Abstract—Hairstyle recognition is a challenging task since hairstyles span a diverse range of appearances in real-world. However, it is possible to start from recognizing the most basic hairstyles then dealing with more complex hairstyles. In this paper, we present a novel hairstyle pattern recognition system based on CNNs. We first give the definitions of four basic hairstyles: straight hairstyle, curly hairstyle, kinky hairstyle, and braid hairstyle. Then we leverage the power of the pre-trained CNN model to learn the distinct features of those basic hair patterns from representative hairstyle patch dataset. The CNN model can classify different hairstyles from patches and perform hairstyle recognition in full hair images. Patch-based recognition makes our system very flexible to be applied to hair images that captured from multi-views. The experiment results show that our system can perform recognition for both simple hairstyle images and complex hairstyle images.

I. INTRODUCTION

Hairstyle is a very important part of a persons appearance in the real world, it can help to provide certain identification and personality. Moreover, the hairstyle can be a supplementary feature for human recognition tasks[1], [2].

However, hairstyle recognition remains one of the most challenging tasks due to the characteristics of hair, such as omnipresent occlusion, specular appearance, and complex discontinuities [3]. Researchers usually recover the 2D/3D hair strand information based on single/multiple-view hair images [4] [5]. However, for certain hairstyles, hair strand information is very difficult to obtain directly from images. For example, the structure of the kinky hairstyle, as shown in Figure 1, is nearly impossible to discover due to the tiny coiled structure and (usually) the dark color. Thus the recognition of different hairstyles usually remains a manual work.

In addition, real-world hairstyles span a diverse range of appearances. Hairstyles can be divided into two main categories: simple hairstyles and complex hairstyles. The simple hairstyles only contain a single hairstyle pattern. However, even the simple hairstyle category can be divided into dozens of sub-categories based on certain criteria (e.g. the curly hairstyle can be divided by the curliness, and the length of the hair strands, etc). Thus it's very difficult to give the precise definitions of all hairstyle patterns. Following on the divided and conquer methodology, it is possible to start with recognizing the most basic hairstyle patterns, then improve the classification by utilizing more detailed information in each category. Since our system aims at recognizing four basic hairstyle patterns, we first provide the definitions of four basic hairstyle patterns as:

- Straight hairstyle: hair is normally straight and do not hold a curl;
- Curly hairstyle: hair contains spirals or inwardly curved forms, or has a definite "S" pattern;
- Kinky hairstyle: hair is tightly coiled with a less visible curl pattern;
- Braid hairstyle: usually two to four hair strands inter-lacing with each other to form a complex structure or pattern.

The examples of the four basic hairstyles are shown in Figure 1.

Moreover, the complex hairstyles are in the form of the combination of the basic hairstyles. As shown in Figure 2, the waterfall hairstyle contains three different basic hairstyles: straight hairstyle (indicated by the red stroke), curly hairstyle (indicated by the yellow stroke), and a braid (indicated by the green stroke) that lies between them. Thus, it is reasonable to achieve rough hairstyles segmentation based on the hairstyle recognition results for the complex hairstyle.

Since each basic hairstyle is composed of certain repeated patterns according to their definitions. It is reasonable to...
design our system to train on hairstyle patches which contain the distinct pattern features, then to perform recognition on full hair images. This strategy improves the flexibility of our method when dealing with complex hairstyles. It is suitable to be applied to hair images that captured from multi-views.

Extract the unique feature of each hairstyle is crucial in hairstyle recognition. However, it is very difficult to use traditional image processing methods to obtain representative features for certain hairstyle. Thus, we apply the convolutional neural networks (CNNs) and leverage its strength to extract hairstyle features and perform hairstyle recognition. CNNs are often combined with large-scale datasets (e.g. ImageNet [6]) to surpass other methods for image recognition and segmentation.

In our system, we have four classes and each class contains approximate 1000 patches for training, 300 patches for evaluation, and 200 patches for testing. The small-scale hairstyle patch dataset itself makes the hairstyle recognition a more challenging learning problem, although it is very realistic to obtain the small-scale dataset in real-world use cases. Thus, we applied the pre-trained CNNs with a final layer retrained to our own hairstyle patch dataset to perform hairstyle recognition.

