CrowdBuy: Privacy-friendly Image Dataset Purchasing via Crowdsourcing

Lan Zhang*, Yannan Li*, Xiang Xiao*, Xiang-Yang Li*, Junjun Wang*, Anxin Zhou*, Qiang Li*
*University of Science and Technology of China

***Abstract*—In recent years, advanced machine learning techniques have demonstrated remarkable achievements in many areas. Despite the great success, one of the bottlenecks in applying machine learning techniques in real world applications lies in the lack of a large amount of high-quality training data from diverse domains. Meanwhile, massive personal data is being generated by mobile devices and is often underutilized. To bridge the gap, we propose a general dataset purchasing framework, named CROWDBUY and CROWDBUY++, based on crowdsourcing, with which a buyer can efficiently buy desired data from available mobile users with quality guarantee in a way respecting users' data ownership and privacy. We present a complete set of tools including privacy-preserving image dataset quality measurements and image selection mechanisms, which are budget feasible, truthful and highly efficient for mobile users. We conducted extensive evaluations of our framework on large-scale images and demonstrate that the system is capable of crowdsourcing high quality datasets while preserving image privacy with little computation and communication overhead.**

I. INTRODUCTION

Recently years, advanced machine learning techniques have demonstrated remarkable achievements in many areas, especially in recognizing and classifying images. Most state-of-the-art learning systems heavily rely on large amounts of labelled training data. The quality, quantity and diversity of training data have a significant impact on the accuracy and generalization capability of trained models. Data collection has mainly relied on web crawlers and data annotation has mainly been done by experts (e.g., doctors for medical images) or hired workers. But such methods have their limitations. First, we have little control on the type of images we could collect from the web for free. Many websites limit the access of web crawlers or put watermarks on the datasets. Images about rare objects or on sensitive subjects are often hard to obtain, for example, biometric and medical images. There are also some issues on ownership and privacy involved. The urgent need of massive high-quality image datasets from diverse backgrounds becomes a bottleneck for many real world applications.

Meanwhile, there are nearly unlimited data in personal devices. A large number of images of diverse categories are being generated by mobile devices every day. According to InfoTrends's report, there are 3.9 trillion images generated in 2016. Most of these images are underutilized or may never been used. We plan to leverage them to satisfy the urgent dataset needs from machine learning by using crowdsourcing.

Crowdsourcing is a paradigm for utilizing the power of a crowd of devices. It has been used to exploit the embedded sensors on mobile devices for sensing and computation [1], [32], termed as participatory sensing [9]. Most of these systems focused on simple structured data such as noise or illumination level, location or mobility data, for applications such as noise monitoring [1] and traffic analysis [2]. Only a few designs are about crowdsourcing based image collection. They either collect data without quality guarantee [23] [9], or require subjective image quality assessment [14], [22]. To provide objective quality measurement, Cao et al. [20], [21] propose to leverage images' metadata and mobile phones' sensor data to collect images of a target object or target area. Those work didn't provide an efficient way to evaluate the quality of image content, especially for large-scale images for deep learning applications. Further, although there are many efforts devoted to privacy-preserving data aggregation, they all focus on numerical data [11], [19], [32]. Few studies have considered image privacy and ownership at all. Due to the complex structures in images, the problem of how to collect a large image dataset with content quality guarantee and considerations of ownership/privacy is still not solved.

In this paper we take a first step towards image dataset purchasing via crowdsourcing to bridge the urgent need of massive high-quality image datasets in diverse categories and the unlimited underutilized images in personal devices. We propose a framework CROWDBUY. A buyer can issue a request describing his/her desired image content and monetary budget. Potential sellers can check their local albums and upload information about matched images. Then image selections are conducted by our system to maximize dataset quality within the budget limit. In the end, the selected sellers get their payments and the buyer gets the corresponding images.

We need to address the following critical challenges.

1. Image ownership and privacy protection. An image can be easily copied or disseminated, and may have sensitive information, e.g., in medical settings. During the whole purchase, the buyer should have the right to access an image only after he/she paid for it. The server or any other third parties can neither access any images for resale, nor know their content. Given these restrictions, it is extremely hard for sellers to prove that they do possess images that match the specification, without revealing the original images. It is also challenging for the server to select a high quality dataset without knowing the images directly.

2. Image selection with quality guarantee and budget restriction: Given a limited budget, our framework should maximize the buyer's profit by selecting a high-quality dataset from the image pool. This is a difficult task, not just because

978-1-5386-4128-6/18/$31.00 ©2018 IEEE

the server cannot know images' content, but also malicious users may report false information about their images or misreport prices to obtain higher compensation. Besides, it's hard to answer what makes high-quality datasets for applications like machine learning. So we need proper quality metrics and a truthful[1] and budget feasible[2] selection mechanism.

3. System efficiency: Our framework should be efficient for resource limited mobile users. Expensive encryption based privacy-preserving methods are not applicable in our setting.

Methodology and contributions. In this work, we propose, design and implement CROWDBUY to address aforementioned challenges. To the best of our knowledge, CROWDBUY is the first design implementing a crowdsourcing based image dataset purchasing platform with quality guarantee, meanwhile respecting sellers' data ownership and privacy. CROWDBUY opens a new window to collect high-quality image datasets for tasks like machine learning. In CROWDBUY, we design a novel image selection method based on two-dimension features of images, which are extracted from a pre-trained CNN model and an autoencoder by the sellers. The cloud, using only the two-dimension image features provided by the sellers, can exam if an image meets requirements, and select datasets with maximum quality under budgets. We design a set of truthful and budget feasible image selection mechanisms to maximize three different image dataset quality requirements, i.e., quantity, similarity and diversity. We also propose a method to measure image dataset diversity in the feature space. Our selection method resists up to 50% noisy images from malicious users.

For CROWDBUY, we analyze, in a comprehensive manner, the privacy issues during the whole transaction. We propose two privacy levels to meet different user requirements. For the first level, we protect the original images by transforming them into two-dimension feature vectors. To further protect the feature vectors, we propose a concept named *feature-indistinguishability*, which is a generalization of differential privacy, and design CROWDBUY++ with a light-weight Min-Hash based technique to satisfy ϵr-feature-indistinguishability. Besides, our system also provides a verification mechanism to resist dishonest behaviors by the buyer and sellers. We conducted extensive evaluations of our systems on 222,300 images. The results prove the effectiveness of our system, and show that our mechanism is capable of selecting high quality datasets to train high-accuracy deep learning models.

