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 handcrafted 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 handcrafted 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

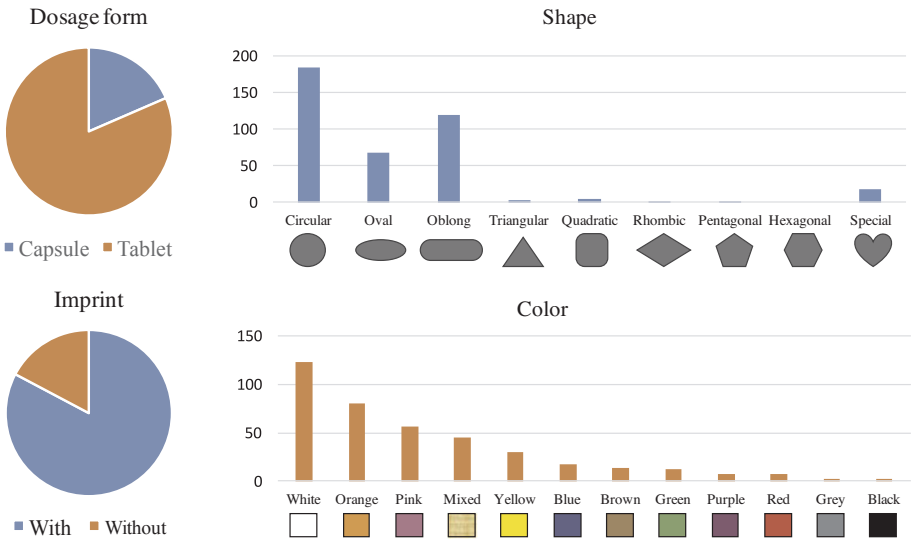

Fig. 1. Pills were categorized in accordance to their dosage form, shape, presence or absence of imprint and color. 81.5% of the pills were tablet and 83% were imprinted with symbols or letters.

$$A_i = \begin{bmatrix} \mathbf{0}^T & -\mathbf{x}_i^T & y_i'\mathbf{x}_i^T \\ \mathbf{x}_i^T & \mathbf{0}^T & -x_i'\mathbf{x}_i^T \end{bmatrix}.$$

The equations can be solved by using linear system solver.

2.2.2. Step 2 – Saliency-driven pill detection

After perspective correction, the pill foreground segment was detected using the manifold ranking-based saliency detection approach [42]. Manifold ranking [43] is a method that exploits the intrinsic manifold structure of data for graph labelling. Mathematically, a manifold ranking function can be represented as

$$\mathbf{f}^* = (\mathbf{I} - \alpha\mathbf{S})^{-1}\mathbf{y},$$

where is an identity matrix, α is a parameter set to 0.99, $\mathbf{S}$ is the normalized Laplacian matrix derived from an affinity graph, which defines the manifold structure of all nodes in a graph, $\mathbf{y}$ denotes an indication vector, which capture the values to be propagated in the graph, and $\mathbf{f}^*$ is the ranking values for all nodes. A detail introduction of manifold ranking can be found in [43].

In this work, the method presented in [42] was adopted due to its robust performance in preserving finer object boundaries compared to other existing methods. Briefly, a superpixel segmentation was first performed to group pixels in proximity into blocks of coherent pixel intensity. An affinity graph was then constructed, whose nodes are the superpixels and each node pairs were weighted by color similarity between the superpixels [42]. Manifold ranking was then performed for two stages consecutively using superpixels at the image boundary as background seeds [43] (Fig. 3A, step 2 and Fig. 3B).

2.2.3. Step 3 – Deep model training

Training the deep model directly with a small quantity of pill images is likely to cause overfitting problem, namely, the deep model performs

Fig. 2. Example of images taken at various angles, from different distances, and under different illumination condition.

