

Dataset Enhancement in Hair Follicle Detection: ESENSEI Challenge

Jan Jakubik

Wrocław University of Science and Technology

Faculty of Computer Science and Management, Department of Computational Intelligence

Wrocław, Poland

Email: jan.jakubik@pwr.edu.pl

Abstract—In this paper, a solution to ESENSEI data mining challenge concerning the analysis of microscopic hair images is described. The task of the challenge was to detect locations of hair follicles in closeup images of a human scalp. The proposed solution is based on a convolutional neural network architecture. To improve generalization performance, we enhance training and test datasets using image transformations applied to both input and output. The chosen transformations are two axis symmetries and switching axes, all of which are possible to apply regardless of resolution without producing interpolation artifacts. Since these can be combined, $2^3 = 8$ possible views of each image can be created to expand both training and test data. We demonstrate the effects of dataset enhancement in both training and classifying on results achievable on the competition dataset. The solution placed 2nd in the final challenge evaluation.

I. INTRODUCTION

IN RECENT years, convolutional neural networks (CNN) have become the standard for machine learning tasks concerning image analysis. Following the introduction [1], breakthrough results in recognition of handwritten digits [2], and significant speed improvements achieved with GPU computing [3], over the past decade the deep learning approach has been widely recognized as the current state of the art in image classification tasks. The capability of convolutional networks to perform well on raw data has freed machine learning researchers of the need to design features with specific invariance properties that are expected in image analysis (invariance to scaling, rotations, transposition). Simultaneously, developments in areas of GPU processing and big data made it more feasible to deploy these complex architectures, requiring large amounts of training data, on real datasets for practical problems.

The weaknesses of convolutional networks related to the complexity of the models have been since an object of attention. In widely publicised result presented in [4], an observation was made that it is possible to completely "confuse" neural networks by altering images in ways imperceptible to humans. This has shown that the CNN's generalization ability still does not directly correspond to our perception of visual data. One possible approach to lessen this problem comes in form of data enhancement: creation of additional samples in the dataset that cover potential cases in which the network could not otherwise learn by generalizing training data. These

can include noise disturbances, adversarial samples, and transformations such as rotation by few degrees or scaling.

In this paper, we describe a CNN-based solution developed during the ESENSEI data mining competition [5] to the task of detecting hair follicles in microscopy images. The crucial insight that allowed the proposed approach to place 2nd in the final evaluation is that expanding the dataset by rotation and symmetrical reflections can significantly improve results. However, due to the nature of the data (specifically, the output being a low-resolution binary image), enhancing the data using transformations that require interpolation does not produce good results. Therefore, the final submission limited transformations to combinations of three different ones that do not require interpolation.

The paper is organised as follows: in section II, challenge data is described and analysed. Section III outlines the implemented approach: preprocessing applied to the data, convolutional network architecture and data enhancement. In section IV we experimentally demonstrate the influence on data enhancement on results. Section V summarizes the conclusions.

II. CHALLENGE DESCRIPTION

The task of ESENSEI data mining challenge was to predict the locations of hair follicles on microscopic images. The dataset consists of 1920x1080 images in color. Fig. 1 shows sample images from the dataset, demonstrating the significant variability in hue, lighting, hair color and the number of follicles on each image. The dataset has 4880 images for training and 1000 for testing. During the competition, leaderboard evaluation was based on 10% of the test set.

Hair follicles are marked on a 16x9 grid. Output variables are binary, i.e., either the 144x144 square corresponding to the position on the grid contains a follicle or does not. The submissions format was the same as annotations, i.e., binary values on a 16x9 grid. These were evaluated using F-score as a metric, with F-score calculated for each image separately and then averaged over the dataset.

III. METHOD DESCRIPTION

The proposed approach uses a fairly standard convolutional neural network architecture with downscaled images as an

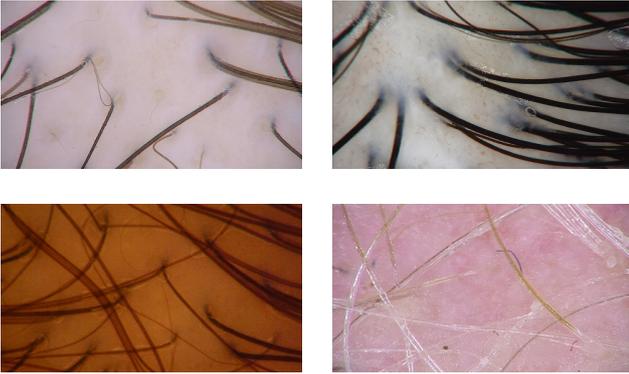


Fig. 1: Sample images from training data with significant differences in hue, lighting, hair color and hair density

input. As CNNs are known to function well with raw data, we strived to minimize the preprocessing steps.