The rest of this paper is organized as follows. In Section II, we describe the problem and design objectives. We introduce our basic system outline in Section III and selection mechanisms in Section IV. To further improve the privacy protection, we present an advanced system in Section V. The experimental results are reported in Section VI. We review related work in Section VII and conclude the work in Section VIII.

[1]A mechanism is truthful if no users can improve its utility by submitting false prices, no matter what others submit.

[2]A budget feasible mechanism requires the payments it makes to agents do not exceed the budget.

II. PROBLEM DESCRIPTION

A. Motivation

To start a purchase, the buyer uploads image requirements and the budget to the cloud. The cloud can broadcast the requirements to a large number of mobile users. Users who have matched images can upload certain description/encryption of those images to the cloud. Then, due to the budget limitation, the cloud selects which images from which sellers to buy for the buyer. After the buyer pays those selected sellers, the sellers send corresponding original images to the buyer. To make this system practical, it is crucial to maximize the buyers benefit as well as protect every sellers data ownership and privacy. The cloud should select images based on the uploaded description/encryption information to produce a high-quality dataset with the budget limit. To motivate the cloud, it can also be paid for every successful trade. During the whole process, only the buyer should have the right to access an image after he/she paid for it, while the cloud or any other third parties can neither access any images for sale, nor know the content of those images. Besides, the overhead of mobile devices should also be minimized. Simultaneously achieving these goals raises great challenges for our system design.

B. Privacy Threats

In our system, we treat images as commodities. Different from traditional commodities, images are easily duplicated and may reveal sensitive information about the seller, e.g, age, income, location and heath condition. In this work, we consider two types of threats to images content, which expose the images information to different extends.

•**Original image privacy.** The most intuitive way to violate a sellers image privacy is to expose the original image itself. So, the original image cannot be transmitted to any party except the buyer who has paid for it.

•**Image features privacy.** Since the cloud needs to select matched and high-quality images for the buyer, some features about to-be-sold images must be provided by sellers. Although the sellers do not upload the original images directly, features extracted from images still contain a lot of information about images' content [5] [15]. Thus, uploaded image features should also be protected from the cloud.

Besides, we also consider the *Identity privacy* of sellers, which means the buyer should not known the identities of those image sellers. Here we do not consider protect sellers' identities from the cloud, since those sellers need to be paid through the cloud [3]. But the cloud doesn't know any content information of seller's images. Note that, we do not consider protecting image content from those buyers who have paid for them, since the sellers decide to sell the image of his/her own free will. We need to carefully design our system to achieve these privacy requirements.

C. Adversary Model

In our system, there are three parties involved, the buyer, the cloud server and sellers. For the cloud, we apply the

[3]The sellers cannot be paid directly by the buyer, because the buyer cannot know their identities

widely adopted semi-honest (honest but curious) assumption, that is, the cloud will follow the protocol but try to get more extra information about the image content. All mobile users can be malicious. During the selection stage, a seller may upload information of arbitrary unmatched images, or even upload forged information to pretend he/she has a matched image. For the image delivery stage, a seller may deliver an image which is inconsistent with the uploaded information during the selection stage. The buyer has the right to verify the consistency of selected images and delivered images, but he/she may lie about the verification results to refuse to pay for valid images. We assume that the cloud will not collude with any user. Thus, besides privacy protection we also need to provide verification mechanisms to resist dishonest users.

D. Design Goals

We design our system to achieve the following three goals:

1) Quality: To select high-quality dataset, first, the selected images should satisfy the buyers target requirements. Then on this basis, our system should select matched images to maximize the quality metric. In our work, we consider three quality metrics, quantity, similarity (to some sample images) and diversity. We propose diversity as a reasonable quality criterion for an image dataset, because intuitively a buyer will not buy a dataset with thousands of very similar images; theoretically, a diverse dataset can train a machine learning model with better generalization capability.

2) Privacy: For threats defined in Section II-B, we introduce two privacy levels. (1) **Level-1 privacy** protects both original image privacy and identity privacy, where the original image can only be transmitted to the buyer who has paid, while image features can only be exposed to the cloud. (2) **Level-2 privacy** protects original image privacy, image feature privacy and identity privacy, where the original image can only be transmitted to the buyer who has paid, while image features cannot be exposed to any parties. For both levels, the sellers identities can only be revealed to the cloud. We design a system CROWDBUY to achieve Level-1 privacy, and an advanced system CROWDBUY++ to achieve Level-2 privacy.

3) Efficiency: Facing massive image data and restricted resources of mobile devices, our system should be efficient for resource limited mobile users in computation and communication overhead. We do not adopt some computation intensive encryption mechanisms, e.g., homomorphic encryptions.

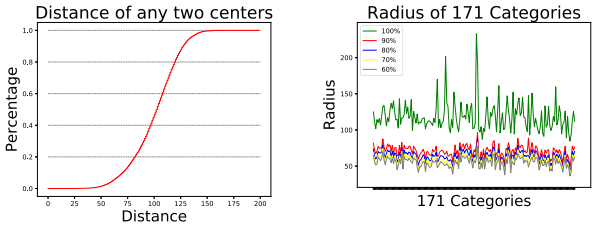
III. CROWDBUY BASIC SYSTEM DESIGN

In this section, we firstly present our basic system CROWDBUY preserving Level-1 privacy.

A. Observation

To protect original images, the image selection cannot be conducted on original images directly. We propose to map images into a feature space and conduct the selection in this feature space. To achieve a reasonable selection criterion, we investigate the feature space in large-scale image dataset.

Feature vector generation: A sophisticated model is required to map diverse images into the same feature space. Simonyan and Zisserman model [16] (VGG) is a well-known deep convolutional networks which achieves significant accuracy for large-scale image classification. We propose to feed an image through a pre-trained VGG-16 network, and extract the eighth (last) fully-connected layer (fc8) as its feature vector, which provides a good abstract of the image content.



(a) Distances between two category centroids. (b) Recall with different radius.

Fig. 1: Distances among different image categories in the feature space for the ImageNet dataset.

Feature space characteristics: To characterize the feature space, we use the well known large-scale image dataset, ImageNet [4]. We generate feature vectors for all images for 171 categories (1300 images for each category), compute a centroid vector for each category, and measure the distances among those vectors. Fig. 1a shows that, in 90% cases, two categories have a distance larger than 75. Fig. 1b shows the recall given a distance to the centroid for each category. Almost for all 171 categories, more than 90% images can be recalled within 75 distance. This results reveal that, in the fc8 feature space, we can effectively distinguish images of different categories of content.

