Journal of Biomedical Informatics 74 (2017) 130–136

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Development of fine-grained pill identification algorithm using deep convolutional network

Yuen Fei Wong[a,*], Hoi Ting Ng[b], Kit Yee Leung[c], Ka Yan Chan[d], Sau Yi Chan[d], Chen Change Loy[c]

[a] Department of Pharmacy, Faculty of Medicine, University of Malaya, Malaysia
[b] Department of Pharmacy, Tuen Moon Hospital, Hong Kong
[c] Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong
[d] Department of Health Sciences, Caritas Bianchi College of Careers, Hong Kong

ARTICLE INFO

Keywords:
Tablet
Capsule
Error
Automatic
Deep learning

ABSTRACT

Objective: Oral pills, including tablets and capsules, are one of the most popular pharmaceutical dosage forms available. Compared to other dosage forms, such as liquid and injections, oral pills are very stable and are easy to be administered. However, it is not uncommon for pills to be misidentified, be it within the healthcare institutes or after the pills were dispensed to the patients. Our objective is to develop groundwork for automatic pill identification and verification using Deep Convolutional Network (DCN) that surpasses the existing methods.

Materials and methods: A DCN model was developed using pill images captured with mobile phones under unconstraint environments. The performance of the DCN model was compared to two baseline methods of hand-crafted features.

Results: The DCN model outperforms the baseline methods. The mean accuracy rate of DCN at Top-1 return was 95.35%, whereas the mean accuracy rates of the two baseline methods were 89.00% and 70.65%, respectively. The mean accuracy rates of DCN for Top-5 and Top-10 returns, i.e., 98.75% and 99.55%, were also consistently higher than those of the baseline methods.

Discussion: The images used in this study were captured at various angles and under different level of illumination. DCN model achieved high accuracy despite the suboptimal image quality.

Conclusion: The superior performance of DCN underscores the potential of Deep Learning model in the application of pill identification and verification.

1. Background

Oral pills, such as tablets and capsules, are pharmaceutical dosage forms that are very commonly used given their superior stability and ease of administration. They often come in various features, such as colors, shapes, scorings, and imprints (letters, numbers, or symbols engraved on the pills) that represent their identity, to a certain extent. Apart from the physical appearance of the actual pills, their packaging also provides information pertaining to the pills, both of which play pivotal roles in pill identification and verification.

Notwithstanding the assorted features that a pill could adopt as well as the self-explanatory information on the packings, misidentification happens occasionally, namely, one pill is mistaken for the other. Not only does pill misidentification lead to catastrophic outcomes to patients [1,2], such mistake contributes towards medication error, which has been inflicting a huge financial burden on healthcare cost worldwide [3].

It is foreseeable that pill misidentification would be particularly apparent in a busy healthcare setting, where the healthcare professionals are overwhelmed with a heavy workload, interspersed with frequent interruptions and distractions [4,5]. Pills that has ambiguous information on their labelling or packaging [4], especially the look-alike sound-alike medications [6–9] are particularly susceptible to misidentification. Therefore, various solutions, for instance, naming system of medications and barcoding, have been recommended and adopted to reduce potential medication error [6–8,10].

This study lays the groundwork, leveraging Deep Learning, for pill identification and verification. Deep Learning learns abstract high-level representation from data through multiple layers of non-linear transformations. The technique has gained remarkable performance in speech recognition, natural language processing, and computer vision. Of note, representation learned by Deep Convolutional Network (DCN) has been shown powerful in capturing abstract concepts invariant to various phenomenon in visual world. Successful applications based on

* Corresponding author.
E-mail address: yfwong@um.edu.my (Y.F. Wong).

http://dx.doi.org/10.1016/j.jbi.2017.09.005
Received 21 April 2017; Received in revised form 14 September 2017; Accepted 14 September 2017
Available online 15 September 2017
1532-0464/

DCN include face recognition [11–13], face alignment [14], image classification [15–17], object detection [18], and image restoration [19]. Here, we explored and established fundamental work, using DCN, for fine-grained pill identification given images captured by off-the-shelf handheld devices such as smartphones.

The approach will serve as an important enabling technology for a cost-effective solution that allows us to conveniently identify and verify pills. It is envisioned that this technology would serve as an invaluable tool for the various stakeholders in healthcare system.

1.1. Related work

1.1.1. Non-computer vision based approach

Various online platforms are now available to serve as an aid in identifying pills, for example, the 'Pillbox' by the United States National Library of Medicine [20], 'Pill Identifier' by Medscape [21], and 'Pill Identification tool' by WebMD [22]. These online platforms share a similar user interface, whereby users are required to manually input or select from the drop-down menus a series of features pertaining to the pill in a query, such as its shape, its color, as well as the presence or absence of imprints and scorings. While these platforms present a valuable database for queries on pills, there are not without pitfalls. First, the choices provided in the dropdown menu may not encapsulate the features queried. This is particularly prominent for the choices of color as, being a continuum feature, it is impossible to literally describe each color and their tones. Second, manual inputting of the information is susceptible to the subjectivity of users, for example on the interpretation of colors. Third, the requirement of manual inputting could be very time-consuming, especially when there are multiple pills that need to be identified.

