1. Robert SZMURŁO[1], 2. Stanisław OSOWSKI[1,2]

Faculty of Electrical Engineering, Warsaw University of Technology (1), Electronic Faculty, Military University of Technology (2)
ORCID: 1. 0000-0001-8041-4438; 2. 0000-0003-3194-4656

doi:10.15199/48.2024.06.05

Relabeling the imperfect labeled data to improve recognition of face images using CNN

***Abstract**. The paper considers the problem of improving the recognition of not-perfectly labeled face images using CNN networks. The proposed solution is based on relabeling the samples of images by applying the KNN classification principle based on the distance between the samples. The original images are first converted to the features and the KNN principle is applied to them. The classes of sample images are relabeled according to the class represented by most neighbors indicated by KNN. The developed system was tested on the problem of face recognition. The dataset was composed of 68 classes of grayscale images. The results of experiments have shown significant improvement in the recognition rate of not perfectly labeled images.*

***Streszczenie.** W artykule poruszono problem poprawy dokładności rozpoznawania obrazów twarzy za pomocą sieci CNN, przy założeniu, że zbiór treningowy zawiera dane z błędnymi etykietami. Zaproponowane rozwiązanie polega na ponownym oznakowaniu próbek obrazów poprzez zastosowanie klasyfikacji KNN opartej na odległości pomiędzy próbkami. Na początku obrazy konwertowane są na wektor cech, do których stosowany jest algorytm KNN. Klasy przykładowych obrazów są ponownie oznaczane zgodnie z klasą reprezentowaną przez większość sąsiadów wskazanych przez KNN. Opracowany system został przetestowany w zastosowaniu do problemu rozpoznawania twarzy. Zbiór danych składał się z 68 klas obrazów w skali szarości. Wyniki eksperymentów wykazały znaczną poprawę skuteczności rozpoznawania niedoskonale oznakowanych obrazów. (**Ponowne etykietowanie niedoskonałych oznaczonych danych w celu poprawy rozpoznawania obrazów twarzy za pomocą CNN**)*

Keywords: relabeling, CNN, KNN, face recognition, feature analysis, deep networks.
Słowa kluczowe: reetykietowanie. CNN, KNN, rozpoznawanie twarzy, analiza cech, sieci głębokie.

Introduction

Automated facial recognition is an important issue in safety engineering. Different institutions use the face in a biometric authentication of a person (so-called Face ID). The police uses face images to identify suspicious individuals. Face recognition is an essential tool in checking the identity of persons in airports. Therefore, this branch of engineering is still of high interest from a research point of view.

The important difficulty in image recognition is the imperfect labeling of training data, followed by the variety of poses, bad resolution, and noisy environment caused by different data acquisition circumstances [1,2]. Many approaches have been proposed in the past to deal with the above problems [1-6]. They apply such techniques as clusterization, nearest neighbors or different approaches to deep learning [7-10].

This paper deals with a problem of wrong labelling by proposing a special technique based on relabeling the images during the preprocessing phase. First, the images are converted to vectorial forms representing their diagnostic features. This is done by applying a convolutional neural network (CNN) transforming an image into its vector form represented by a flattened layer.

The relabeling process assumes, that there is a substantial part of correctly labeled classes in the input images dataset. It means, that their vectorial forms (features) are associated with the correct classes. The suspicious samples of images (also in vectorial form) are compared to them based on the K-nearest neighbor principle (KNN). The class represented by the majority of the nearest neighbors dictate the new label for the suspected image for relabeling.

In the proposed solution, as the preprocessing step in the generation of the image features, we have applied transfer learning based on the Alexnet architecture of CNN. The Alexnet network, presented below, generates the numerical descriptors of the image in the flattened fc6 layer, treated as the features. Their size is 4096.