These observations prove the rationality of our proposal to select adequate images in this feature space. Given a requirement by a buyer, e.g., a sample image, in the feature space, we can select images within a reasonable range, e.g, 75, around the sample image.

B. System Design

Now, we present the overview of our Level-1 privacy preserving system CROWDBUY. As shown in Figure 2, CROWDBUY consists of three parties: the buyer(s), sellers and cloud. There are two stages for image purchase, *selection stage* and *delivery stage*. In the selection stage, sellers upload features of images for selection. In the delivery stage selected sellers transmit encrypted original images to the buyer after get paid.

(1) **Request**: the buyer sends a request $(\mathbb{S}, B, Q, Pk)$ to the server. $\mathbb{S}$ is the collection of sample images, e.g., an image of dog or several images of different kinds of dogs, which means that the desired images should be in the same category as at least one of the sample images. B is the budget for dataset purchase, which requires our selection system to be budget feasible. Q is the preferred quality metric (we will introduce in Section IV-B). Pk is the buyer's public key to encrypt the selected images.

(2) **Broadcasting requirements:** The cloud broadcasts the request, *i.e.*, sample images $\mathbb{S}$, to all potential sellers.

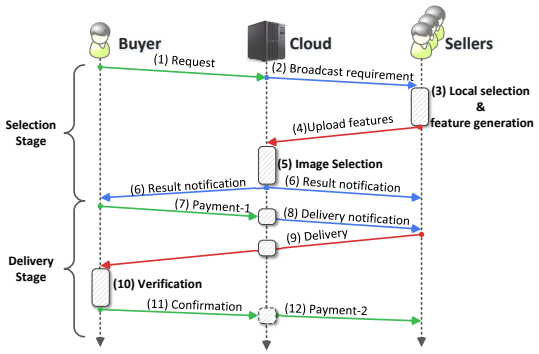


Fig. 2: Design overview of CROWDBUY.

(3) **Local selection & feature generation:** Given $\mathbb{S}$, each seller checks his/her local album to see if there is any image satisfying the requirement. Then the fc8 feature vector F of this image is extracted locally.

(4) **Uploading features:** The seller uploads a bid (F, c) to the cloud, where F is the fc8 feature of his/her image, and c is the declared price of the image which is not necessarily equal to the true price $\hat{c}$. $\hat{c}$ is private information to the seller. The seller may misreport the price to obtain higher compensation, so our mechanism should be truthful to incent all participants to report true prices.

(5) **Image Selection:** Given all feature vectors from different sellers $\mathbb{F} = \{F_1, F_2, \cdots, F_n\}$, the cloud firstly selects images satisfying the requirement by calculating their distances to the sample images, and forms the candidate feature vector set $\mathbb{F}_c$. Then based on the quality requirement Q, the cloud selects images from $\mathbb{F}_c$ to maximize the dataset quality within the budget limitation B. Based on the selected feature vector set $\mathbb{F}_s$, the cloud calculates the total compensation P for this dataset. We will present the detailed image selection mechanism in Section IV

(6) **Result notification:** The cloud then notifies each selected seller which images have been selected. It also notifies the buyer about the total money P he/she needs to pay.

(7) **Payment-1:** The buyer pays the total compensation P to the cloud under certain contract.

(8) **Delivery notification:** After getting paid from the buyer, the cloud sends Pk to all selected sellers and notifies them to deliver the selected images.

(9) **Delivery:** Each selected seller encrypts the corresponding original image by the buyer's public key Pk and uploads the ciphertext to the cloud. The cloud generates a hash value for each image's ciphertext and stores all hash values for potential verification in the future. Then it packs all ciphertext and sends them to the buyer, along with the selected feature vector set $\mathbb{F}_s$.

(10) **Verification (Optional):** The buyer decrypts all images with his/her private key Sk. In case of dishonest sellers, he/she can generate feature vectors for all images or a sampled subset of images. Let the feature vector set be $\mathbb{F}_v$. Then he/she compares $\mathbb{F}_s$ and $\mathbb{F}_v$ to verify if delivered images meet those uploaded feature vectors during the selection phase.

(11) **Confirmation:** If $\mathbb{F}_s$ is consistent with $\mathbb{F}_v$, then the buyer confirms the success of the purchase to the cloud. If there is an image whose feature in $\mathbb{F}_v$ is inconsistent with that in $\mathbb{F}_s$, then the buyer reports the image id to the cloud as well as send the image to the cloud. Receiving the report, in case the buyer is dishonest, the cloud first compares the hash value of uploaded image's ciphertext to the stored one, then generates its feature vector to compare it with that stored in $\mathbb{F}_s$. Only if the image passes the first comparison and fails the second one, the seller is dishonest and will be punished, meanwhile the payment for this image will be refunded. In other cases, the buyer is lying and will get punished.

(12) **Payment-2:** After verification by the buyer and confirmation by the cloud, the cloud pays money to honest sellers and refunds payment for dishonest sellers to the buyer.

C. Privacy Analysis

In CROWDBUY, only paid original images will be exposed to the buyer. No other parties can obtain any original image. All sellers' identities are only revealed to the cloud for the confirmation and payment. Specially, during the confirmation stage, the buyer may upload original images to the cloud to report inconsistence. This doesn't violate original image privacy, since if the buyer is lying, then the uploaded images are not original images provided by the sellers; if the buyer is honest, then these images are some invalid images uploaded by malicious sellers deliberately. The only revealed information about images are their feature vectors. According to [5], a rough contour can be reconstructed from the 1000-dimension fc8 feature vectors. In CROWDBUY, we reduce the feature vector to 2-dimension to greatly improve the efficiency as well as reduce the privacy leakage (see Section IV-D). We propose a strategy to further protect the 2-dimension vectors to preserve Level-2 privacy (see Section V).

IV. IMAGE SELECTION

In this section, given a buyer's request $(\mathbb{S}, B, Q, Pk)$ we design a set of truthful and budget feasible image selection mechanisms to maximize different image dataset quality. The selection is conducted on uploaded feature vectors of images $\mathbb{F} = \{F_1, F_2, \cdots, F_n\}$ and consists of two stages. First the cloud selects features satisfying the requirement to form a candidate feature vector set $\mathbb{F}_c$, and then selects features from $\mathbb{F}_c$ to build an optimal dataset. For the rest of this paper, we use "image" and "its feature" interchangeably.

