. Higher scores indicate a greater certainty of the privacy prediction model regarding the presence of a privacy component. This enables users not only to make informed decisions about whether to rely on the model's output but also to have an initial assessment of the model's accuracy and provide more precise feedback, thus clarifying the direction for optimizing the model.

Subsequently, we measure privacy based on the extraction results, as shown in Figure 6. The measurement algorithm consists of four parts: preprocessing, anonymization, conjunction, and evaluation:

- Preprocessing: Due to inconsistencies and much noise in the data from Table 3, it is necessary to perform denoising. Then, for the reasons stated above, we use the predictive scores of each class from a multi-label classification model as the extraction results of corresponding privacy components. Additionally, to prevent individual attribute values from overly influencing the weight, we construct a ternary interval number with the minimum value $min\Lambda_i$, the maximum value $max\Lambda_i$, and the actual value Λ_i, i.e., $[\Lambda] = [min\Lambda_i, \Lambda, max\Lambda_i]$.
- Anonymization: In the ternary interval, min Λ_i represents the lower limit (also called the small element), max Λ_i represents the upper limit (also called the large element), and Λ_i is the value most likely to be taken on the interval, namely the preference value (also called the special element). However, the information on ternary intervals

has strong regularity, making it easy for attackers to identify. To better establish data privacy protection, we convert the ternary intervals into connection numbers.

- Conjunction: We aggregate privacy data, such as biometric and personal information, from the set pair connection number database. There are two common methods for data information aggregation: the mean value method and the set pair logical connection method. The mean value method calculates the mean connection number from the mean values of identity, discrepancy, and opposition in each set pair connection number. Given the minimal value of the privacy amount determines its privacy level, we adopt the conjunctive aggregation method to calculate the privacy amount.
- Evaluation: For the aggregated set pair connection numbers, we perform effective privacy measurement by defining the potential of the connection number $\mathrm{Shi}(\mu)$ and the scoring function $S(\mu)$. The potential and scoring function of connection numbers can categorize privacy protection issues into identity, discrepancy, and contrary, guiding the feasibility of privacy measurement publication. The potential or scoring function of a connection number reflects the trend of connection between information. When $\mathrm{Shi}(\mu) > 1$, the processing plan is considered feasible; when $\mathrm{Shi}(\mu) = 1$, it is average; and when $\mathrm{Shi}(\mu) < 1$, the plan is not feasible. Similarly, the scoring function $S(\mu)$ also reflects the trend of change in the connection between data information, with higher $S(\mu)$ values indicating closer connections and lower values indicating more distant relations. We present the results of the average connection number's potential and scoring function and their analysis in Table 4.

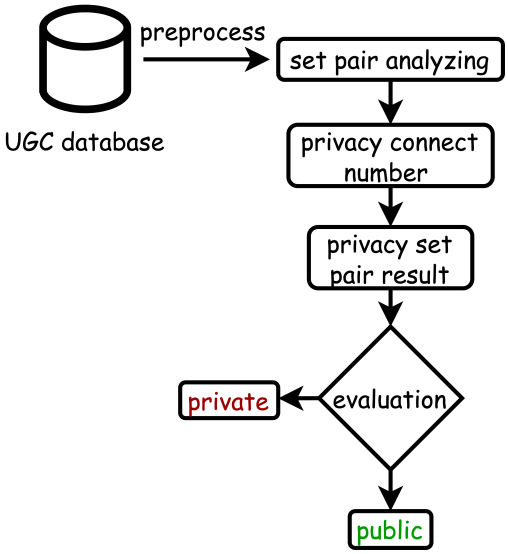

Figure 6. Set pair analysis.

Table 4. Set pair analysis contact numbers of private data after conversion.