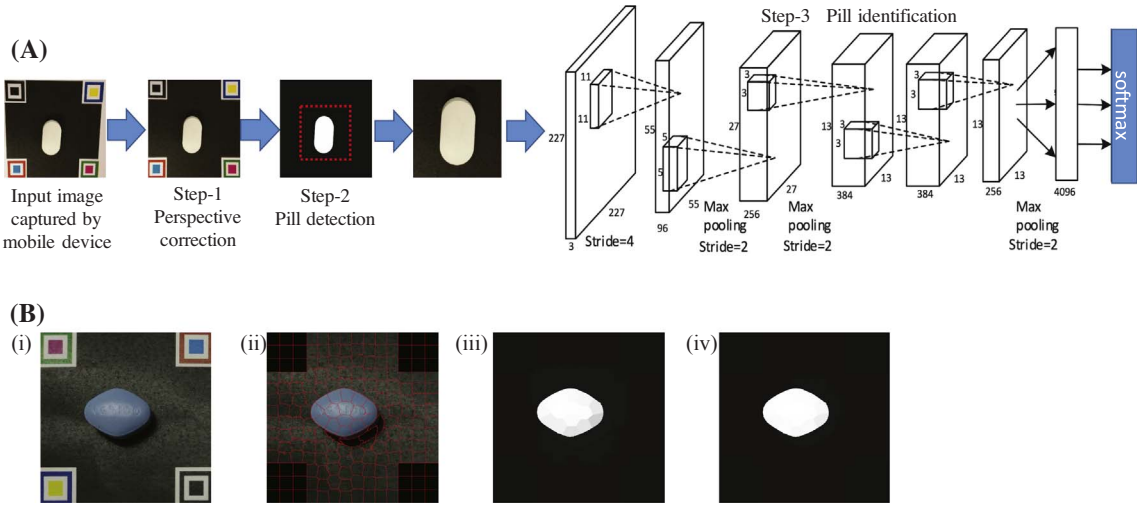

Fig. 3. (A) The general pipeline for pill identification: (Step-1) Perspective correction, (step-2) pill detection, (step-3) pill identification using a trained deep model. (B) From left to right: Pill image after (i) perspective correction, (ii) superpixels, (iii) first stage of saliency detection, and (iv) second stage of saliency detection. The foreground segment was then obtained by performing adaptive thresholding on the second-stage output of saliency detection.

well on the training set but yields poor results on unseen images. With this in mind, a deep model with a structure similar to AlexNet [14] was adopted, but differing from AlexNet, the fully-connected layer fc7 was removed to reduce the number of parameters in order to prevent overfitting given few pill images. The architecture of the deep model is shown in Fig. 3A, step 3. In addition, the model was pre-trained with the large-scale ImageNet dataset [44]. Moreover, the pill training images were augmented by introducing rotation, translation, and flipping to generate additional images to obtain a larger quantity of training images. The deep model pre-trained with ImageNet data was then fine-tuned on this augmented pill training set. As in AlexNet [14], the model was trained using stochastic gradient descent with a momentum of 0.9 and weight decay of 0.0005. A batch size of 256 examples, rather than 128 in AlexNet [14], to increase the convergence time. The update rule for weight ω is

$$\Delta_{i+1} = 0.9\cdot\Delta_i - \eta\cdot\frac{\partial L}{\partial W_i},\ W_{i+1} = W_i + \Delta_{i+1}$$

where W are the network parameters, i is the iteration, η is the learning rate, and $\frac{\partial L}{\partial w_i}$ is derivative of the softmax loss. The base learning rate was set to 0.001 and Caffe (http://caffe.berkeleyvision.org/) was used as the deep learning framework.

Pre-training the model with ImageNet data [44] is important to avoid overfitting. Experiments showed that training a deep model from scratch using the pill images would lead to a drop of over 2% in accuracy rate in comparison to fine-tuning from an ImageNet pre-trained model.