1 'data' Image Input 227x227x3 images with 'zerocenter' normalization
2 'conv1' Convolution 96 11x11x3 convolutions with stride [4 4] and padding [0 0]
3 'relu1' ReLU
4 'norm1' Cross Channel Normalization with 5 channels per element
5 'pool1' Max Pooling 3x3 max pooling with stride [2 2] and padding [0 0]
6 'conv2' Convolution 256 5x5x48 convolutions with stride [1 1] and padding [2 2]
7 'relu2' ReLU
8 'norm2' Cross Channel Normalization with 5 channels per element
9 'pool2' Max Pooling 3x3 max pooling with stride [2 2] and padding [0 0]
10 'conv3' Convolution 384 3x3x256 convolutions with stride [1 1] and padding [1 1]
11 'relu3' ReLU
12 'conv4' Convolution 384 3x3x192 convolutions with stride [1 1] and padding [1 1]
13 'relu4' ReLU
14 'conv5' Convolution 256 3x3x192 convolutions with stride [1 1] and padding [1 1]
15 'relu5' ReLU
16 'pool5' Max Pooling 3x3 max pooling with stride [2 2] and padding [0 0]
17 'fc6' Fully Connected 4096 signals (numerical descriptors)
18 'relu6' ReLU
19 'drop6' Dropout 50% dropout
20 'fc7' Fully Connected K neurons % K set individually by the user
21 'relu7' ReLU
22 'drop7' Dropout 50% dropout
23 'fc8' Fully Connected M neurons
24 'prob' Softmax
25 'output' Classification results of M classes

The important point in the solution is the choice of the value of the K parameter in the KNN classification process. This was solved in our approach by initial experiments, assuming different values and accepting the one leading to the best results.

The numerical experiments have been carried out on the database of face images representing 68 classes. The obtained results have confirmed a large improvement in class recognition after applying the relabeling technique in the presence of label noise on different levels in the samples.

The database used in experiments

The developed system used the database containing faces of 68 people (68 classes) all in a grayscale. Each class contained 20 representatives of the same person. The photographs were taken in different positions of the person's face, at different illumination and with different face expressions. Part of the photographs contain whole faces and some only part of them. To make the database more difficult they are differing also by the presence or absence of glasses. The size of the original images was the same and equal to 100×100 pixels. Fig. 1 presents examples of one-person images, illustrating the difficulties in the recognition process.

Fig.1. The examples of face images representing one class of data. They depict the differences in their acquisition.

The same person in the set is photographed in different poses and scales representing either a full-face or only a limited part of it. Some images show the face with glasses and some without them. The differences among the representatives of the same class are clearly visible.

To investigate the differences among the images we have calculated the average values of the structural similarity index for the images forming all analyzed classes using the function *ssim* in Matlab. Fig. 2 presents its values for all 68 classes of data. The values on the diagonal represent the self similarity index for images representing the class, and the off diagonal values the inter-class structural similarity. Its relatively small diagonal values confirm a large variety of samples forming the same class. The mean self-similarity index presented on the diagonal of the distribution is equal to 0.32 whereas the mean value of the inter-class similarity index is equal to only 0.19.

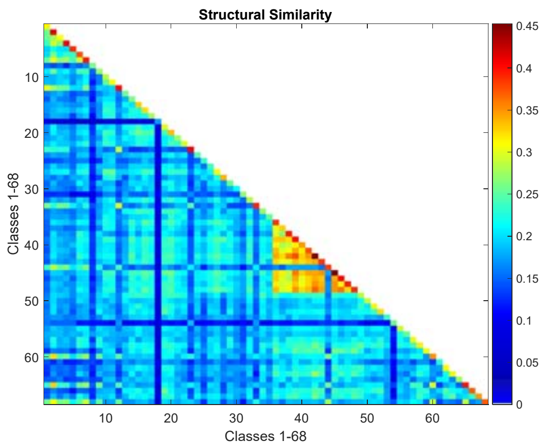


Fig. 2 The distribution of average structural similarity index values of samples belonging to 68 classes of data.