Index	Contact Number	*Shi (μ)*	*S(μ)*	Result
1	–	–	–	public
2	0.6117 + 0.0550i + 0.3333j	1.8353	0.2784	private
3	0.4906 + 0.0650i + 0.4444j	1.1040	0.0462	private
4	0.3028 + 0.0305i + 0.6667j	0.4542	−0.3639	not sure
5	0.3903 + 0.0541i + 0.5556j	0.7025	−0.1653	not sure

In this study, we introduce the concept of "Synpotence" derived from set pair analysis theory, which offers a novel perspective to comprehend the multifaceted significance of varying types of privacy information embedded within images. Synpotence functions as a holistic measure, capturing the overall degree or weight of privacy information present in

images. It serves as a versatile tool that enables us to quantify and compare the relative importance of different privacy features, facilitating more informed decisions in the realm of image privacy protection. By employing synpotence, we can systematically assess the privacy content of images, taking into account both the intrinsic value of individual privacy features and their inter-relationships within the image context. This comprehensive approach allows us to identify patterns and trends that may not be immediately apparent through traditional analysis methods. The concept of synpotence is intricately linked with other technologies and theories, such as machine learning and privacy-preserving techniques. By integrating these advancements, synpotence enhances its capabilities to accurately evaluate privacy risks while preserving the integrity of sensitive information. For instance, when dealing with privacy-sensitive data, synpotence can facilitate the selection of appropriate anonymization or encryption strategies, ensuring that valuable insights are gained without compromising individual privacy. Specifically, the calculation of synpotence involves treating different types of privacy information as distinct datasets and then obtaining the synpotence value through a_A / c_A. Synpotence reflects the degree of convergence between two sets in relation to the research question, where "two sets" refers to the sets of privacy components extracted from privacy, and the "research question context" refers to the degree of privacy in visual privacy data, thereby indicating the level of privacy. The size of synpotence can reflect the magnitude of the privacy degree implied by the data.

Based on the results shown in Table 3, we have the following analysis: (i) For data that do not involve privacy components after privacy extraction, we directly conclude that the data are public. (ii) For data that contain privacy components after privacy extraction, we note that the data indexed 2 and 3 have higher *Shi* values, greater than 1, and *S* values greater than 0, indicating that these datasets contain a significant amount of privacy, consistent with the original data in Table 4. This implies a high level of privacy components, demonstrating the effectiveness of set pair analysis theory as a means of privacy measurement in this context. (iii) For the fourth and fifth sets of data, because the *Shi* values are less than 1 and the *S* values are negative, we categorize these datasets as uncertain regarding their privacy status, due to the lesser amount of privacy components contained, allowing users to decide based on the actual data whether to classify them as private.

4.1.4. Future Work in Privacy Measurement

This study achieved the quantification of privacy levels on a self-constructed visual privacy dataset; however, due to the poor generalizability of the multi-label classification model, it did not achieve optimal results. Therefore, in the future, we need to collect and construct richer and more diverse visual privacy datasets, including images under different scenarios, lighting conditions, etc., to improve the model's generalization ability. Moreover, we need to carry out and consider the following work in the future: Explore different neural network models and structures for privacy feature extraction, such as GANs, autoencoders, etc., to learn more abstract and semantic representations of privacy. Explore multi-task learning and transfer learning methods, using additional information to aid in privacy feature extraction and measurement, for example, auxiliary classification tasks. Design new measurement metrics, considering the accuracy, completeness, and attack resistance of privacy extraction, etc. Concepts like trustworthiness could be referred to. Explore how to explain privacy measurement results, provide intuitive privacy security tips, and improve user experience. Consider different privacy preferences, setting reasonable privacy measurement boundaries according to legal and ethical norms. Moreover, since the importance of privacy attributes also varies, future work should consider applying

weighted set pair analysis theory to calculate the privacy quantity of data, ensuring more precise judgments for the last two groups of data in Table 3. Enhancing the distinction in privacy measurement is also a major direction for future research.

In addition, the integration of PSSA-SPA introduces additional computational overhead, which can be challenging for real-time applications. Currently, we have not specifically tested whether the model can run in real time or if it is too resource-intensive for real-world deployment and suitable for edge devices. The privacy-specific spatial attention (PSSA) mechanism increases the complexity of feature extraction, while the set pair analysis (SPA) framework involves additional computations to quantify identity, discrepancy, and opposite relationships. Future research should focus on optimizing the efficiency of these processes by exploring model pruning, quantization, and hardware acceleration (e.g., GPU and TPU inference optimization) to reduce computational time.