In summary, the main contributions of this paper are:

- We build a hairstyle dataset that contains basic hairstyles patches for hairstyle recognition task.
- We introduce a novel hairstyle pattern recognition system to recognize four basic hairstyles. Especially, it is the first time that the kinky hairstyle is included in hairstyle recognition.
- The strategy, which is training on hairstyle patches and performing recognition in full hairstyle image, facilitates the recognition for both simple hairstyles and complex hairstyles. The hairstyle recognition system can be applied to multi-view hair images.

II. RELATED WORKS

A. Hair recognition for human identity recognition

Researchers use human hair as a supplementary feature for human identification recognition.

Yacoob estimated a set of attributes of the hair from a single image. They developed algorithms and associated metrics that enable detection, representation, and comparison of the hair of different subjects. Their experiment results showed that the hair attributes can help to improve the human identification [1]. In their work, they provided important information for hair detection and description by introducing the hair attributes (e.g. length, volume, surface area, dominant color, etc). However, since the purpose of their work is human identification, the images in their experiments are human frontal face images. In addition, their work aim at measuring the attributes of the whole hair region. On the contrary, our work aims at recognizing different hairstyle patterns inside the hair region from multi-view hair images.

Dass used unsupervised learning method to discover distinct hairstyles, namely the whole hair regions, from a large number of frontal face images. Their learning method involved unsupervised clustering of hair regions, where they do not need to assume any pre-determined number of clusters. For each hairstyle region cluster, they generate a style template, which is a probability mask indicating the probability of hair at a certain position of a facial image. The templates are subsequently used to recognize hairstyles. In their experiments, they collected human face images randomly from the Internet. Clustering on these selected images resulted in five clusters of hairstyles. The five different hairstyles probability masks are generated and the accuracy of the classification is 75.62% [2].
as the work of [1], their experiments are based on frontal face images. And they focus on the recognition of complete hair regions rather than the detailed segmentation of different hairstyles inside the hair region.

To sum up, research works on the detailed hairstyle pattern recognition have not been fully explored. Furthermore, compared to unsupervised learning algorithms, supervised learning methods can be used in hairstyle recognition by providing high-level features leading to more reliable recognition results.

B. Convolutional neural networks

Convolutional Neural Networks (CNNs) have been widely adopted in many classification and segmentation tasks, including object recognition [7] etc., and demonstrated to provide superior performance than traditional classification and segmentation systems. There have been many successful CNNs models that developed and tested by state-of-the-art research works [8] [9] [10] [11] [12]. Usually, CNNs require a large amount of training data (e.g. ImageNet) in order to reach the best performance and avoid overfitting. Trained on large datasets, the CNN can learn useful features and leverage such features to reach a better accuracy than other methods that rely on the small datasets. However, for our hairstyle recognition system, only a small amount of training data is available. In order to avoid over-fitting, we obtain the CNN model that already trained on a larger dataset from ImageNet. Once the network parameters have converged, we can perform an additional training step using our own data to fine-tune the network weights. Furthermore, in order to make the best use of our data, we applied the augment technique to increase our datasets. Thus the final layer of our CNN model can be retrained on our own hairstyle patches dataset.

III. Hairstyle pattern recognition

The overview of the hairstyle recognition system is shown in Figure 3. The whole system can be divided into two parts:

- Hairstyle patch recognition.
- Full hairstyle image recognition.

A. Datasets

All hairstyle images used in our system are obtained from Flickr. Those hairstyle images are captured from multi-views and show a variety of colors, length, and volumes, etc. The average hairstyle image size in our dataset is approximate (450 - 600) pixels × (450 - 600) pixels. Those hairstyle images are then separated into two groups. In each group, the hairstyle images are evenly sampled. There is no overlapping images between those two groups.