A. Matched Image Decision

As we mentioned in Section III-A, we can measure images' similarity by the distance between their feature vectors. Here we define feature vector distance as $\mathcal{D}(F_i, F_j) = \sqrt{\sum_{k=0}^{D-1}(F_{ik} - F_{jk})^2}$. Let the feature vectors of sample images $\mathbb{S} = \{S_1, S_2, \cdots, S_m\}$ be $\mathbb{O} = \{O_1, O_2, \cdots, O_m\}$, and treat them as target centroids. R is the distance threshold, which indicates that if the distance between a feature F_i and any centroid O_j is larger than R, then this feature doesn't meet the requirement and the image is unmatched. Specially,

if the distance equal 0, the image is also unwanted, since the buyer won't pay for some images he/she already has. Formally, a feature is unmatched if $\min\{\mathcal{D}(F_i, O_j)\} > R$ or $\min\{\mathcal{D}(F_i, O_j)\} = 0$ for $1 \le j \le m$, and the candidate set is $\mathbb{F}_c = \{F_i | 0 < \min\{\mathcal{D}(F_i, O_j)\} \le R, 1 \le j \le m\}$.

B. Dataset Quality Metrics

We consider three reasonable dataset quality requirements:

1) Quantity metric: For the quantity metric, the buyer wants to buy as many images as possible within his/her budget. The quantity of the selected image dataset is $|\mathbb{F}_s|$.

2) Matching degree metric: For the matching degree metric, the buyer wants to buy images which are most similar to the sample images, e.g., different angles of the same person's face.

3) Diversity metric: For the diversity metric, the buyer wants to buy images as different as possible, provided they are all matched images, e.g., instead of images of one person's face, the buyer may prefer images of different people' faces to train a better face recognition model. This is an important metric especially for machine learning, because a diverse dataset can help the model to mitigate overfitting and improve the generalization capability.

For a feature F_i, we can define its matching degree according to the comparison between $\mathcal{D}(F_i, O_j)$ and R, where 1 is perfectly matched while 0 means unmatched.

Definition 1 (Matching Degree) *Given target centroids $\mathbb{O} = \{O_1, \cdots, O_m\}$, the matching degree of F_i on $\mathbb{O}$ is*

$$e_{F_i|\mathbb{O}} = \begin{cases} 1 - \frac{\min_{1\le j\le m}\{\mathcal{D}(F_i,O_j)\}}{R}, & \frac{\min_{1\le j\le m}\{\mathcal{D}(F_i,O_j)\}}{R} \le 1 \\ 0, & \frac{\min_{1\le j\le m}\{\mathcal{D}(F_i,O_j)\}}{R} \ge 1 \end{cases}$$

So, for a selected set of images $\mathbb{F}_s$, its matching degree is defined as the total matching degree of all images in this set, which is defined as: $e_{\mathbb{F}_s|\mathbb{O}} = \sum_{F_i \in \mathbb{F}_s} e_{F_i|\mathbb{O}}$.

To maximize the content diversity of a selected image dataset within the budget limit, we should select images within R as dispersed as possible. Here we model the diversity quantification problem as a coverage problem, by giving each feature vector a convex coverage range around it. Specifically, we model the coverage range as a hypercube with $2d$-length sides, whose center is F_i and every side is parallel to one axis. As a result, the diversity of an image set can be measured by the total volume of the union of all hypercubes. Formally, we define image *diversity utility* as the volume of its feature's hypercube $v(F_i)$.

Definition 2 (Diversity Utility) *Given target centroids $\mathbb{O} = \{O_1, \cdots, O_m\}$, the diversity utility of S_i on $\mathbb{O}$ is defined as,*

$$U_{F_i|\mathbb{O}} = \begin{cases} v(F_i) & e_{F_i|\mathbb{O}} > 0 \\ 0 & e_{F_i|\mathbb{O}} = 0 \end{cases}$$

Accordingly, the utility of a selected set of images $\mathbb{F}_s$ on $\mathbb{O}$ is the union volume of all hypercubes in this set, which is

$U_{\mathbb{F}_s|\mathbb{O}} = \bigcup_i v(F_i), F_i \in \mathbb{F}_s$ *and* $e_{F_i|\mathbb{O}} > 0$.

Figure 3 gives an example in 2D features space. By Definition 2, the utility of F_4 is 0. For a single matched image, its utility is the 2D volume $4d^2$. For the image set $\{S_1, S_2\}$, its utility is less than $8d^2$, since they have overlap.

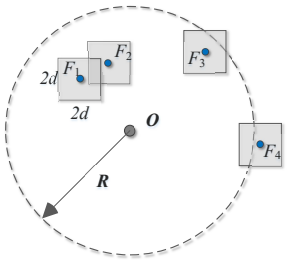


Fig. 3: Diversity-driven feature selection model.

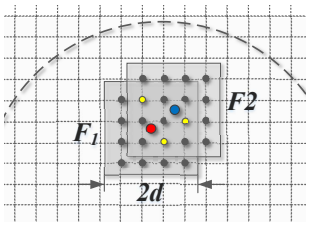


Fig. 4: Example of the feature indistinguishability model.

C. Quality-driven Image Selection and Payment

Driven by different dataset quality metrics in Section IV-B, we design different image selection and payment mechanisms to satisfy the budget feasible and truthful requirements, to maximize the dataset value, i.e.,

$$\textbf{Objective: } Maximize\ V(\mathbb{F}_s) \quad \textbf{Subject to: } \sum_{F_i \in \mathbb{F}_s} p_i \le B.$$

Here, $V(\mathbb{F}_s)$ is the value of the selected image set $\mathbb{F}_s$ given the target $\mathbb{O}$, which depends on specific image quality metric. p_i is the payment for image S_i.

1. Quantity driven selection: For the quantity dataset metric, the value function is $V_q(\mathbb{F}_s) = |\mathbb{F}_s|$.

If there are n uploaded features, each feature F_i has a declared price c_i. The optimal solution to maximize this quantity value function is by greedily choosing the lowest-priced images until the budget is exhausted. Based on [17], the truthful and budget feasible mechanism works as follows: we sort the n bids to make $c_1 \le c_2 \le \cdots \le c_n$, and let k be the largest index for which $c_k \le B/k$. Then the selected set is the first k lowest-priced images, and the payment for each seller is $\min\{B/k, c_{k+1}\}$.

Theorem 1 *We have a truthful budget-feasible crowdsourcing mechanism for maximizing the number of selected images that is two-approximation for the achieved utility.*

2. Matching degree driven selection: The matching degree value function is $V_m(\mathbb{F}_s) = e_{\mathbb{F}_s|\mathbb{O}} = \sum_{F_i \in \mathbb{F}_s} e_{F_i|\mathbb{O}}$. Given a set of selected images $\mathbb{F}_s$, the marginal value of an image $F_i \notin \mathbb{F}_s$ is $V_{\mathbb{F}_s}(F_i) = V(\mathbb{F}_s \cup \{F_i\}) - V(\mathbb{F}_s)$, which is $e_{\mathbb{F}_i|\mathbb{O}}$.