2.3. Development of hand-crafted features as baseline for comparison

Caban et al. [23] achieved remarkable pill identification accuracy on pill images captured with controlled illumination and viewpoint, through exploiting hand-crafted features and k-Nearest-Neighbor (k-NN) algorithm. Therefore, this method was chosen to serve as baseline for the performance comparison. To adopt this method on our pill dataset, which presents with perspective distortion and varied illumination, a more robust baseline built upon Caban et al.'s approach was developed for a fairer comparison. Specifically, geometric transformation [42] and manifold ranking-based saliency detection [43] steps, applied in the deep learning approach, was applied in this method to correct the perspective distortion and to detect foreground region. Next, shape, color, and texture features were extracted. Rotational-invariant shape features were extracted as reported by Caban et al. [23] For color features, reference colors at the four corner blocks of the reference board were used to normalize the color distribution of the pill images. Texture features were then extracted based on rotational-invariant local binary pattern (LBP), which has proven robust in many computer vision tasks [32].

Apart from the original k-NN method adopted in Caban et al. [23], a random forest method [45] was also evaluated. Briefly, the forest size was set to 100 trees, because there was no significant difference in its performance with a larger forest size. The number of feature variables, m_{try}, was set as $m_{try} = \sqrt{d}$, where d is the data feature dimension [45]. Linear data separation was employed as the test function for node splitting. The decision tree was grown until some stopping criterion was satisfied, e.g., leaf nodes were formed when no further split can be achieved given the objective function, or the number of training samples arriving at a node was smaller than the predefined node size, ∅. In this experiment, ∅ was set as 1 for capturing sufficiently fine-grained data structure.

The performance was assessed by the accuracy of each model in assigning the correct class to the 400 test pill images. In other words, the number of classes was set to the desired pill classes to be identified (400 in this study). Upon confronted with a queried pill image, the algorithms would do the background matching and rank the retrieved results, namely, the most likely match would be ranked number 1, the second likely match would be ranked number 2, the third likely match would be ranked number 3, and so on. After the ranking was done, one image for each query that was ranked number 1 would be labelled as the "Top-1" retrieval; five images for each query that ranked from number 1–5 would be labelled as the "Top-5" retrieval; 10 images for each query that ranked from number 1–10 would be labelled as the "Top-10" retrieval. The mean accuracy rates of the Top-1, Top-5 and Top-10, averaged across five repeated random sub-sampling (also known as Monte Carlo cross-validation), were then computed for each algorithm, respectively, for comparison.

3. Results and discussion

The pill identification mean accuracy, averaged across five repeated random sub-sampling, is summarized in Table 1. The training and test splits for each fold was formed by drawing samples randomly from the dataset. The mean accuracy rate for the hand-crafted visual representation in k-NN approach was the lowest, which is not unexpected. This is because hand-crafted visual representation was overwhelmed by redundant and noisy features, and the k-NN approach lacks

Table 1
Comparison of mean accuracy rates (%) and standard deviations across five repeated random sub-sampling (mean ± s.d).

Methods	Top-1	Top-5	Top-10
Features [23] with ***k***-NN	70.65 ± 0.72	85.50 ± 0.94	89.95 ± 1.02
Features [23] with random forest	89.00 ± 1.08	98.55 ± 0.65	99.50 ± 0.40
Deep Convolutional Network (DCN)	95.35 ± 0.22	98.75 ± 0.35	99.55 ± 0.21

an effective mechanism to choose the right features for identification. By employing random forest in hand-crafted feature, the mean accuracy rate was increased to 89.00%, 98.55%, and 99.50% at Top-1, Top-5, and Top-10 returns, respectively. This is largely because random forest performs implicit feature selection through maximizing information gain in its splitting nodes.

While the adoption of random forest in hand-crafted feature resulted in an improvement on pill identification, the performance of DCN model was more superior, particularly at Top-1 return. Specifically, the mean accuracy rate of DCN model at Top-1 return was further improved to 95.35%. The Top-5 and Top-10 returns of DCN model were 98.75% and 99.55%, respectively. Of note, unlike images that were fed to the baselines, the input images to DCN were not pre-processed for color normalization, which highlights the prominent advantage of DCN model in extracting meaningful representation despite the illumination variation.