A. Data preprocessing

All images are first resized to 1/8 of the original height and width. This partially removes significant noise present in the challenge images, while making processing of the images significantly less resource-heavy. Then, simple normalization is applied by first scaling all channels to (0,1) interval by dividing them by 255, and then transposing each channel to zero mean. This is done for each image separately to alleviate differences in hue and lighting between images.

B. Network Architecture

The neural network employed for follicle detection consists of convolutional layers and pooling only. Each convolutional layer utilizes ReLU activation and 3x3 filters. Layer size was set to 40 for all layers. Full network is shown in Fig. 2. A single residual connection [6] between layers is used, adding the output of the 1st layer to the output of layer 4 as the input of layer 5. This allows faster training given the model depth. We built this architecture through iterative deepening, i.e., we started with 6 layers of convolution with 2 pooling layers in between, and then added new layers as long as performance improvements were seen on both validation set during training, as well as leaderboard after uploading test set results. The residual connection was added when problems in propagating gradients first appeared at 10 layers depth.

With two pooling layers, the image is downscaled 15 times, to 16x9, which means the network output can be compared directly with the ground truth grid. As the problem is effectively mapping images onto images, we did not find it necessary to use any fully connected layers on top of the network. For optimization objective, we chose standard Mean Square Error loss. The network weights are optimised with Adadelta adaptive gradient descent method [7].

As the output is real-valued, thresholding must be applied to obtain locations marked as follicles. The threshold that has to be exceeded in order to mark a particular pixel in the 16x9

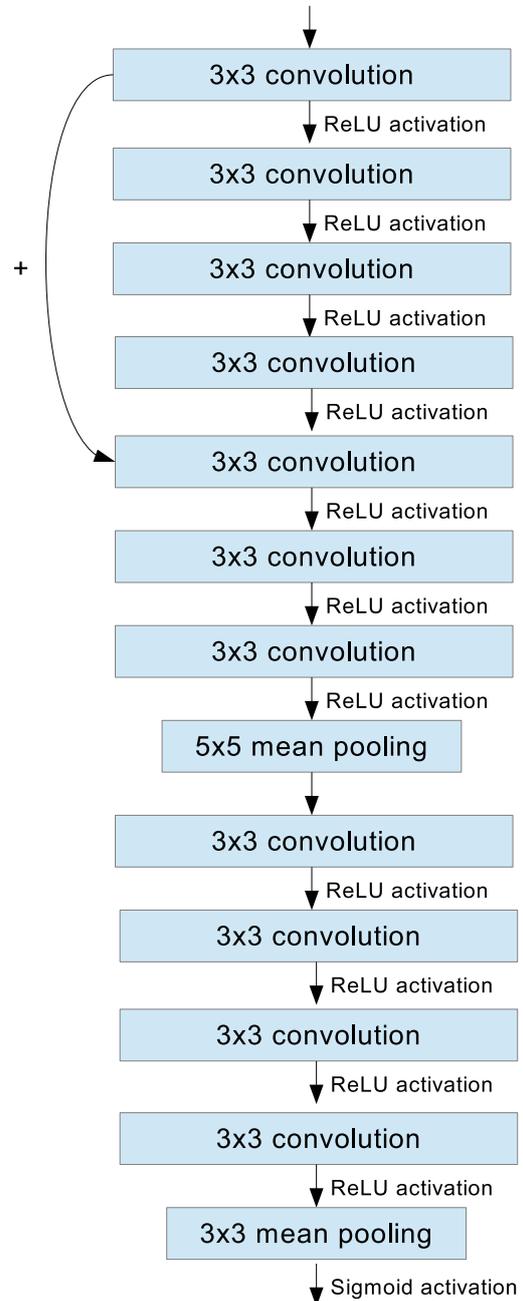


Fig. 2: CNN architecture for the final submission. Plus sign denotes a residual connection.

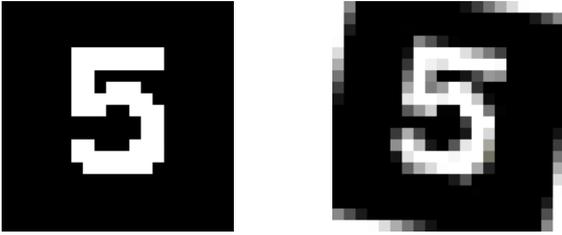


Fig. 3: 20x20 resolution picture and its rotation by 10 degrees. Interpolation results are clearly seen and will affect training of a CNN.

output as a follicle is chosen based on validation set results. We test threshold values from 0.01 to 0.5 (every multiple of 0.01) and choose the one that maximises F1 score on the validation set. The split of data into training and validation is done randomly, using 4400 images out of 4880 for training.