B. Hairstyle patch recognition

In order to prepare the hairstyle patches training dataset, we crop the hairstyle images from the first group into hairstyle patches and assign the label for each of them. The criterion of cropping is that we need to reserve the distinguish structures of different hairstyles. Take the braid hairstyle for example, if the cropping windows are very small, the image patches will lose the ability to represent their unique interlacing structure pattern and every image patch will look like the straight hairstyle. On the contrary, if the cropping window is very large, it may contain several hairstyles inside it and make the recognition difficult. Thus, instead of using a fixed-size window for hairstyle image patch cropping, we made the size of the cropping window adjustable and can capture the unique hairstyle structure of its own kind. After the cropping procedure, we adjust the size of each hairstyle image patch into 75 pixels × 75 pixels. The hairstyle image patches are shown in Figure 4. The first row shows the straight hairstyle patches, the second row shows the curly hairstyle patches, the third row shows the kinky hairstyle patches, and the fourth row shows the braid hairstyle patches.

Then we separate all the hairstyle patches into training dataset and testing dataset. For each hairstyle, we use 1000 images for training, 300 images for validation, and 200 images for testing. Given our small training dataset, augment methods are used in order to avoid the over-fitting problem. Those methods including rotation, vertical shift, horizontal shift, shearing transformation, and horizontal flip. The data argument results for a braid hairstyle patch are shown in Figure 5. During the training stage, we apply the random transformations and normalization operations on our hairstyle patch dataset and generate augmented hairstyle patches and their labels at the same time.

In our system, we applied the Inception V3 network [13] with a final layer retrained to our hairstyle image patches dataset. The original Inception V3 network is trained on ImageNet. We add a final layer retrained to our own hairstyle dataset to learn features for different hairstyles. The hairstyle patch CNN model runs 4000 training steps. Each step chooses 10 hairstyle patches at random from the training set to predict the labels. Those predictions are then compared against the actual labels to update the final layers weights through the back-propagation process. Then we run the testing procedure based on a set of hairstyle patches that kept separate from the training and validation patches. The final testing accuracy reaches 93.4%.
C. Full-image hairstyle recognition

The full images used in hairstyle recognition are selected from the second group. Since those images contain both hair regions and non-hair regions (e.g. faces, backgrounds, etc), the non-hair region should be removed before the hairstyle recognition procedure. Thus we select points on the boundary of the hair region to generate the hair region mask and obtain only the hair region, as shown in Figure 6.

Usually, the sliding window technique is used to search and recognize the target [14]. However, for hairstyles images, there are no strong and clear boundaries between each basic hairstyle region. We need to find the rough boundaries between the hairstyle regions in order to perform recognition. Thus, the segmentation method we choose need to satisfy the following demands: 1) the algorithm should utilize the rich features of hair images. 2) the results should emphasize potential boundaries between different hairstyles. 3) the computational space and time should be reasonable. We segment the hair region into super-pixels by applying the simple linear iterative clustering (SLIC) algorithm [15]. The SLIC algorithm, which adapts a k-means clustering approach to efficiently generate superpixels. The SLIC algorithm is very simple to implement. Moreover, the algorithm is both time and memory efficient. The superpixels generated by SLIC are compact, uniform in size, and adhere well to region boundaries. The superpixel segmentation results are shown in Figure 6.

Given the superpixel segmentation results, we need to generate hairstyle patches satisfy the input requirements of the hairstyle recognition CNN model. Thus we calculate the centroid points of the superpixels using Equation 1 and Equation 2.

$$X_i = \frac{\sum_{n=1}^{m} x_{in}}{m}$$

$$Y_i = \frac{\sum_{n=1}^{m} y_{in}}{m}$$

where $X_i$ and $Y_i$ are the coordinates of the centroid point of the $i$-th superpixel and $m$ is the amount of pixels in the $i$-th superpixel.

Regardless of the shape of the superpixel, we wrap all the pixels in a tight bounding box around each centroid point to the required size (75 pixels × 75 pixels). Noting that the patch size is just a little bit larger than the superpixel. Thus we have the superpixels superpixel$_i$ and their corresponding image patches patch$_i$. The generated image patches are shown in the third row of Figure 7. The red lines indicate the boundaries of the superpixels. The blue dots are the centroid points of the superpixels. The green boxes indicate the hairstyle image patches that generated based on the centroid points. As shown in Figure 7, the first superpixel lies on the boundary of the hair region, so the information that contained in the superpixel is limited. However, the corresponding hairstyle patch is able to include more hairstyle information.