To further determine whether the computed error rate truly represents the performance of the various features, a two-dimensional (2D) embedding figure of manually designed features and of DCN was plotted using multi-dimensional scaling (Fig. 4). The deep features, which each vector has a dimension of 4096, were extracted from the fully-connected layer of the DCN shown in Fig. 3A. As shown in Fig. 4, the hand-crafted features displayed points clustered together without a clear separation between different pills. On the other hand, the DCN model was more discriminative, characterized by a more dispersed pattern. In particular, similar looking pills lie closer to each other, for example, round, white color pills were closer in their coordinates; oblong, orange color pills were closer to each other in their coordinates but were at a distance from their round, white color counterparts. The 2D visualization provided a solid evidence that DCN model could identify pills at a much higher accuracy compared to the baselines.

In order to examine the performance of each feature, we looked into some of the pill queries that were misclassified (Fig. 5). It was apparent that the manually designed features with random forest were confused over shapes and imprints of the queried pills (Fig. 5A). Such elements represent the most distinctive elements of a given pill, which a non-professional would be able to discriminate easily. Whereas, these distinctive elements were captured by DCN and hence classification was made correctly. When examining the pill queries that DCN misclassified, it was found that those were extremely challenging cases, due to either a close resemblance of pills' appearance or blurry images (Fig. 5B). Identification of such pills based on the images could be arduous, even for experienced professional healthcare workers.

While there have been previous efforts working on pill identification, such studies focused on pill images captured in a controlled environment, namely with a specific angle, at pre-defined camera distance, and restrained under certain illumination [23,28]. It is thus conceivable that images captured under such constrained environment are of superb quality, hence facilitating the downstream identification task. In contrast, recognizing pills from images captured "in-the-wild" is non- trivial. As shown in Fig. 2, the pill images in this study were rarely captured in parallel with the capturing plane. In addition, the lighting condition was different, and some of the images were blurry due to camera shake. Attempting conventional methods [23–29] based on manually designed features on such challenging dataset would yield poor performance. The suboptimal performance is largely anticipated because the inherent algorithms are susceptible towards such noise. In particular, these methods first pinpoint the pill location in the image, separate the pill from the background by foreground segmentation algorithm, extract the shape statistics (e.g., area, perimeter, and roundness), color distribution, and imprint features, finally employ a classifier such as k-NN algorithm for pill classification. The segmentation algorithm is susceptible to illumination and shadow cast on the pill images. In addition, the shape statistics of the pill cannot be well described given the failed foreground segmentation. This was also clearly demonstrated in our study, whereby the original baseline and the improved version of it were inferior to the DCN model.

From the clinical practicality point of view, an accuracy rate of 95.35%, 98.75%, and 99.55% for Top-1, Top-5, and Top-10 return,