1.1.2. Computer vision based approach

Recognizing the need for accurate recognition of pills, several approaches have been proposed to identify pill automatically through computer vision [23–29]. All the existing studies design some visual features for pill identification. Color features are usually based on hue, saturation, value (HSV) color profile due to its robustness to illumination variation. Apart from color, shape is among the most popular hand-crafted visual features. Caban et al. [23] propose the use of rotational invariant shape features. The method involves detecting the contour of pill as the first step. Points are then uniformly sampled on the contour and their distances to the pill's mass centre are computed, forming a distance vector that represents the shape of the pill. Statistics such as maximum, minimum, and standard deviation can then be extracted as features. Alternatively, cross-correlation of distance vectors between any two pills can be computed to return a score to represent the shape resemblance of the two pills. Hu moments [30] of shape contour have been adopted too [25,27]. Several studies focus on imprint extraction and representation for pill recognition. For instance, Chen and Kamata [24] and Yu et al. [28,29] employ mainly imprint feature of pills based on modified stroke width transform for recognition. Other study [27] adopts Scale Invariant Feature Transform (SIFT) [31] and Multi-scale Local Binary Pattern (MLBP) [32] to describe the imprint pattern.

With hand-crafted features and k-Nearest-Neighbor (k-NN) algorithm, Caban et al. [23] report an accuracy of 91.13% on 568 pill classes. We argue that manually designed features work well in a controlled environment but would yield poor performance in unconstrained settings such as with images captured by mobile devices. More precisely, manually designed features require one to have strong domain knowledge when crafting the features. The design is not data-driven and may generalize poorly if a wrong design is chosen [33]. This drawback can be avoided if good features can be learned automatically using a general-purpose learning procedure. Deep convolutional network is chosen in this study because it has shown strong performance in many real-world computer vision tasks such as image classification [33]. Its remarkable capability of learning representations that are important for discrimination makes it a preferable option than other existing algorithms that rely on hand-designed features.

More recently, two other independent groups had published their work using deep learning for pill identification. Wang et al. [34] attempted automatic pill identification by employing the GoogLeNet Inception Network [35] with elaborated data augmentation technique. On the other hand, Zeng et al. [36] adopted a more complex learning framework by first constructed three independent DCNs, using color, gray, and gradient images as inputs, respectively. The networks were trained using a triplet loss function [37–39] that is useful in learning a deep representation space. They then compressed the three DCNs into a single smaller network with knowledge distillation strategy [40]. The major difference between our work and the aforementioned two is that we have employed a QR-like board, as a reference of pill sizes, as well as to rectify geometric and color distortions, resulting in higher reliability.

2. Materials and methods

2.1. Data collection

A total of 400 commonly used tablets and capsules were collected from the dispensing laboratory at the Department of Health Sciences, Caritas Bianchi College of Career. These include pills used in the cardiovascular system (28.5%), nervous system (18.8%), gastrointestinal system (9.2%), endocrine system (8.8%), infection (7.7%), blood and nutrition (6.9%), musculoskeletal system (6.7%), respiratory system (6.3 %), genitourinary (3.5%), immune system and malignant disease (0.6%), dermatology (0.2%), and others (2.9%). The pills were categorized based on their dosage forms, presence or absence of imprints, shapes and colors, as shown in Fig. 1.

All pills were laid on a reference board and images were taken using two mobile devices of different operating system, at resolutions of 72 pixels/inch. Pills were arbitrarily placed at random spots, as long as they were within the boundary of the reference board. Pill images were deliberately captured at various angles, from different distances, and under different illumination conditions, to better reflect the real-world usage condition. Example of the images taken is shown in Fig. 2. Ten to twenty-five pictures were taken for each pill, including front and back images, amounting to 5284 images in total. The pill dataset was randomly divided into training-test partitions of 4884 and 400 images, respectively.

2.2. Development of DCN for pill identification

The pipeline of the proposed approach of using DCN for pill identification is depicted in Fig. 3. Prior to training the deep model, the images were subjected to two pre-processing steps:

2.2.1. Step 1 – geometric transformation

A pill image can be distorted due to arbitrary image capturing angle. The reference board provided an easy identification of the required registration for correcting the perspective distortion (Fig. 3A, step 1).

Specifically, the reference board consists of finder patterns at its four corners, as in a conventional QR code [41]. The four corners provide us with a spatial guide to perform geometric transformation. Specifically, given N tuples each consists of a pair of 2D data-points extracted from finder patterns, namely, $\{<\mathbf{x}_i, \mathbf{x}_i'>, i = 1,2,\cdots,N\}$, where $\mathbf{x}_i$ is the position of a detected pattern, and $\mathbf{x}_i'$ is the corresponding point in the data matrix to be reconstructed. In perspective projection, $\mathbf{x}_i$ is the homogeneous coordinate representation, and each pair of corresponding points gives two linear equations A_i H= 0, where H is the transformation matrix to be estimated and