C. Dataset Enhancement

As convolutional networks do not possess invariance properties regarding scaling or rotation naturally and are expected to learn these from data, we utilize data enhancement to improve results.

For a model where (x, y) is a valid input-output training sample, we seek transformations T for which $(T(x), T(y))$ is also a valid input-output training sample that can be added to the data. Rotations and symmetries are among these transformations - the locations of follicles in a rotated image should be possible to obtain by rotating the follicle locations on the original image, and the same is true for symmetric reflections. However, due to the very small resolution of the output grid, interpolation artifacts produced by rotations that are not multiples of 90 degrees could influence results in a significant way. Fig. 3 presents results of rotation requiring interpolation on a 20x20 resolution picture. Due to concerns of interpolation affecting the output values, we limit the transformations to three that do not require interpolation:

- X axis symmetry
- Y axis symmetry
- switching X and Y axes

By combining these, we can obtain 8 possible transformations, including rotations by multiples of 90 degrees. These are shown in Fig. 4.

Additionally, some other enhancement transformations that could be considered are scaling and random noise distortion. However, we decided not to include these in the dataset based on our intuitions concerning the data. Enhancing the training dataset is applicable only when the enhancement improves generalization ability on the test data. For rotations and symmetries, it can be intuitively justified. As the hair on training and test images grow in different directions and at different angles, we would like training samples to cover most

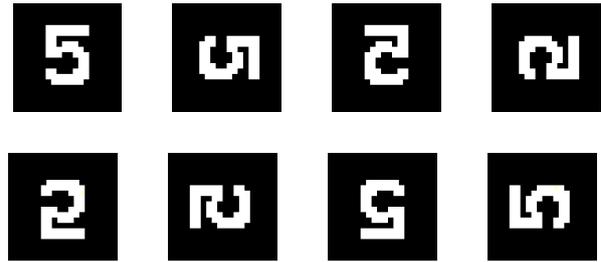


Fig. 4: Possible transformations resulting from different combinations of two axis symmetries and axis switching.

angles possible, and one way to do this is to expand the data with rotations and symmetries.

For random noise and scaling, such intuitions are not in place. Random noise distortions would be justifiable if there was a noise pattern present in test data that was missing from training data and can be replicated. We failed to identify such a pattern. Scaling, similarly to rotations, would require interpolation, with the possible exception of scaling by an integer number. However, an upscaled picture will show hair at a different zoom level than the original photos. Our knowledge of microscopic hair imaging is insufficient to tell whether widely varying zoom levels are expected between training and test data.

In practice, the dataset enhancement is implemented as follows: during training, for a particular batch, we apply each of the three listed transformations with 50% probability. During testing, all 8 possible views are considered by the same neural network, and a position on the grid is considered a follicle if in more than half of the views it was marked as a follicle.

IV. RESULTS

Fig. 5 demonstrates the performance on validation set over the duration of training. For obtaining these results, the same training-validation split was used for each experiment.

The comparison includes proposed approach applied to the training data, test data or both, as well as a baseline with no dataset enhancement. We also test enhancement with random rotations with rotation angle uniformly sampled from $(-10, 10)$ interval, applied in addition to already mentioned symmetries and axis switching. This is to test whether the problem of interpolation outweighs the potential gains from expanding the data by a wider range of rotations.

It can be seen that the biggest boost in performance is achieved by enhancing the training set. Without it, the network starts overfitting very fast. This can be seen in the graph as decreasing performance on the validation set from epoch 10 onwards. Networks trained with training data enhancement do not overfit within the 100 epochs period, but it can be seen that adding an expanded range of rotations worsens performance instead of increasing it.