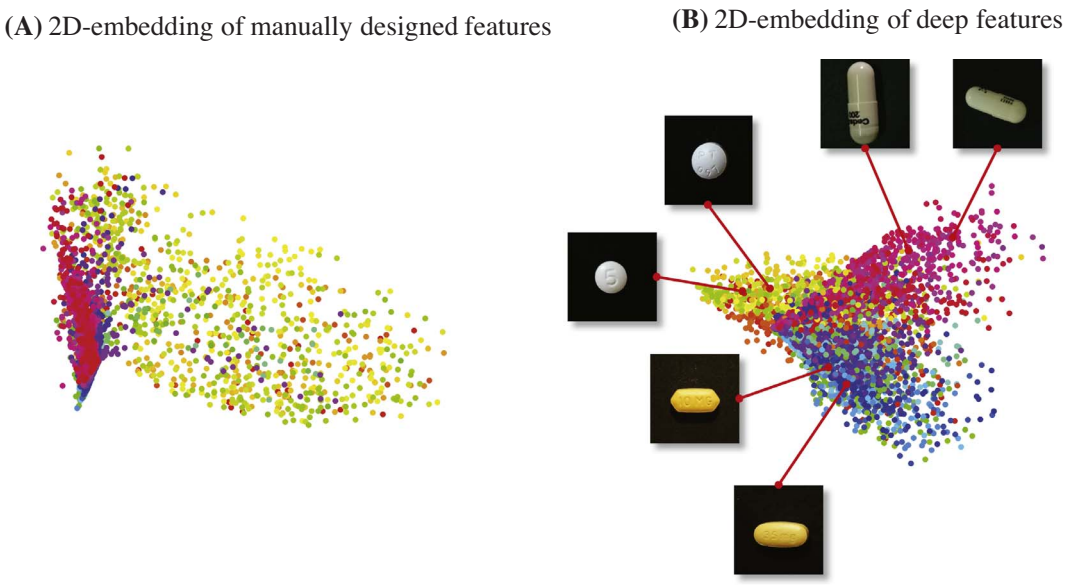

Fig. 4. Two-dimensional embedding of pill patterns obtained using multi- dimensional scaling. Each point corresponds to representation of a pill obtained through (A) manually designed features and (B) hidden features extracted from the fully connected layer of deep convolutional network. Every point is encoded by color based on its associated class. The positions of six distinct pills are also shown in (B).

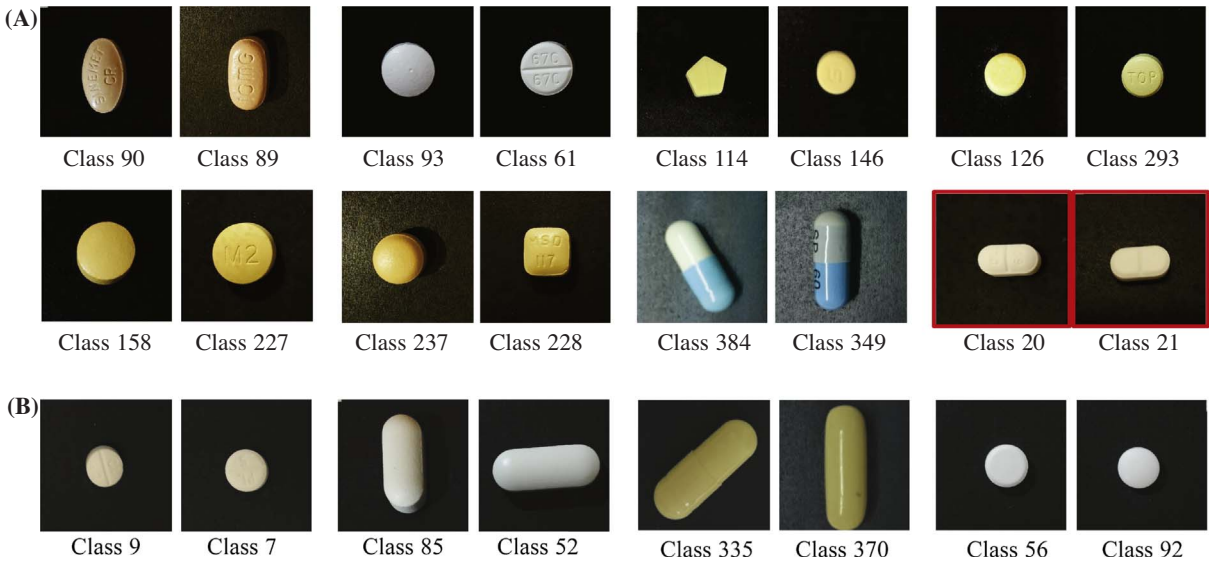

Fig. 5. (A) Failure cases of baseline that exploits manually designed features with random forest as classifier. The baseline fails on all these while Deep Convolutional Network (DCN) succeeds in all except for the last one, where pill Class 20 was mistaken as Class 21 (highlighted in red borders). (B) Pill queries that DCN misclassify were mostly those that bear minimal identification features, namely of standard shapes and without any imprint. The left image of each image pair is the query, while the right one is the search result returned by the baseline or the DCN.