Lemma 2 *The matching degree value function V_m is submodular and non-decreasing.*

Proof The proof is omitted due to space limitation. ∎

3. Diversity driven selection: Based on the image diversity utility definition, the diversity value function is:

$$V_d(\mathbb{F}_s) = U_{\mathbb{F}_s|\mathbb{O}} = \bigcup_i v(F_i), F_i \in \mathbb{F}_s \text{ and } e_{F_i|\mathbb{O}} > 0$$

Given a set of selected images $\mathbb{F}_s$, the marginal value of image $F_i \notin \mathbb{F}_s$, *i.e.*, $V_{\mathbb{F}_s}(F_i) = V(\mathbb{F}_s \cup \{F_i\}) - V(\mathbb{F}_s)$, is $v(F_i) - v(F_i) \cap v(\mathbb{F}_s)$ if $e_{F_i|\mathbb{O}} > 0$.

Lemma 3 *The diversity value function V_d is submodular and non-decreasing.*

Proof The proof is omitted due to space limitation. ∎

Mechanism for matching degree metric and diversity metric. Since both matching degree value function V_m and diversity value function V_d are submodular and non-decreasing, we can leverage the state-of-the-art result, e.g., BEACON [31], to achieve truthful, budget feasible, and individual rational [4] mechanisms. More specifically, we find the image with the largest value as a candidate, compare its value with that of the greedy selected set (which iteratively selects the image with the maximum marginal contribution per cost $V_{\mathbb{F}_{i-1}}(F_i)/c_i$ until currently considered image F_i violates budget feasible allocation condition), then select the larger one. After the selection, as the second stage of the Myerson Lemma, we can find the critical payments for selected sellers.

Theorem 4 *Our mechanism is truthful budget-feasible for maximizing the matching degree or the diversity of selected images that is at least $\frac{1-e}{5e}$ of the optimum utility.*

D. Further Cost Reduction and Privacy Improvement

Using the 1000 fc8 feature vector achieves high accuracy while may cause large computation and communication cost when facing large-scale images. Especially, the diversity driven selection needs to solve expensive intersection reporting problem in D-dimension feature space. We optimize the efficiency by significantly reducing the 1000-dimension space to a 2-dimension space. Then we can use optimal solution for the diversity driven selection in 2-dimension space, where $O(n \log n + k)$ time and $O(n)$ space suffice to report the k intersecting pairs. To achieve this goal, we propose to use a pre-trained autoencoder [3] to learn the 2-dimension vectors of fc8 feature vectors. Here we use about 200,000 images of different categories from the ImageNet dataset to train the autoencoder, which includes three fully connected layers and two softplus for the encoder. We also investigate image distance distribution in the 2-dimension space, which is similar to Fig. 1, only with a different distance scale. So, we can still set a reasonable threshold R, e.g, 4000, in the 2-dimension feature space, to achieve high precision and recall. By reducing feature vectors to 2-dimension, we greatly reduce computation and communication cost as well as privacy leakage.

V. Advanced System Model for Better Privacy

In the basic system design, CrowdBuy protects original images while only reveals 2-dimension feature vectors of images. However, the autoencoder is trained in an encoder-decoder manner, with the help of the decoder an adversary can reconstruct the 1000-dimension feature with some loss. So, in this section, we further improve our system to CrowdBuy++ to provide Level-2 privacy, i.e., protecting the 2-dimension vectors. Note that, CrowdBuy++ only modifies the image selection phase while other steps remain the same as CrowdBuy. To minimize the cloud and users' cost, we do not adopt homomorphic encryption based methods, while we design a MinHash based strategy to achieve feature-indistinguishability.

A. Feature-indistinguishability

Intuitively, to protect each feature vector, we want a feature vector indistinguishable with other neighboring feature vectors within a range r. Towards this, we propose the notion *l-feature-indistinguishability* which is a generalization of differential privacy. Here l indicates the level of privacy for the range r. And the smaller l is, the higher the privacy. We require $1/l$ be proportional to r to achieve useful quality metric.

Definition 3 (Feature-indistinguishability) *A mechanism satisfies l-feature-indistinguishability iff for any range $r > 0$, the feature vector preserves ϵr-privacy within r.*

To satisfy l-feature-indistinguishability, we propose a MinHash based mechanism to map a feature vector's neighboring range to k MinHash values in this range. And each seller uploads the set of k hash values instead of the feature vector to the cloud. The cloud conducts selection on the hash value sets to maximize dataset quality for different metrics, and the selection mechanisms are still truthful and budget feasible.

B. Feature Vector Hash

Definition 4 (MinHash $h_{(k)}(\mathcal{T})$) *Given a set $\mathcal{T}$, let $h \rightarrow [0,1]$ be a hash function that maps members of $\mathcal{T}$ to distinct numbers drawn uniformly at random from interval $[0,1]$. For any set $T \subset \mathcal{T}, |T| \geq k$, define the value $h_{(k)}(T)$ to be the subset of the k members of T with the smallest values of h.*

To hash a feature vector F_i, we first map F_i's neighboring range within a distance d into a neighboring set $\mathbb{F}_i^d$. For computation efficiency, the neighboring range is defined the same as the coverage range in CrowdBuy. Then $\mathbb{F}_i^d$ is the set of integer points within F_i's neighboring range. Fig. 4 gives an example of neighboring ranges and neighboring sets of F_1 and F_2 in 2D space. Given a feature vector F_i, we apply MinHash on its neighboring set $\mathbb{F}_i^d$ to generate its MinHash signature $h_{(k)}(\mathbb{F}_i^d)$, which is a set of k integer points in $\mathbb{F}_i^d$ producing smallest hash values. For example, in Fig. 4, the yellow points are the MinHash signatures of both $\mathbb{F}_1^d$ and $\mathbb{F}_2^d$.

C. Quality-driven Image Selection and Payment Mechanism

Each seller uploads the MinHash signature $h_{(k)}(\mathbb{F}_i^d)$ of his/her image feature F_i to the cloud. Then cloud conducts selection on all MinHash results.