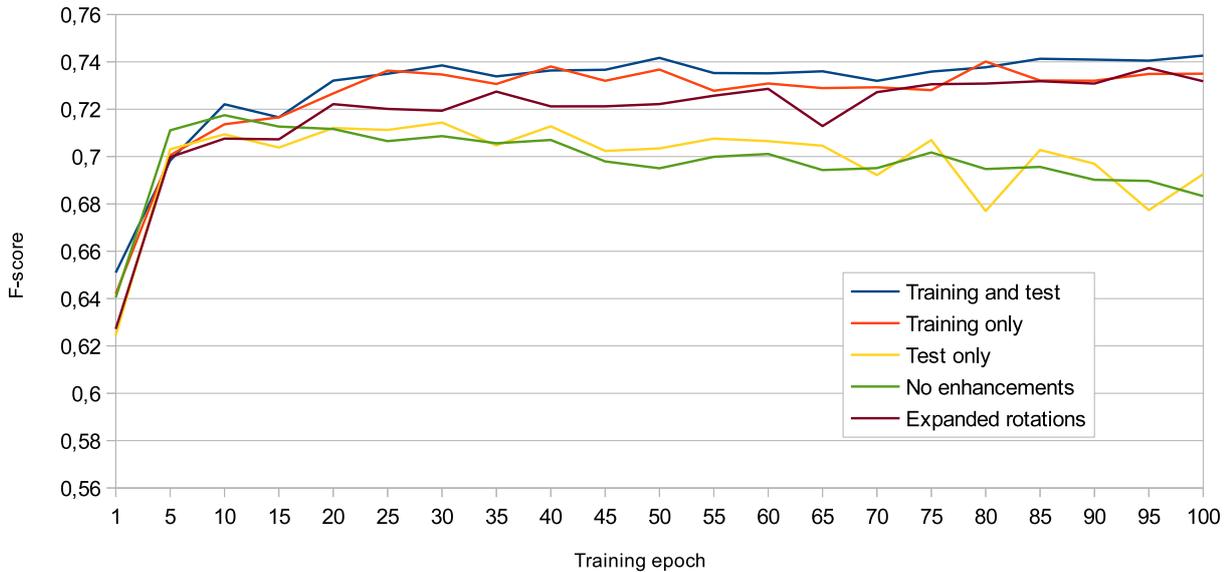


Fig. 5: Performance over the course of training, depending on where the data enhancement was applied. Expanded rotations include rotations between -10 and 10 degrees, possibly requiring interpolation.

TABLE I: Results of the competition - top submissions

Team	RMSE
rs	0.76299467
jj	0.74934757
podludek	0.71388834

Based on these results, the approach of enhancing training and test data with symmetries and axis switching was chosen for the final competition submission. Scores achieved by top 3 submissions in the competition are shown in Table 1.

V. CONCLUSIONS

We have described an approach to the task of hair follicle detection, based on a CNN trained with dataset enhancement. The proposed approach to the task has proven successful in ESENSEI data mining challenge, achieving 2nd place out of 16 submissions. Experimental results show that the good results are largely reliant on the chosen approach to enhancing the training and test data. In general, the described approach can be applied to any kind of image data, regardless of resolution, due to the fact that all applied transformations require no interpolation.

Our results showcase the importance of dataset enhancement for training CNN. However, care should always be taken when choosing specific ways to enhance data. Tests with transformations that require interpolation on the competition dataset show that the produced artifacts have a negative impact on performance as expected, likely due to the small

resolution of the desired output image. However, the observed improvement from utilizing rotations and symmetries is also an effect specific to the competition data. For hair imaging, the choice of symmetries and rotations by multiples of 90 degrees results from two basic insights. First is that we need to expand the range of angles at which hair grow in training pictures, and second is that the very low resolution of the output grid prevents us from using transformations with interpolation. In practice, any approach to enhancing a dataset should be reliant on reasonable intuitions concerning the particular type of data.

ACKNOWLEDGEMENTS

We would like to thank eSensei and Knowledge Pit for providing the data and a platform for the competition.

REFERENCES

- [1] LeCun, Yann; Leon Bottou; Yoshua Bengio; Patrick Haffner (1998). "Gradient-based learning applied to document recognition". *Proceedings of the IEEE*. 86 (11): 2278-2324
- [2] Ciresan, Dan; Meier, Ueli; Schmidhuber, Jurgen (June 2012). "Multi-column deep neural networks for image classification". 2012 IEEE Conference on Computer Vision and Pattern Recognition.
- [3] Dave Steinkraus; Patrice Simard; Ian Buck (2005). "Using GPUs for Machine Learning Algorithms". 12th International Conference on Document Analysis and Recognition (ICDAR 2005). pp. 1115-1119.
- [4] Goodfellow, Ian, Shlens, Jonathon, Szegedy, Christian. (2014). Explaining and Harnessing Adversarial Examples. arXiv 1412.6572.
- [5] <https://knowledgepit.fedcsis.org>
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, Deep Residual Learning for Image Recognition, arXiv preprint arXiv:1512.03385, 2015
- [7] M. D. Zeiler, "ADDELTA: An Adaptive Learning Rate Method", arXiv preprint arXiv:1212.5701, 2012.