respectively, is still rather low to allow unsupervised, fully automated pill identification. Nonetheless, it would be invaluable to serve as an assisting tool for the pharmacists, the pharmacy technicians, and other healthcare professionals, to double check or to verify the medications prior to dispensing to the patients. Notwithstanding the superiority of DCN model compared to the baseline methods, there are still rooms for improvement. Specifically, the current study had only included 400 commonly used pills, each of which contributed ten to twenty-five pictures for the training of the DCN model. It is foreseeable that by enlarging the training dataset through including more pills, as well as by capturing more photos at different angles and under different illuminations, the performance of DCN model could be greatly improved.

4. Conclusion

We have shown in this study that DCN achieves exceptional results in identifying pills captured under unconstraint environment. As far as we know, this is the first study that leverages DCN for pill identification. The highly promising results justify further expansion of the work, by enlarging the scale of the current dataset and improving the deep network with contemporary techniques, such as multi-tasks learning [16]. The saliency detection step can be skipped by performing direct pill segmentation on the captured images [46].

Acknowledgements

The authors would like to thank Mr. Fong Alfred Ching To for helping on data collection, and Dr. Chen Huang, Mr. Xintao Wang, and Ms. Zhiyi Cheng for helping to process the data and fine-tune the deep network.

Competing interests

The authors declare no competing interests.

References

[1] N. Collins, Grandmother Dies after Receiving Wrong Prescription. The Telegraph 2014 (accessed 4 January 2017).
[2] A. Crook, M. Walker, Baby rushed to hospital unconscious after being given anti psychotic drugs in pharmacy mix up. Mirror 2016 (accessed 4 January 2017).
[3] C. Andel, S.L. Davidow, M. Hollander, et al., The economics of health care quality and medical errors, J. Health Care Finance 39 (1) (2012) 39–50.
[4] ASHP guidelines on preventing medication errors in hospitals Am. J. Hosp. Pharm. 1993, 50(2), 305–314.
[5] A. Beso, B.D. Franklin, N. Barber, The frequency and potential causes of dispensing errors in a hospital pharmacy, Pharm. World Sci. 27 (3) (2005) 182–190.
[6] L.M. Emmerton, M.F. Rizk, Look-alike and sound-alike medicines: risks and 'solutions', Int. J. Clin. Pharm. 34 (1) (2012) 4–8.
[7] S. Gabriele, The role of typography in differentiating look-alike/sound-alike drug names, Healthc Q 9 (2006) Spec No: 88-95.
[8] R. Ostini, E.E. Roughead, C.M. Kirkpatrick, et al., Quality use of medicines-Medication safety issues in naming; look-alike, sound-alike medicine names, Int. J. Pharm. Pract. 20 (6) (2012) 349–357.
[9] J.M. Hoffman, S.M. Proulx, Medication errors caused by confusion of drug names, Drug Saf. 26 (7) (2003) 445–452.
[10] A. Berman, Reducing medication errors through naming, labeling, and packaging, J. Med. Syst. 28 (1) (2004) 9–29.
[11] Y. Sun, Y. Chen, X. Wang, et al., Deep learning face representation by joint identification-verification, Adv. Neural Inform. Proc. Syst. (2014) 1988–1996.
[12] F. Schroff, D. Kalenichenko, J. Philbin, FaceNet: A unified embedding for face recognition and clustering, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
[13] Y. Wen, K. Zhang, Z. Li, et al. A discriminative feature learning approach for deep face recognition, in: European Conference on Computer Vision (ECCV), 2016.
[14] Z. Zhang, P. Luo, C.C. Loy, et al., Learning deep representation for face alignment with auxiliary attributes, IEEE Trans. Pattern Anal. Mach. Intell. 38 (5) (2016) 918–930.
[15] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inform. Proc. Syst. (2012) 1106–1114.
[16] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2015 arXiv, 1409–1556.
[17] X. Zhang, Z. Li, C.C. Loy, et al. PolyNet: A pursuit of structural diversity in very deep networks, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017.
[18] W. Ouyang, X. Wang, X. Zeng, et al. DeepID-Net: Deformable deep convolutional neural networks for object detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
[19] C. Dong, C.C. Loy, K. He, et al., Image super-resolution using deep convolutional networks, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2) (2016) 295–307.
[20] Pillbox. http://pillbox.nlm.nih.gov/pillimage/search.php (accessed 4 March 2017).
[21] Pill Identifier. http://reference.medscape.com/pill-identifier (accessed 4 March 2017).
[22] Pill Identification Tool. http://www.webmd.com/pill-identification/ (accessed 4 March 2017).
[23] J.J. Caban, P. Rheingans, T. Yoo, Automatic identification of prescription drugs using shape distribution models, in: IEEE International Conference on Image Processing (ICIP), 2012.
[24] Z. Chen, S.-I. Kamata, A new accurate pill recognition system using imprint information, in: International Conference on Machine Vision, 2013.
[25] A. Cunha, T. Adão, P. Trigueiros, HelpmePills: A mobile pill recognition tool for elderly person, Proc. Technol. 16 (2014) 1523–1532.
[26] D.H. Andreas, C. Arth, Computer-vision based pharmaceutical pill recognition on mobile phones, Central European Seminar on Computer Graphics, 2010.