Scalability is another critical factor for real-world deployment. To enable edge computing applications, such as privacy-aware content moderation on social media or mobile devices, lightweight model variants need to be explored. This could include reducing the number of attention heads in PSSA, approximating SPA computations with faster heuristics, or implementing a hierarchical privacy risk assessment approach that balances accuracy and efficiency. We plan to conduct comprehensive evaluations to determine the model's performance in real-time scenarios, resource requirements for deployment, its viability on edge devices, and incorporate the results into the revised manuscript.

In addition, latency analysis should be performed to assess the feasibility of real-time processing. Measuring the inference time per image and analyzing its impact on large-scale deployments can help determine whether the model is suitable for interactive privacy recommendation systems or batch processing scenarios. Addressing these challenges will be critical in making the proposed privacy risk assessment framework practical and efficient in real-world applications.

5. Conclusions

This study proposes a novel visual privacy measurement method integrating deep learning and set pair analysis (SPA) theory, featuring a framework to evaluate privacy risks through similarity, discrepancy, and opposition in privacy data subsets. It also includes a privacy-specific spatial attention (PSSA) mechanism, which is used to enhance focus on privacy-sensitive regions and improve fine-grained feature extraction, validated by superior performance on the VISPR dataset, advancing automated privacy assessment and privacy-aware systems in complex visual environments.

Author Contributions: In this study, H.J. was responsible for conceptualization, research methodology design, software programming, writing the original draft, and conducting review and editing. J.Z. provided supervision and guidance, ensuring the research direction and overall quality. Teacher Y.L. was in charge of acquiring research funding, which provided financial support for the smooth progress of the study. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the National Natural Science Foundation of China Youth Project (Grant No. 62402057) focusing on 'Personalized Privacy Metrics for Large-Scale Intelligent Connected Vehicles' and China's National Key Research and Development Program (Grant No. 2022YFB3104900) for 'Network Data Security Monitoring and Control in Intelligent Vehicles'. The APC was funded by the same funding sources. The authors thank Professor Lu and Dr. Zuo for their invaluable support, as well as the anonymous reviewers for their constructive feedback that enhanced this paper.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in this study.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: During the preparation of this manuscript, the authors primarily drafted the content without significant reliance on generative AI (GenAI) tools. However, AI-assisted tools were utilized for specific purposes. ChatGPT (OpenAI) was used for language refinement, grammar checking, and improving the clarity of technical descriptions. Zotero (or another citation manager) was employed for managing and formatting references. All AI-generated suggestions were carefully reviewed and revised by the authors to ensure technical accuracy and adherence to scientific integrity. The authors have reviewed and edited the final output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Fernandes, N.; Dras, M.; McIver, A. Generalised differential privacy for text document processing. In *Principles of Security and Trust, Proceedings of the 8th International Conference, POST 2019, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2019, Prague, Czech Republic, 6–11 April 2019*; Proceedings 8; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; pp. 123–148.
2. Humphreys, L.; Gill, P.; Krishnamurthy, B. How much is too much? Privacy issues on Twitter. In Proceedings of the Conference of International Communication Association, Singapore, 22–26 June 2010.
3. Gross, R.; Acquisti, A. Information revelation and privacy in online social networks. In Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society, Alexandria, VA, USA, 7 November 2005; pp. 71–80.
4. Li, M.; Cao, N.; Yu, S.; Lou, W. Findu: Privacy-preserving personal profile matching in mobile social networks. In Proceedings of the 2011 Proceedings IEEE INFOCOM, Shanghai, China, 10–15 April 2011; pp. 2435–2443.
5. Liu, Y.; Gummadi, K.P.; Krishnamurthy, B.; Mislove, A. Analyzing facebook privacy settings: User expectations vs. reality. In Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference, Berlin, Germany, 2–4 November 2011; pp. 61–70.
6. Sleeper, M.; Cranshaw, J.; Kelley, P.G.; Ur, B.; Acquisti, A.; Cranor, L.F.; Sadeh, N. "I read my Twitter the next morning and was astonished" a conversational perspective on Twitter regrets. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France, 27 April–2 May 2013; pp. 3277–3286.
7. Ganhör, C.; Penz, D.; Rekabsaz, N.; Lesota, O.; Schedl, M. Unlearning protected user attributes in recommendations with adversarial training. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, 11–15 July 2022; pp. 2142–2147.
8. Mosallanezhad, A.; Beigi, G.; Liu, H. Deep reinforcement learning-based text anonymization against private-attribute inference. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 2360–2369.
9. Xu, Q.; Qu, L.; Xu, C.; Cui, R. Privacy-aware text rewriting. In Proceedings of the 12th International Conference on Natural Language Generation, Tokyo, Japan, 29 October–1 November 2019; pp. 247–257.
10. Orekondy, T.; Schiele, B.; Fritz, M. Towards a visual privacy advisor: Understanding and predicting privacy risks in images. In Proceedings of the IEEE international Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3686–3695.
11. Tran, L.; Kong, D.; Jin, H.; Liu, J. Privacy-cnh: A framework to detect photo privacy with convolutional neural network using hierarchical features. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; Volume 30.
12. Sun, X.; Wu, P.; Hoi, S.C. Face detection using deep learning: An improved faster RCNN approach. *Neurocomputing* **2018**, *299*, 42–50. [CrossRef]
13. Zhao, X.; Yuan, J.Z.; Liu, H.; Zhou, J.S. Improved AdaBoost Algorithm for Robust Real-Time Multi-face Detection. *J. Softw.* **2017**, *12*, 53–61. [CrossRef]
14. Molina-Moreno, M.; González-Díaz, I.; Díaz-de María, F. Efficient scale-adaptive license plate detection system. *IEEE Trans. Intell. Transp. Syst.* **2018**, *20*, 2109–2121. [CrossRef]
15. Sun, Q.; Schiele, B.; Fritz, M. A domain based approach to social relation recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3481–3490.
16. Ng, W.W.; Zheng, T.M.; Chan, P.P.; Yeung, D.S. Social relationship discovery and face annotation in personal photo collection. In Proceedings of the IEEE 2011 International Conference on Machine Learning and Cybernetics, Guilin, China, 10–13 July 2011; Volume 2, pp. 631–637.