Matched image decision: For a MinHash signature $h_{(k)}(\mathbb{F}_i^d)$, let the centroid of set $\mathbb{F}_i^d$ be $O(\mathbb{F}_i^d)$. Compared with CrowdBuy, here the distance $\mathcal{D}(F_i, O_j)$ can be approximated by $\mathcal{D}(O(\mathbb{F}_i^d), O_j)$.

Quantity driven selection: Uploading MinHash signatures doesn't cause any change to the value function (image number), so the selection mechanism is the same as that of the basic system CrowdBuy.

[4]Individual rationality means each participating user will have a non-negative utility.

Matching degree driven selection: In CROWDBUY++, to measure the matching degree of each feature F_i, we use the centroid $O(\mathbb{F}_i^d)$ to represent F_i.

Diversity driven selection: Since the cloud only has k integer points for each F_i, it cannot calculate the accurate overlapped volume of two feature vectors. As a result, we redefine the coverage of a range as the number of distinct integer points covered by its MinHash signature. Then, the utility of an image F_i is $v(F_i) = |h_{(k)}(\mathbb{F}_i^d)|$, and the diversity value function is

$$V_d(\mathbb{F}_s) = |\bigcup_i h_{(k)}(\mathbb{F}_i^d)|, F_i \in \mathbb{F}_s \text{ and } e_{F_i|\mathbb{O}} > 0.$$

We can prove that, the matching degree value function V_m and diversity value function V_d are still monotone submodular functions. So, the selection and payment mechanisms for them are the same as that of the basic system CROWDBUY.

In CROWDBUY++, for the verification and confirmation, we require each user to upload the hash value of the feature vector along with the MinHash signature.

D. Privacy Analysis

Theorem 5 *For two sets $\mathbb{F}_i^d$ and $\mathbb{F}_j^d$ where $\mathbb{F}_i^d$ and $\mathbb{F}_j^d$ are r-step away, that is they have r different rows or r different columns, the probability of the output of the MinHash signatures $h_{(k)}(\mathbb{F}_i^d)$ and $h_{(k)}(\mathbb{F}_j^d)$ satisfies:*

$$Pr\{h_{(k)}(\mathbb{F}_i^d) = H\} \leq e^{\epsilon r} \times Pr\{h_{(k)}(\mathbb{F}_j^d) = H\},$$

$$\epsilon r = \frac{2dk}{4d^2 - k + 1} r$$

Proof The proof is omitted due to space limitation. ∎

So, when $r = 1$, $\mathbb{F}_i^d$ and $\mathbb{F}_j^d$ are neighboring data set, and our mechanism satisfies ϵ-privacy for their feature vectors. And according to Definition 3, our mechanism also satisfies l-feature indistinguishability. Requiring ϵr-privacy forces the MinHash signature to be similar to features close to each other, while relaxing the constraint for those far away from each other, allowing the cloud to select images maximizing the matching degree and diversity. To achieve better privacy (smaller l), the system should use smaller k and larger d. Same as CROWDBUY, during the whole transaction, the original images are only revealed to the buyer who has paid and all sellers' identities are only revealed to the cloud.

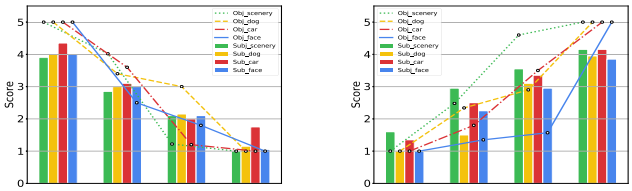
VI. EVALUATION

CROWDBUY preserves Level-1 privacy and CROWDBUY++ preserves Level-2 privacy, which both protect sellers' data ownership and preserve image content privacy to different extents. In this section, we evaluate selected dataset quality in different settings, as well as report the efficiency.

A. Experiment Setting and Parameter Decision

For all following experiments, we use 171 diverse categories (1000 images for each category) from dataset "ILSVRC2012_img_train", one of ImageNet datasets.

There are three parameters in CROWDBUY and CROWDBUY++, the distance threshold R, coverage/neighboring range d and MinHash parameter k. Based on analysis of ImageNet, on average 80% images of the same category are within the range of 4000 from their centroid, while about 90% categories



(a) Matching degree. (b) Diversity.

Fig. 5: Objective and subjective measurements of image dataset matching degree and diversity.

are more than 4000 away from each other. So, R is set to 4000 to achieve both high precision and reasonable recall, since for image purchase precision is more important. d is both coverage range of an image in CROWDBUY and neighboring range in CROWDBUY++. Bigger d and smaller k provide better privacy, which can be adjusted for different scenarios to meet different privacy requirements. In our experiments, the average distance for vectors within R is about 125. To provide sufficient protection to vectors and adequate selection accuracy, we set $k = 4$ and $d = 300$, then ϵ is about $\frac{1}{150}$.

B. Metric Validation

We propose to select images in the 2-dimension feature space to optimize dataset quality. Quantity can be measured accurately in any feature space. Before we present more evaluation results, we validate the matching degree and diversity metrics to answer two questions: (1) do two close images in the feature space have similar content? (2) does larger coverage in the feature space mean better diversity of image content?

First, our analysis of distances among images in ImageNet dataset shows that images with similar content have smaller distance than images with different content, as shown in Fig. 1. Then, we select four diverse categories, including human face, dog, car and scenery. Based on our metrics, for each category, we select images from it to build four datasets with different matching degree levels and four datasets with different diversity levels. We showed 32 image sets of four categories to 20 volunteers and asked them to rate the matching degree and diversity of each image set between $1 \sim 5$. 1 means worst matching degree and worst diversity. We also normalize our objective metrics to $[1, 5]$. Fig. 5 proves that for diverse categories of images, our matching degree and diversity metrics are in accord with subjective evaluation. The result confirms the feasibility to select high quality datasets in the 2-dimension feature space.

C. Dataset Quality

We first measure the precision of the candidate dataset selected by our matched image decision mechanism when there are dishonest sellers uploading irrelevant images. Then, we evaluate the quality of dataset selected by our three selection mechanisms. For Fig. 7, Fig. 8, Fig. 9 and Fig. 10, the bars and lines presents results of CROWDBUY++.

For the purchase, without losing generality, we assume the buyer needs images of dogs. We use images of 10 different categories of dogs as matched images (13,000 images in

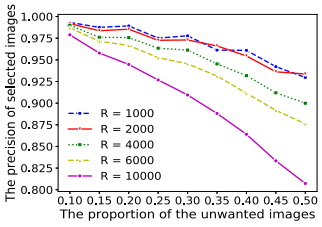


Fig. 6: Selected datasets precision against different noise level.