On the other side, we have observed high values of the similarity index for samples representing different classes. In an extreme case of two classes, its value was close to 0.38 (comparable to the highest value for samples of the same class which was 0.45). These numbers illustrate the additional difficulty in problem-solving.

In Table 1. Statistical measures like mean, standard deviation, energy, skewness and kurtosis for extracted features for inter-class comparison are also presented. The purpose of this research was to analyze how the dataset was difficult with respect to inter-class similarities. To prepare the comparison firstly images with similar poses, like eyes direction and face expression were selected for each class. Then, for the selected subset of 68 total images, the features were extracted with Alexnet convolutional layers. The features were then subjected to statistical analysis.

In table 1 the classes 46 and 49 present significantly similar parameters and the similarity can be observed between them in Figure 3a. In contrary, the classes 2 and 20 present different all parameters except the Kurtosis. The evident difference in classes can be observed in Figure 3b visually.

Table 1. Statistical comparison of features for similar and different classes

Class	Mean	Std	Energy	Skewness	Kurtosis
46	-0,139	235,443	0,195	0,313	2,964
49	-0,147	246,909	0,196	0,257	2,996
2	-0,167	264,894	0,192	0,338	3,116
20	-0,201	384,653	0,232	0,255	3,050

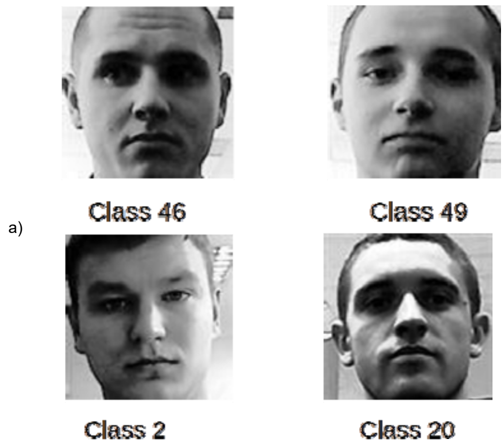


Fig. 3 a) Similar and b) different classes images.

Summing up, the section describing the dataset clearly proves that the dataset used for testing the proposed solution is not trivial.

The proposed solution

The proposed solution applies the CNN layers for automatic extracting features. It is used for two times. Firstly, it is applied for preprocessing of the data during relabeling phase and then for the second time as a part of the final class recognition pipeline.

During preprocessing the results of the CNN in the form of a flattened vector are cooperating with the KNN classifier in relabeling the data. The KNN classifier using the Euclidean distance metric:

$$D(\mathbf{x}_i, \mathbf{x}_j) = \sum_{l=1}^{N} \left| x_{il} - x_{jl} \right| \quad , \tag{1}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are vectors of features for i and j images. The parameter K ranging from 4 to 5 is used to find majority of classes for each sample image. Then a new class label is assigned to each image based on the assigned by KNN model.

The relabeled images are then delivered again to the full CNN working as the final class recognition unit. The general scheme of the proposed solution is presented in Fig. 4. The procedure contains the following phases: CNN preprocessing, KNN model training, relabeling of data and the final CNN classifier.

