

Braid Hairstyle Recognition based on CNNs

Chao Sun and Won-Sook Lee

EECS, University of Ottawa, Ottawa, ON, Canada
{csun014, wslee}@uottawa.ca

Keywords: Braid Hairstyle Recognition, Convolutional Neural Networks.

Abstract: In this paper, we present a novel braid hairstyle recognition system based on Convolutional Neural Networks (CNNs). We first build a hairstyle patch dataset that is composed of braid hairstyle patches and non-braid hairstyle patches (straight hairstyle patches, curly hairstyle patches, and kinky hairstyle patches). Then we train our hairstyle recognition system via transfer learning on a pre-trained CNN model in order to extract the features of different hairstyles. Our hairstyle recognition CNN model achieves the accuracy of 92.7% on image patch dataset. Then the CNN model is used to perform braid hairstyle detection and recognition in full-hair images. The experiment results shows that the patch-level trained CNN model can successfully detect and recognize braid hairstyle in image-level.

1 INTRODUCTION

Hairstyle, which can help to provide unique personality, is considered as one of the most important features of a human being in real-world. Moreover, in computer games and animation films, different hairstyles represent different identifications of virtual characters. However, hairstyle recognition remains one of the most challenging tasks due to the characteristics of the hair (e.g. the texture, colors, etc), the variety of appearances under different environments (e.g. lighting conditions, etc), as well as countless combinations of different hairstyles.

Most of the researchers who work on 3D hair modelling examine the characteristics of hair based on single-view or multiple-view hair images and try to obtain hair strands structure information (e.g. orientation of hair strands). For certain hairstyles, such as straight hairstyle, this kind of information is relatively easy to obtain since the straight hair strands share the same direction. However, for more complex hairstyle, such as the braid hairstyle, the corresponding recognition procedure is more challenging, and is usually performed by human. Thus, an automatic braid hairstyle recognition system is needed in order to facilitate the hair modelling procedure.

The main challenges for braid hairstyle recognition are:

- The braid hairstyle spans a diverse range of appearances in real-world, it is very difficult to use hand-designed image features to recognize. Ex-



Figure 1: Different braid hairstyles.

amples of braid hairstyle are shown in Figure 1. They are "french braid", "reverse french braid", "fishtail braid", and "four-strand braid".

- The braid hairstyle often co-exist with other hairstyles, thus the hair strands usually share similar appearance. As shown in Figure 2, the hair image contains three different hairstyles: straight



Figure 2: The combination of different hairstyles.

hairstyle (indicated by the blue stroke), curly hairstyle (indicated by the yellow stroke), and a braid (indicated by the green stroke) that lies between those two regions. The only difference is the structure or pattern that formed by hair strands.

- The boundaries between the braid hairstyle and other hairstyles are difficult to detect. As shown in Figure 1, the hair strands gradually merge into the braid region and become a part of the braid.

Braid hairstyle is defined as two to four hair strands interlacing with each other to form a complex structure or pattern. Since the braid hairstyle is composed by certain repeated patterns, indicating that the most distinguish pattern lies in the interlacing area. Thus, if we can detect the interlacing pattern in the hair images, then we can locate the braid area in the full hair images.

Since braid hairstyle is usually co-exist with other hairstyles, it is reasonable that we develop a recognition system that can learn features from both the braid hairstyle and non-braid hairstyle and separate them based on those features. Thus, we include three other hairstyles in our system. They are 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), and kinky hairstyle (hair is tightly coiled with a less visible curl pattern).

Due to the characteristics of different hairstyles, traditional image processing methods usually failed to extract the structure features from hair images directly (e.g. the kinky hairstyle). Thus, we leverage the strength of the Convolutional Neural Networks (CNNs) to automatically learn features of different hairstyles. Usually, the CNNs are trained on large-scale image dataset (e.g. ImageNet(Deng et al., 2009), etc), however, our hairstyle patch dataset is a

relatively small one. There are four hairstyle classes and each class has approximate 1000 image patches, including 800 patches for training and 200 patches for testing. When dealing with small dataset, which is realistic in real-world use cases, overfitting is the main problem we need to avoid. although. Thus, we apply the transfer learning via a pre-trained CNNs with a final layer retrained to our own hairstyle dataset to learn features for different hairstyles.