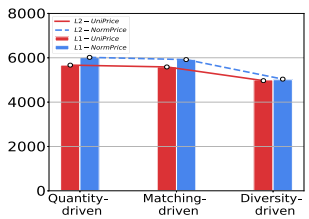


Fig. 7: Quantity of datasets under different price models.

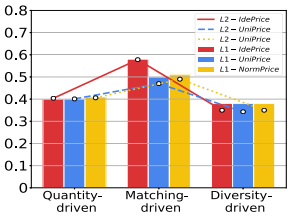


Fig. 8: Average matching degree under different price models.

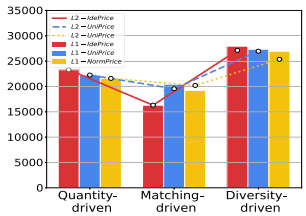


Fig. 9: Diversity of datasets under different price models.

total), including Labrador retriever, basset hound, Scottish deerhound, etc. For each purchase, the buyer's 10 sample images are randomly drawn from 10 categories respectively. Honest sellers' images are randomly drawn from the 10 categories, while dishonest sellers' images are randomly drawn from the other 161 categories(noise images). For the declared prices, we consider three commonly used models, identical price (denoted as IdePrice), uniformly distributed price (denoted as UniPrice) and normally distributed price (denoted as NormPrice). The mean values of both uniform distribution and normal distribution equal to the identical price, which is 1 in our experiment, and variances are 0.4. The budget is 5000.

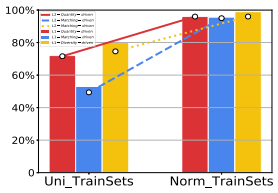
1) Dataset precision: We evaluate our selction precision against different noise levels. As plotted in Fig. 6, when the distance threshold R is less than 4000, as the noise percentage increases from 10% to 30%, the precision declines slightly from 99% to 95%. Even when there are 50% noise images, the precision is still above 90%, because with our mechanism, most unmatched images won't fall into the target range.

2) Quantity: With the same budget, for the IdePrice, the buyer can always buy 5000 images. Fig. 7 presents the dataset quantity using different quality driven mechanisms for UniPrice and NormPrice. Obviously, quantity driven mechanism purchases most images in all cases, which is 5669 for UniPrice and 6019 for NormPrice with both CROWDBUY and CROWDBUY++. The image quantity of diversity driven mechanism is the smallest, which are 5000 and 5011 for CROWDBUY, and 4969 and 5038 for CROWDBUY++.

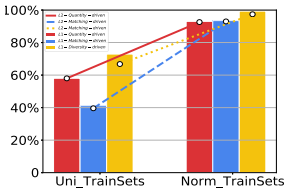
3) Matching degree: As presented in Fig. 8, our matching degree driven mechanism produces datasets most similar to the sample images in all cases. With CROWDBUY, for IdePrice, its average matching degree is 0.58, and for UniPrice and NormPrice, they are 0.5 and 0.51 respectively. The quantity driven mechanism has smaller matching degree, which is about 0.4 for all price models. The diversity driven mechanism has the smallest matching degree (about 0.37), since it needs to maximize image content diversity. CROWDBUY++ performs quite similarly with CROWDBUY. Though three mechanisms produce datasets with different matching degrees, the matching degrees are all above 0.34, which guarantees the lower bound of similarity between selected images and image samples.

D. Diversity and data utility

As present in Fig.9, in all cases, while our diversity driven mechanism achieves best diversity, the matching degree driven mechanism achieves worst diversity. Dataset selected by the quantity driven mechanism has a fair diversity due to the randomness of selected images in the feature space.



(a) Test known categories.



(b) Test unknown categories.

Fig. 10: Deep learning model accuracy trained with dataset selected by three mechanisms for different data distributions.

To better demonstrate the dataset content diversity, we use the selected data of different mechanisms to train a (Convolutional Neural Network) CNN model. For test images, we use dog images of known categories (unselected but matched images of sample images) and dog images from unknown categories. Here, we only consider the IdePrice to eliminate the effect caused by uneven selected image dataset sizes. We consider different image distributions on the seller side. As shown in Fig.10, in all cases, the diversity driven mechanism achieves best accuracy due to the better generalization capability. For dog images from known categories, when training (selected) images from 10 categories are uniformly distributed, the diversity driven mechanism produces most accurate model (80% for CROWDBUY and 75% for CROWDBUY++), while the matching degree mechanism achieves worst accuracy (53% for CROWDBUY and 49% for CROWDBUY++). When the training images are normally distributed, three mechanisms achieve comparably high accuracy (more than 90%), where the accuracy for diversity driven mechanism is 99.3% for CROWDBUY and 96.2% for CROWDBUY++. For dog images from unknown categories, accuracy of three mechanisms degrade to different extents. The degradation of the diversity driven mechanism is the smallest, and it still achieves 99.2% accuracy for CROWDBUY and 97.4% for CROWDBUY++. The results prove that our diversity driven mechanism is capable of selecting high-quality datasets for tasks like deep learning.

E. Efficiency

We run our components on a Huawei mate9 mobile phone. For each image, on average, it takes about 1s to extract the fc8 feature, 0.405s to generate the 2-dimension feature vector and 58ms to generate the MinHash signature for each image(k=4). Since we use only 2-dimension vector or 4 hash values for image selection, communication cost is also extremely small.

VII. Related work

Crowdsourcing Data Collection. Many existing crowdsourcing systems focus on issues of data aggregation and computation [7], [26], [29], incentive mechanisms [8], [18], [24], [27], [30], participatory sensing [9], etc. Most of these work focused on simple structured data like noise and motion data, [1], [2]. When it comes to images, there are some platforms collecting image data without quality guarantee [23] [9]. Some previous work collect image data and measure them by annotation quality [6], [13], or subjective test [22] [14]. Cao et al. [20], [21] leverage metadata as measurement. Those work couldn't efficiently evaluate the quality of image content, especially for large-scale images. Besides, most of them didn't consider image privacy and ownership, or malicious participants. Thus, how to collect large image sets with content quality guarantee and privacy/ownership respect is still very challenging.

Privacy-preserving Data Aggregation. For many crowdsourcing systems, sensitive and private user data could be exposed. Most existing work use encryption to protect data privacy, for example, BGV homomorphic encryption [32], additive homomorphic encryptions [10], [11], [28], homomorphic message authentication code [28] and ABE [12], [25]. All of those work focus on numerical value aggregation and computing, which cannot be applied to image data directly. Besides, those secure but expensive methods are considered not applicable in our large-scale image data setting.