For each hairstyle image patch patch$_i$, the hairstyle recognition result is composed of the class labels and the corresponding scores (label$_n$, score$_n$). Noting that the scores satisfy $\sum_{n=1}^{4} score_n = 1$. If the score score$_n$ of the $i$-th image patch is larger than the predefined threshold value threshold (= 0.75), then corresponding label label$_n$ will be assigned to the $i$-th image patch as well as the $i$-th superpixel. For example, if score$_4$ of the image patch patch$_{50}$ is larger than the threshold, then label$_4$ which indicates the

Fig. 6. Hair region mask and superpixel segmentation result.

Fig. 7. Hairstyle patches generation results. The second column shows the hairstyle superpixels. The third row are the corresponding 75 pixels × 75 pixels hairstyle patches.

Fig. 8. The kinky hairstyle recognition results.
braid hairstyle will be assigned to the corresponding super pixel superpixels50. The recognition results for two hairstyle image patches from Figure 7 are shown in Table I.

For the first hairstyle patch, the recognition result is (Curly, 0.574778092). Since the score is less than the threshold (= 0.75), it is recognized as NAN. The recognition result of the second hairstyle patch (Kinky, 0.999615). The recognition result for the complete hairstyle image is shown in Figure 8.

IV. EXPERIMENTS AND RESULTS

We conduct experiments on different hairstyle images, including both simple hairstyle images and complex hairstyle images that captured from multi-views.

A. Simple hairstyle recognition

The straight hairstyle recognition results are shown in the first row of Figure 9. The curly hairstyle recognition results are shown in the second row of Figure 9. There are several hairstyle patches are recognized as kinky hairstyle with high scores. However, by careful observation, we found that those hairstyle patches actually belong to the kinky hairstyle. The kinky hairstyle recognition results are shown in the third row of Figure 9. The recognition results indicate that our system can recognize the kinky hairstyle. The braid hairstyles are shown in the last row of Figure 9. The boundary of the braid is recognize as curly hairstyle because of the S pattern in those image patches.

B. Complex hairstyle recognition

The recognition results of the combination of straight hairstyle and braid hairstyle are shown in the first row of Figure 10. We notice that there is one misclassified patch (in green color) at the top of the hairstyle image and another one (in blue color) at the bottom of hairstyle image. The first hairstyle patch is very near the dividing line in the hair region. Since the appearance of the region near the dividing line shows an interlacing structure, the hairstyle patch is considered as braid hairstyle. The second one only contains limited hair region and the limited region is very similar to the straight hairstyle. The recognition results of the combination of curly hairstyle and straight hairstyle are shown in the second row of Figure 10. The curly hairstyle in Figure 10 can be described as very loose and contains the "S" pattern. From the recognition results, we can find that the upper part of the hair region is recognized as straight hairstyle and the lower part is recognized as curly hairstyle. The recognition results of a more complex braid structure that combined with straight hairstyle are shown in the last row of Figure 10.

V. CONCLUSION

In this paper, we present a novel hairstyle recognition system. We first provide the definitions for four basic hairstyles. Then we create the hairstyle patch dataset in order to perform hairstyle patch recognition and full hair image recognition. We leverage the power of the pre-trained Convolutional Neural Networks to extract the features for each hairstyle. During the training procedure, data augment techniques and transfer learning are applied to deal with the problem of over-fitting that cause by our small hairstyle dataset. The experiment results show that our hairstyle pattern recognition system can distinguish four basic hairstyles. Moreover, the strategy of training on patch-level and testing on image-level can facilitate the recognition for multi-view complex hairstyles. In the future, we need to increase our datasets to include more hairstyles. Furthermore, we need to carefully examine the detailed characteristics (e.g. the curliness for curly hairstyle, or how many stands are there in a certain braid hairstyle region, etc) for different hairstyles.

REFERENCES

Fig. 9. Simple hairstyle recognition results.

Fig. 10. Complex hairstyles recognition results.