To sum up, the contributions of this paper are:

- A novel hairstyle recognition system that can detect the unique features of braid hairstyle and recognize braid hairstyle in full hair images.
- The strategy of patch-level feature learning and image-level recognition can facility the recognition for complex hairstyles. The hairstyle recognition system can be applied to the front-view hair images, the side-view hair images, as well as the back-view hair images.

2 RELATED WORK

2.1 Hair Recognition in Human Identification

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

Yacoob et. al estimated a set of attributes (e.g. length, volume, surface area, dominant color, coloring, etc) of the head 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 shown that the hair attributes can improved the human identification results(Yacoob and Davis, 2006). In their work, they provided some important information for hair detection and description by introducing the hair attributes, however, since the purpose of their work is human identification, the images in their experiments are all human frontal face images with hair regions.

Dass et. al 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 clustering of hair regions, where they do not need to assume any pre-determined number of clusters. For each hair-style 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 the hairstyle of

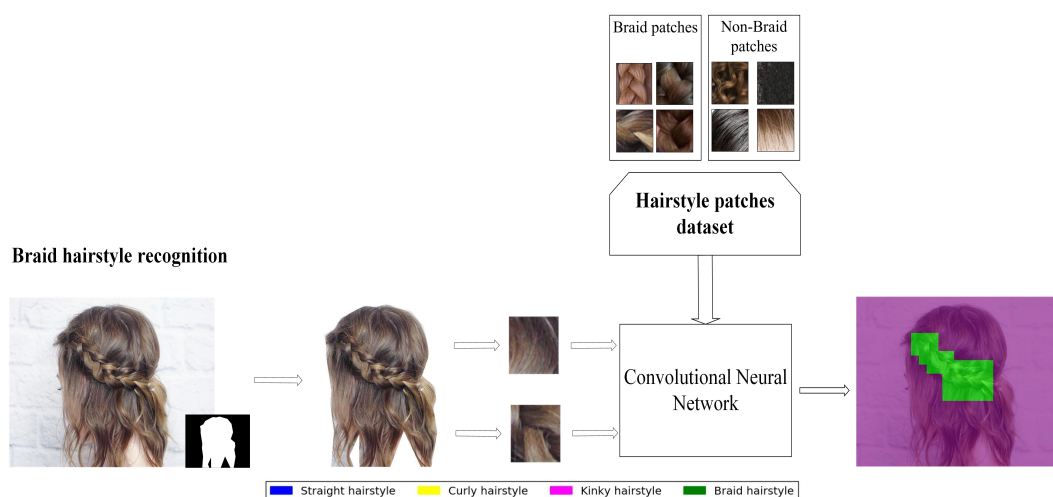


Figure 3: Braid hairstyle recognition system overview.

a person. The hair distribution of a person is compared with the templates to recognize hairstyles. In their experiments, they collected male and female 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% (Dass et al., 2013). All their experiments are based on frontal face images and unsupervised learning algorithm (clustering) and they focus on the recognition of complete hair regions from face regions rather than the detailed partition of different hairstyles inside the hair region. To sum up, research works on more complicated hairstyles, especially from the back-views images or side-views images, have not been explored. Furthermore, supervised learning algorithm can also be used in hairstyle recognition in order to provide high-level features and more reliable recognition results.

2.2 3D Braid Modelling

In the area of hair modelling, research obtained the 3D braid models by fitting the captured hair braid 3D point cloud with the pre-generated 3D braid models. Hu et.al propose a data-driven method to automatically reconstruct braided hairstyles from input data obtained from RGB-D camera. They produced a database of 3D braid patches and use a robust random sampling approach for data fitting. The experiment results demonstrated that using a simple equipment is sufficient to effectively capture a wide range of braids with distinct shapes and structures (Hu et al., 2014).

2.3 Materials Recognition

Research works on material recognition usually applied hand-designed image features to classify different materials.