17. Wang, G.; Gallagher, A.; Luo, J.; Forsyth, D. Seeing people in social context: Recognizing people and social relationships. In Proceedings of the Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Greece, 5–11 September 2010; Proceedings, Part V 11; Springer: Berlin/Heidelberg, Germany, 2010; pp. 169–182.
18. Beigi, G.; Liu, H. A survey on privacy in social media: Identification, mitigation, and applications. *ACM Trans. Data Sci.* **2020**, *1*, 1–38. [CrossRef]
19. Gupta, R.; Saraf, D. Privacy and Security Challenges in Online social media: A Case Study Analysis. *Rev. Rev. Index J. Multidiscip.* **2023**, *3*, 1–7. [CrossRef]
20. Hu, X.; Zhu, T.; Zhai, X.; Zhou, W.; Zhao, W. Privacy data propagation and preservation in social media: A real-world case study. *IEEE Trans. Knowl. Data Eng.* **2021**, *35*, 4137–4150. [CrossRef]
21. Wang, R.; Li, Z.; Cao, J.; Chen, T.; Wang, L. Convolutional recurrent neural networks for text classification. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–6.
22. Shaham, S.; Ding, M.; Liu, B.; Lin, Z.; Li, J. Transition-Entropy: A novel metric for privacy preservation in location-based services. In Proceedings of the IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Paris, France, 29 April–2 May 2019; pp. 1–6.
23. Guo, Z.; Zhang, Y.; Teng, Z.; Lu, W. Densely connected graph convolutional networks for graph-to-sequence learning. *Trans. Assoc. Comput. Linguist.* **2019**, *7*, 297–312. [CrossRef]
24. Biega, A.J.; Saha Roy, R.; Weikum, G. Privacy through solidarity: A user-utility-preserving framework to counter profiling. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Tokyo, Japan, 7–11 August 2017; pp. 675–684.
25. Bottou, L. Large-scale machine learning with stochastic gradient descent. In Proceedings of the COMPSTAT'2010: 19th International Conference on Computational Statistics, Paris, France, 22–27 August 2010; Keynote, Invited and Contributed Papers; Springer: Berlin/Heidelberg, Germany, 2010; pp. 177–186.
26. Jevremović, A.; Jovanović, Z.; Jovanović, N. Complex networks analysis by spectral graph theory. In Proceedings of the Sinteza 2017-International Scientific Conference on Information Technology and Data Related Research, Belgrade, Serbia, 21 April 2017; Singidunum University: Belgrade, Serbia, 2017; pp. 182–185.
27. Feng, J.; Wu, Y.; Sun, H.; Zhang, S.; Liu, D. Panther: Practical Secure Two-Party Neural Network Inference. *IEEE Trans. Inf. Forensics Secur.* **2025**, *20*, 1149–1162. [CrossRef]
28. Fiesler, C.; Dye, M.; Feuston, J.L.; Hiruncharoenvate, C.; Hutto, C.J.; Morrison, S.; Khanipour Roshan, P.; Pavalanathan, U.; Bruckman, A.S.; De Choudhury, M.; et al. What (or who) is public? Privacy settings and social media content sharing. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, Portland, OR, USA, 25 February–1 March 2017; pp. 567–580.