[27] Y.-B. Lee, U. Park, A.K. Jain, et al., Pill-ID: Matching and retrieval of drug pill images, Pattern Recogn. Lett. 33 (7) (2012) 904–910.
[28] Yu J, Chen Z, Kamata S-i. Pill recognition using imprint information by two-step sampling distance sets, in: International Conference on Pattern Recognition 2014, pp. 3156–3161.
[29] J. Yu, Z. Chen, S.-i. Kamata, et al., Accurate system for automatic pill recognition using imprint information, IET Image Proc. 9 (12) (2015) 1039–1047.
[30] M.-K. Hu, Visual pattern recognition by moment invariants, IRE Trans. Inform. Theory (1962) 179–187.
[31] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vision 60 (2) (2004) 91–110.
[32] T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution gray-scaleand rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 971–987.
[33] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[34] Y. Wang, J. Ribera, C. Liu, et al. Pill recognition using minial labeled data, in: IEEE Conference on Multimedia Big Data, 2017.
[35] C. Szegedy, W. Liu, Y. Jia, et al. Going deeper with convolutions, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
[36] X. Zeng, K. Cao, M. Zhang, MobileDeepPill: A small-footprint mobile deep learning system for recognizing unconstrained pill images, in: International Conference on Mobile Systems, Applications, and Services (MobiSys) 2017.
[37] J. Wang, Y. Song, T. Leung, et al. Learning fine-grained image similarity with deep ranking, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.
[38] F. Schroff, D. Kalenichenko, J. Philbin, FaceNet: A unified embedding for face recognition and clustering, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.
[39] C. Huang, Y. Li, C.C. Loy, et al. Learning deep representation for imbalanced classification, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016.
[40] G. Hinton, O. Vinyals, J. Dean, Distilling the knowledge in a neural network 2015 arxiv.org/abs/1503.02531.
[41] Yang Z, Cheng Z, Loy CC et al. Towards robust color recovery for high-capacity color QR codes. IEEE International Conference on Image Processing (ICIP) 2016.
[42] C. Yang, L. Zhang, H. Lu, et al. Saliency detection via graph-based manifold ranking, in: Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 3166–3173.
[43] D. Zhou, J. Weston, A. Gretton, et al., Ranking on data manifolds, Adv. Neural Inform. Proc. Syst. (2004) 169–176.
[44] J. Deng, W. Dong, R. Socher, et al. ImageNet: a large-scale hierarchical image database, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.
[45] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
[46] Z. Liu, X. Li, P. Luo, et al. Semantic image segmentation via deep parsing network, in: IEEE Conference on Computer Vision (ICCV), 2015.