Liu et. al proposed an augmented Latent Dirichlet Allocation (aLDA) model to combine the rich set of low and mid-level features under a Bayesian generative framework and learn an optimal combination of features. Experimental results show that the system performs material recognition reasonably well on a challenging material database, outperforming state-of-the-art material/texture recognition systems (Liu et al., 2010).

Hu et. al empirically study material recognition of real-world objects using a rich set of local features. They applied the Kernel Descriptor framework and extend the set of descriptors to include material-motivated attributes using variances of gradient orientation and magnitude. Large-Margin Nearest Neighbor learning is used for a 30-fold dimension reduction. They also introduce two new datasets using ImageNet and macro photos (Hu et al., 2011).

Qi et.al introduced the Pairwise Transform Invariance (PTI) principle, and then proposed a novel Pairwise Rotation Invariant Co-occurrence Local Binary Pattern (PRICoLBP) feature, and further extend it to incorporate multi-scale, multi-orientation, and multi-channel information. The experiments demonstrated that PRICoLBP is efficient, effective, and of a well-balanced tradeoff between the discriminative power and robustness (Qi et al., 2014).

Cimpoi et.al identified a rich vocabulary of forty-seven texture terms and use them to describe a large dataset of patterns collected in the wild. The result-

ing Describable Textures Dataset (DTD) is the basis to seek for the best texture representation for recognizing describable texture attributes in images. They applied the Improved Fisher Vector (IFV) to texture recognition. The experiment results showed that their method outperformed other specialized texture descriptors in established material recognition datasets (FMD and KTHTIPS-2) benchmarks(Cimpoi et al., 2013).

Bell et.al introduced a new, large-scale, open dataset of materials in the wild, the Materials in Context Database (MINC), and combine this dataset with deep learning to achieve material recognition and segmentation of images in the wild. For material classification on MINC, they achieved 85.2% mean class accuracy. They combined these trained CNN classifiers with a fully connected conditional random field (CRF) to predict the material at every pixel in an image and achieving 73.1% mean class accuracy(Bell et al., 2014).

The differences between material recognition and the hairstyle recognition are:

- The material recognition emphasizes on the recognition of different classes of materials. For example, they want to tell the hair from the skin. Thus, it is a inter-class recognition problem.
- Our braid hairstyle recognition focus on distinguishing hairstyle structures inside the hair class. The differences between hairstyles are caused not only by the characteristics of different hair fibres, but also by the structures that the hair strands formed. Thus, it is more like a intra-class recognition problem.

2.4 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) have been widely adopted in classification and segmentation tasks, including object recognition (Krizhevsky et al., 2012), hair region detection(Chai et al., 2016), and demonstrated to provide superior performance than traditional classification and segmentation systems. CNNs usually require a large amount of training data in order to reach the best performance and avoid overfitting. However, for our braid hairstyle detection and recognition system, only a small amount of training data is available. In order to avoid overfitting, we trained our CNN on a larger data set from a related domain (ImageNet). Trained on large dataset, the CNN can learned useful features and leverage such features to reach a better accuracy than other methods that rely on the small dataset. We perform an additional training step using our own data to fine-tune the trained

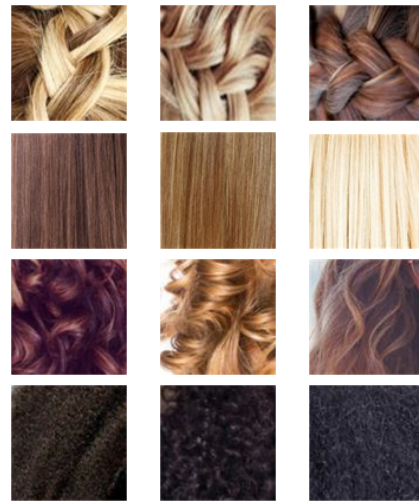


Figure 4: Hairstyle patches.

network weights. The model in our system is the Inception V3 network with a final layer retrained on our own hairstyle patch dataset.

3 BRAID HAIRSTYLE RECOGNITION

The overview of the braid hairstyle recognition system is shown in Figure 3.