29. Ganguly, D.; Roy, D.; Mitra, M.; Jones, G.J. Word embedding based generalized language model for information retrieval. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, 9–13 August 2015; pp. 795–798.
30. Guan, X.; Cheng, Z.; He, X.; Zhang, Y.; Zhu, Z.; Peng, Q.; Chua, T.S. Attentive aspect modeling for review-aware recommendation. *ACM Trans. Inf. Syst.* **2019**, *37*, 1–27. [CrossRef]
31. Giles, C.L.; Kuhn, G.M.; Williams, R.J. Dynamic recurrent neural networks: Theory and applications. *IEEE Trans. Neural Netw.* **1994**, *5*, 153–156. [CrossRef]
32. Su, F.; Li, P.; He, X.; Elumalai, V. Set pair analysis in earth and environmental sciences: Development, challenges, and future prospects. *Expo. Health* **2020**, *12*, 343–354. [CrossRef]
33. Han, R.; Tong, L.; Tong, W.; Yu, J. Research on vulnerability assessment of human-land system of Anshan city based on set pair analysis. *Prog. Geogr.* **2012**, *31*, 344–351.
34. Li, J.; Shan, H.; Pan, Y.; Wen, J.; Wu, D. Urban contaminated sites risk classification based on the analytic hierarchy process and set pair analysis. *Environ. Eng.* **2013**, *31*, 89–94 .
35. Song, X.; Wang, X.; Nie, L.; He, X.; Chen, Z.; Liu, W. A personal privacy preserving framework: I let you know who can see what. In Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, Ann Arbor, MI, USA, 8–12 July 2018; pp. 295–304.
36. Huang, W.Q.; Xia, J.F.; Yu, M.; Liu, C. Personal privacy metric based on public social network data. *J. Phys.* **2018**, *1087*, 032007. [CrossRef]
37. Dutta, A.; Gupta, A.; Zissermann, A. VGG Image Annotator (VIA). 2016. Available online: http://www.robots.ox.ac.uk/~vgg/software/via/ (accessed on 13 January 2025).
38. Zhang, N.; Paluri, M.; Taigman, Y.; Fergus, R.; Bourdev, L. Beyond frontal faces: Improving person recognition using multiple cues. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 4804–4813.

39. Korshunov, P.; Ebrahimi, T. PEViD: Privacy evaluation video dataset. In *Applications of Digital Image Processing XXXVI*; SPIE: Bellingham, WA, USA, 2013; Volume 8856, pp. 578–586.
40. Spyromitros-Xioufis, E.; Papadopoulos, S.; Popescu, A.; Kompatsiaris, Y. Personalized privacy-aware image classification. In Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval, New York, NY, USA, 6–9 June 2016; pp. 71–78.
41. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
42. Chen, Z.M.; Wei, X.S.; Wang, P.; Guo, Y. Multi-label image recognition with graph convolutional networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 5177–5186.
43. Chen, T.; Xu, M.; Hui, X.; Wu, H.; Lin, L. Learning semantic-specific graph representation for multi-label image recognition. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 522–531.
44. Guo, H.; Zheng, K.; Fan, X.; Yu, H.; Wang, S. Visual attention consistency under image transforms for multi-label image classification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 729–739.