VIII. Conclusion

To meet the urgent requirements for massive high-quality image datasets, as well as make good use of unlimited image data in personal devices, we took a first step towards image dataset purchasing via crowdsourcing. We propose, design and implement CrowdBuy and CrowdBuy++, with which a buyer can efficiently buy desired images with a certain quality guarantee, meanwhile respecting users' data ownership and privacy. A complete set of methods are designed to enable privacy-preserving image selection and quality measurement, which are also truthful and budget feasible. Our evaluation on large-scale images shows that our system is promising to collect high quality datasets for tasks like deep learning.

Acknowledgments

Xiang-Yang Li is the corresponding author. The research is partially supported by the National Key R&D Program of China 2017YFB1003000, China National Funds for Distinguished Young Scientists with No. 61625205, NSFC No. 61572281, No. 61472218, No. 61520106007, Key Research Program of Frontier Sciences, CAS, No. QYZDY-SSW-JSC002, NSF ECCS-1247944, and NSF CNS 1526638.

References

[1] "Noisetube," www.noisetube.net.
[2] "vtrack," www.vtrack.in/.
[3] Y. Bengio *et al.*, "Learning deep architectures for ai," *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.
[4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR*, 2009.
[5] A. Dosovitskiy and T. Brox, "Inverting visual representations with convolutional networks," in *IEEE CVPR*, 2016.
[6] P.-Y. Hsueh, P. Melville, and V. Sindhwani, "Data quality from crowdsourcing: a study of annotation selection criteria," in *NAACL HLT*, 2009.
[7] H. Jin, L. Su, H. Xiao, and K. Nahrstedt, "Inception: incentivizing privacy-preserving data aggregation for mobile crowd sensing systems." in *MobiHoc*, 2016, pp. 341–350.
[8] T. Jung, X.-Y. Li, W. Huang, J. Qian, L. Chen, J. Han, J. Hou, and C. Su, "Accounttrade: Accountable protocols for big data trading against dishonest consumers," in *INFOCOM 2017-IEEE Conference on Computer Communications, IEEE*. IEEE, 2017, pp. 1–9.
[9] S. S. Kanhere, "Participatory sensing: Crowdsourcing data from mobile smartphones in urban spaces," in *IEEE MDM*, 2011.
[10] Q. Li and G. Cao, "Efficient and privacy-preserving data aggregation in mobile sensing," in *IEEE ICNP*, 2012.
[11] ——, "Privacy-preserving participatory sensing," *IEEE Communications Magazine*, vol. 53, no. 8, pp. 68–74, 2015.
[12] X.-Y. Li and T. Jung, "Search me if you can: privacy-preserving location query service," in *INFOCOM, 2013 Proceedings IEEE*. IEEE, 2013, pp. 2760–2768.
[13] S. Nowak and S. Rüger, "How reliable are annotations via crowdsourcing: a study about inter-annotator agreement for multi-label image annotation," in *ACM MIR*, 2010.
[14] F. Ribeiro, D. Florencio, and V. Nascimento, "Crowdsourcing subjective image quality evaluation," in *IEEE ICIP*, 2011.
[15] Y. Shu, M. Zhu, K. He, J. Hopcroft, and P. Zhou, "Understanding deep representations through random weights," *arXiv preprint arXiv:1704.00330*, 2017.
[16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
[17] Y. Singer, "Budget feasible mechanism design," *ACM SIGecom Exchanges*, vol. 12, no. 2, pp. 24–31, 2014.
[18] A. Singla and A. Krause, "Truthful incentives in crowdsourcing tasks using regret minimization mechanisms," in *ACM WWW*, 2013.
[19] W. Wang, L. Ying, and J. Zhang, "The value of privacy: Strategic data subjects, incentive mechanisms and fundamental limits," *arXiv preprint arXiv:1603.06870*, 2016.
[20] Y. Wu, Y. Wang, and G. Cao, "Photo crowdsourcing for area coverage in resource constrained environments," in *IEEE INFOCOM*, 2017.
[21] Y. Wu, Y. Wang, W. Hu, and G. Cao, "Smartphoto: a resource-aware crowdsourcing approach for image sensing with smartphones," *IEEE TMC*, vol. 15, no. 5, pp. 1249–1263, 2016.
[22] Q. Xu, Q. Huang, and Y. Yao, "Online crowdsourcing subjective image quality assessment," in *ACM MM*, 2012.
[23] T. Yan, M. Marzilli, R. Holmes, D. Ganesan, and M. Corner, "mcrowd: a platform for mobile crowdsourcing," in *ACM SenSys*, 2009.
[24] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing," in *ACM Mobicom*. ACM, 2012.
[25] K. Yang, K. Zhang, J. Ren, and X. Shen, "Security and privacy in mobile crowdsourcing networks: challenges and opportunities," *IEEE Communications Magazine*, vol. 53, no. 8, pp. 75–81, 2015.
[26] L. Zhang, T. Jung, K. Liu, X.-Y. Li, X. Ding, J. Gu, and Y. Liu, "Pic: Enable large-scale privacy preserving content-based image search on cloud," *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 11, pp. 3258–3271, 2017.
[27] L. Zhang, X.-Y. Li, J. Lei, J. Sun, and Y. Liu, "Mechanism design for finding experts using locally constructed social referral web," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 8, pp. 2316–2326, 2015.
[28] L. Zhang, X.-Y. Li, Y. Liu, and T. Jung, "Verifiable private multi-party computation: ranging and ranking," in *INFOCOM, 2013 Proceedings IEEE*. IEEE, 2013, pp. 605–609.
[29] L. Zhang, K. Liu, X.-Y. Li, C. Liu, X. Ding, and Y. Liu, "Privacy-friendly photo capturing and sharing system," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2016, pp. 524–534.
[30] D. Zhao, X.-Y. Li, and H. Ma, "Budget-feasible online incentive mechanisms for crowdsourcing tasks truthfully," *IEEE/ACM TON*, vol. 24, no. 2, pp. 647–661, 2016.
[31] Z. Zheng, F. Wu, X. Gao, H. Zhu, G. Chen, and S. Tang, "A budget feasible incentive mechanism for weighted coverage maximization in mobile crowdsensing," *IEEE TMC*, 2016.
[32] G. Zhuo, Q. Jia, L. Guo, M. Li, and P. Li, "Privacy-preserving verifiable data aggregation and analysis for cloud-assisted mobile crowdsourcing," in *IEEE INFOCOM*, 2016.