All hairstyle images used in our system are downloaded via Internet. The hairstyle images contain different hair colors, lengths, and volumes, etc. In addition, those hairstyle images are captured from different point of views, including front-view hairstyle image, side-view hairstyle images, and back-view hairstyle images. Moreover, we avoid very small-size hairstyle images since the quality is relative low and the details of the hair structure tend to be vague. We also reduce the size of very high resolution hairstyle images. Thus the average width of the hair region is in the range of 450 pixels to 600 pixels.

The hairstyle images are then separated into the following two categories. Noting there is no overlapping between those two sets.

- Dataset-I: Hair images that needed to be cropped into hairstyle patches to form the dataset for training the hairstyle recognition system.
- Dataset-II: Hair images that used to perform the full-image hairstyle recognition.

3.1 Training Procedure

In order to prepare the hairstyle patches dataset for training the braid hairstyle recognition model, we manually crop hairstyle patches from Dataset-I and label them. During the cropping procedure, we need to control the size of the cropping window in order to reserve the distinguish structures of the braid hairstyles. Given the characteristics of the braid hairstyle, if the cropping windows are very small, then the image patches will lose the ability to represent the unique interlacing structure and every image patch will look like the straight hairstyle. On the other hand, if the cropping window is very large, it may contain several different hairstyles and make the recognition difficult. Thus, instead of using a fixed-size window for hairstyle patch cropping, we made the size of the cropping window adjustable in order to capture the unique braid structure.

After the cropping procedure, we adjust the size of each hairstyle image patch into $50 \text{ pixels} \times 50 \text{ pixels}$. The hairstyle patch samples are shown in Figure 4. The first row shows the braid hairstyle patches, the rest are the non-braid hairstyle patches, including: the straight hairstyle patches (the second row), the curly hairstyle patches (the third row), and the kinky hairstyle patches (the last row).

Then we separate all the hairstyle patches into the training dataset and the testing dataset to train the braid hairstyle recognition model. The details of the training dataset and testing datasets are shown in Table 1.

As shown in Table 1, our training and testing datasets only contain a small amount of hairstyle patches. In order to prevent over-fitting and help the hairstyle recognition model generalize better, we need to make the most of our few training examples by "augmenting" the hairstyle image patches via a number of random transformations, including: rotation, vertical shift, horizontal shift, shearing transformation, and horizontal-flip. The hairstyle patch augment results are shown in Figure 5, the first row is the original braid patch. The second to sixth rows are augmented braid patches. We notice that the augment procedure reserves the basic structure of the braid, it also increases the diversity of the braid by changing the direction of the braid, modifying the width of the braid, etc. Since all the augmented patches can be found in real-world hairstyles, thus the augmented results are reasonable. During the training stage, we apply the random transformations and normalization operations on our hairstyle image patch dataset and generate augmented hairstyle image patches and their corresponding labels.

Table 1: Hairstyle patch dataset.

Hairstyle	Index	# Training	# Testing
Braid	1	800	200
Straight	2	800	200
Curly	3	800	200
Kinky	4	800	200

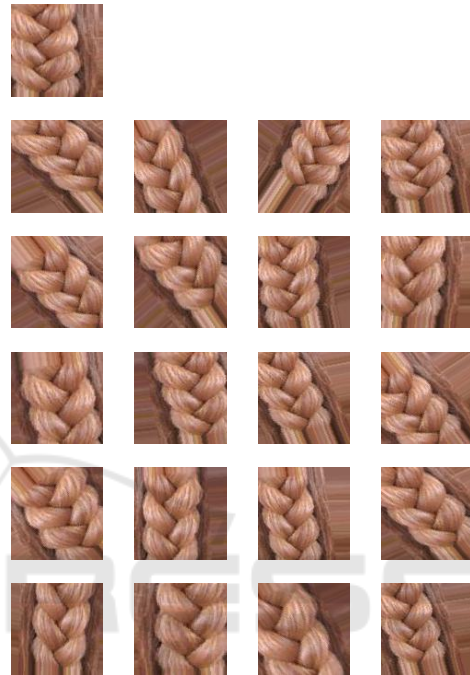


Figure 5: Data augmentation results (braid hairstyle).

After obtain the hairstyle patch dataset, we applied the Inception v3 network (Szegedy et al., 2016) with a final layer retrained on it. The original Inception v3 network is trained on ImageNet (Deng et al., 2009), which provides enough knowledge of real-world objects. We add a final layer retrained to our own hairstyle dataset to learn features for different hairstyles. Our hairstyle recognition system reaches the accuracy of 92.7%.

3.2 Full-image based Braid Hairstyle Recognition

3.2.1 Hair Region Mask Generation

During the procedure for full image braid hairstyle detection and recognition, the input images of our system are selected from the Dataset-II that mentioned before. Those hair images contain both hair regions and non-hair regions (e.g. faces, backgrounds, etc). We manually select points on the boundary of the hair

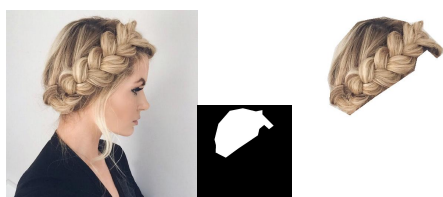


Figure 6: Hair region mask.

region to generate the hair mask and obtain the hair region, the results as shown in Figure 6.

We apply the sliding window method inside the hair region. The size of the sliding window is W pixels \times H pixels (e.g. $W = 50$ pixels and $H = 50$ pixels). The stride of the window is S pixels (e.g. $S = 15$ pixels). Then we can obtain the hairstyle prediction for every window patch.

For each hairstyle patch $patch_i$, the braid hairstyle recognition system will provide the class labels and the corresponding scores ($label_n, score_n$). Noting that n indicates the label index in Table 1 and the scores satisfy $\sum_{n=1}^4 score_n = 1$. Although our system aims to detect and recognize braid hairstyle, we keep all the labels and scores for different hairstyles. However, there are overlapping regions between the adjacent windows, the scores and labels updating procedure is shown in Figure 7. The red window indicates the original patch, the blue window and the green window indicate current patch when the sliding window moves 15 pixels horizontally and 15 pixels vertically, respectively. We compare the scores of the original score with the current score of the overlapping region. If the current score (0.994485) is less than the original score (0.996278), we keep the original label (*straight*) and score (0.996278) for the overlapping part, otherwise, we update the score (0.998383) and the corresponding label (*straight*) according to the score and label of the current window.

After the score and label updating procedure, we compare the score of each pixel with the predefined threshold value $threshold (= 0.88)$, if the score is larger than the threshold value, we accept the recognition result. Otherwise, we reject the recognition result.

3.3 Experiment Results

We conduct experiments on full hair images that selected from Dataset-II.

As shown in Figure 9, our system can detect braid region in full hair image. The first column shows the original full hair images, the second column shows the hair region mask, the third column shows the hair region images, the last column shows the braid hairstyle recognition results inside the hair regions.

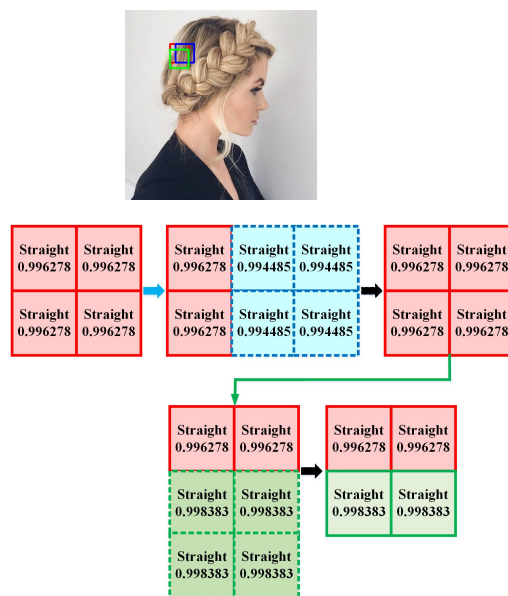


Figure 7: Hairstyle label and score update.

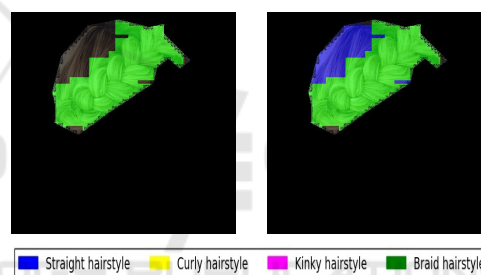


Figure 8: The braid hairstyle recognition results.

The braid hairstyle regions are highlighted with color green.

In the first row of Figure 9, the size of the full hair image is 458 pixels \times 504 pixels. The size of the sliding window is 50 pixels \times 50 pixels, the stride of the sliding window is 25 pixels.

In the second row of Figure 9, the size of the full hair image is 517 pixels \times 678 pixels. The size of the sliding window is 50 pixels \times 50 pixels, the stride of the sliding window is 25 pixels. There are mainly three curly hairstyle patches are recognized as braid hairstyle, as shown in Figure 10. The patches contain the patterns that are very similar to the strands interlacing structure of the braid hairstyle. The results indicate that the braid hairstyle recognition is relatively more difficult than the recognition of other hairstyles.

In the third row of Figure 9, the size of the full hair image is 653 pixels \times 1129 pixels. The size of the sliding window is 60 pixels \times 60 pixels, the stride of the sliding window is 30 pixels. Since the braid in this hair image is simpler than other full hair images,

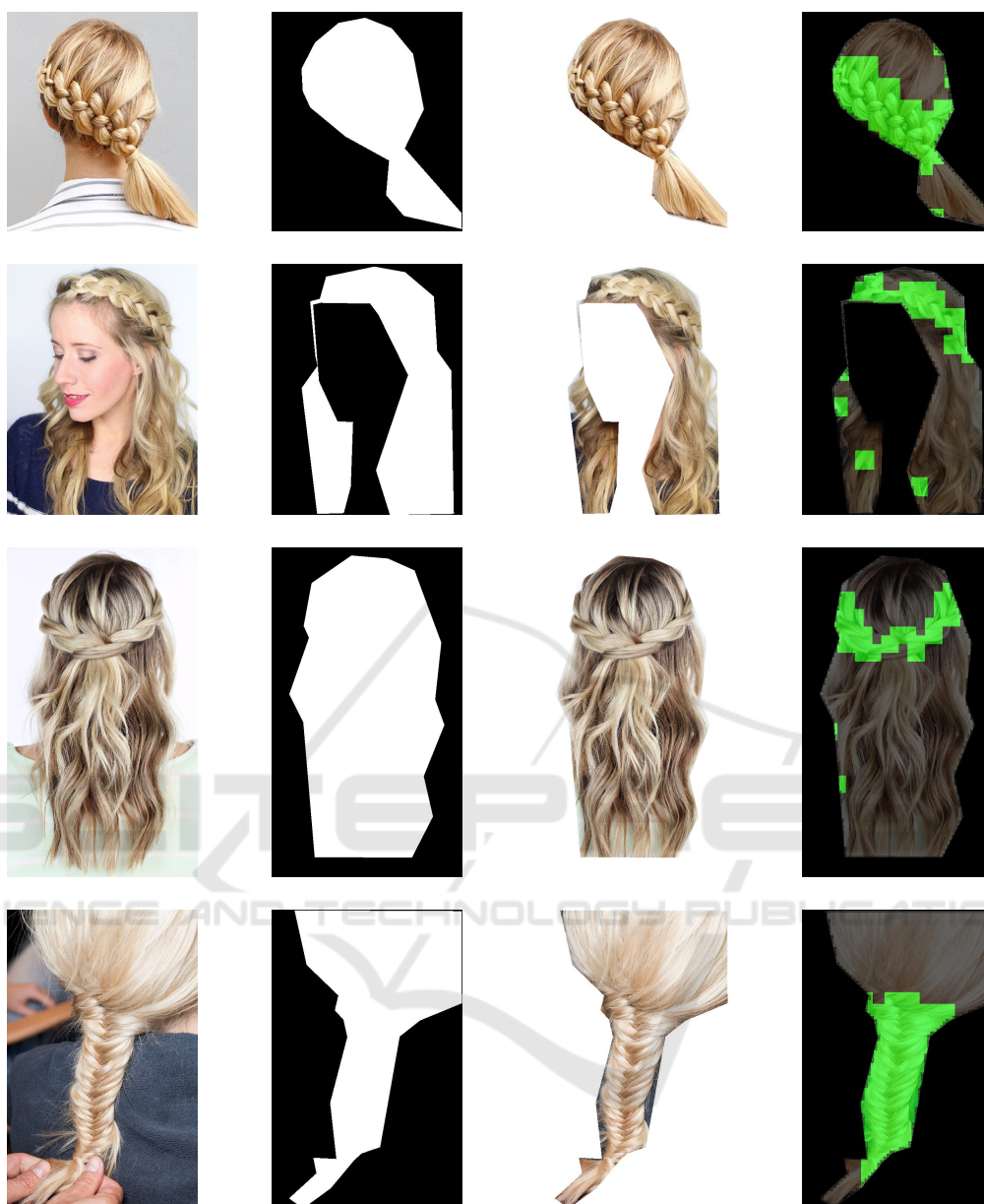


Figure 9: Braid hairstyles recognition results.

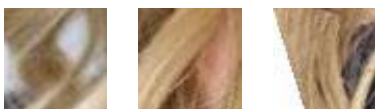


Figure 10: Mis-classified hair patches.

a slightly large sliding window will contain more information for braid recognition.

The "fishtail" braid recognition results is shown in the fourth row of Figure 9, the size of the full hair image is 488 pixels × 763 pixels. The size of the sliding window is 60 pixels × 60 pixels, the stride of

the sliding window is 30 pixels.

The experiment results indicate that the braid hairstyle recognition system can successfully recognize braid hairstyle in full hair images.

4 CONCLUSIONS AND FUTURE WORKS

In this paper, we present a novel braid hairstyle recognition system. We leverage the power of the

pre-trained Convolutional Neural Networks to learn the features of braid hairstyle as well as non-braid hairstyles. However, due to our small-scale dataset, data augment techniques and transfer learning are applied to deal with the problem of overfitting. The experiment results show that our system is capable to recognize four basic hairstyles, including braid hairstyle, straight hairstyle, curly hairstyle, and kinky hairstyle, however, we focus on recognize braid hairstyle in this paper. Moreover, the strategy of training on patch-level and performing recognition on image-level can facilitate the recognition procedure for complex hairstyles. In addition, since the system is based on image patches, it can be used to recognize hairstyle not only in the front-view hair images, but also in the side-view hair images, as well as the back-view hair image.

In the future, we need to increase our data to include more braid hairstyles. Furthermore, we need include the spacial information as the global information in order to eliminate mis-classified patches.

REFERENCES

- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Susstrunk, S. (2012). Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.*
- Bell, S., Upchurch, P., Snavely, N., and Bala, K. (2014). Material recognition in the wild with the materials in context database. *CoRR*, abs/1412.0623.
- Chai, M., Shao, T., Wu, H., Weng, Y., and Zhou, K. (2016). Autohair: Fully automatic hair modeling from a single image. *ACM Trans. Graph.*
- Cimpoi, M., Maji, S., Kokkinos, I., Mohamed, S., and Vedaldi, A. (2013). Describing textures in the wild. *CoRR*, abs/1311.3618.
- Dass, J., Sharma, M., Hassan, E., and Ghosh, H. (2013). A density based method for automatic hairstyle discovery and recognition. In *Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2013 Fourth National Conference on.*
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*.
- Hu, D., Bo, L., and Ren, X. (2011). Toward robust material recognition for everyday objects. In *BMVC*, pages 1–11.
- Hu, L., Ma, C., Luo, L., Wei, L.-Y., and Li, H. (2014). Capturing braided hairstyles. *ACM Trans. Graph.*
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*.
- Liu, C., Sharan, L., Adelson, E. H., and Rosenholtz, R. (2010). Exploring features in a bayesian framework for material recognition. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 239–246.
- Qi, X., Xiao, R., Li, C. G., Qiao, Y., Guo, J., and Tang, X. (2014). Pairwise rotation invariant co-occurrence local binary pattern. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(11):2199–2213.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*.
- Yacoob, Y. and Davis, L. S. (2006). Detection and analysis of hair. *IEEE Trans. Pattern Anal. Mach. Intell.*