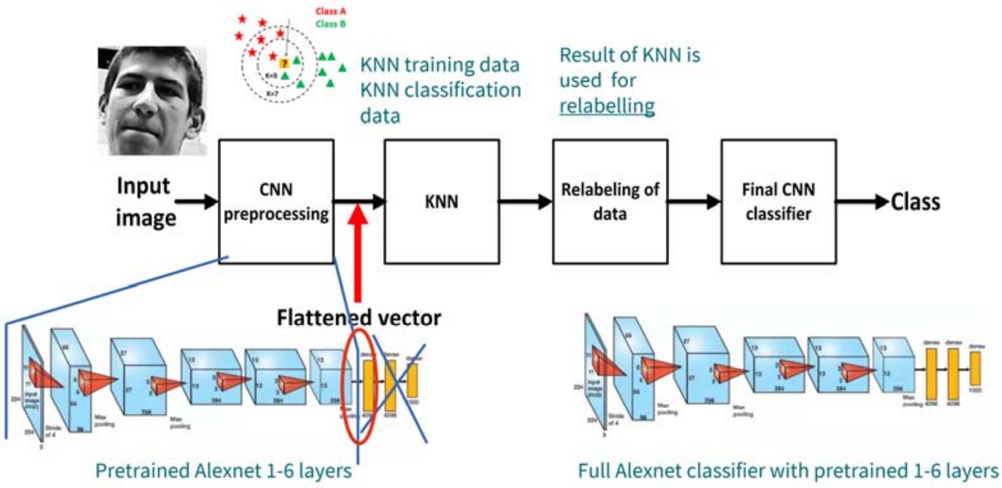


Fig.4. The general structure of the relabeling system used in image recognition.

In the practical implementation of the system, we have applied a pretrained Alexnet as the CNN structure, although any other CNN architecture can be used. The cross-entropy in objective function definition has been applied. The learning algorithm used in experiments was stochastic gradient descent with momentum (SGDM).

Results of numerical experiments

The numerical experiments have been carried out using the database presented in section II. The samples of all classes have been split into two parts: one subset of correctly defined true class membership and the remaining randomly changed labels. The rate between sizes wrong and correct labels subsets was changed from 5% to 50%.

In the first test, the contents of the raster images were original (without any artificial noise) and only the wrong labeling was inserted. The graphical results comparing the accuracy of class recognition by the system without relabeling and after application of the proposed method of relabeling is presented in Fig. 5.

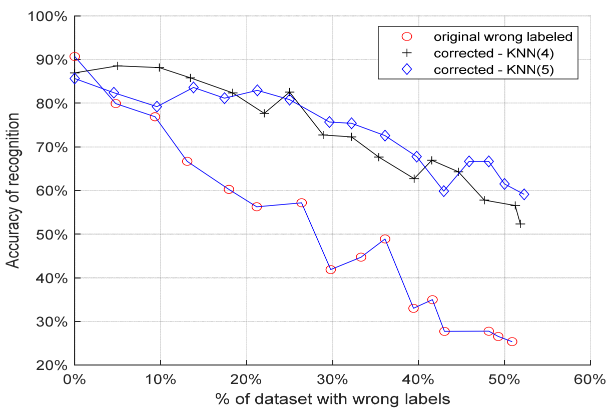


Fig. 5 The comparison of the accuracy of class recognition for the data set containing the wrong labeling of samples (no noise present in input samples). The horizontal axis represents the ratio of wrongly labeled samples.

Two different numbers of neighbors in KNN are shown: K equal to 4 and to 5. These two values have been found as the most successful in the ranges of the ratio of label distortion. A significant improvement in system performance is observed, especially at a high ratio of labeling distortion (above 20%). However, one can observe slightly worse results of the test (with KNN corrected labels) compared to the original (uncorrected) dataset when there were no wrong labels (ratio 0%). In this case, when the ratio of wrong labels in the dataset was lower than 5% the drop in accuracy was between 5% and 10%. This indicates that the KNN correction procedure is not ideal and deteriorates the results for perfectly correctly labeled datasets.

The next experiments included some artificial noise introduced to the input raster images. The samples with true labels were not distorted, and the Gaussian white noise of the variance 0.01 was applied to the images. The MATLAB function *imnoise* with 0.01 variance was used in this experiment. In Fig. 6 three samples of images with and without distortion are presented. We can notice that distortions are significant and observable in the images.

Two sets of experiments have been done. In the first case, the Gaussian noise was affecting all samples, including true labeled data and in the second case only the wrong-labeled samples were distorted. The accuracy of class recognition without and with relabeling in these two cases is presented in Fig. 7.

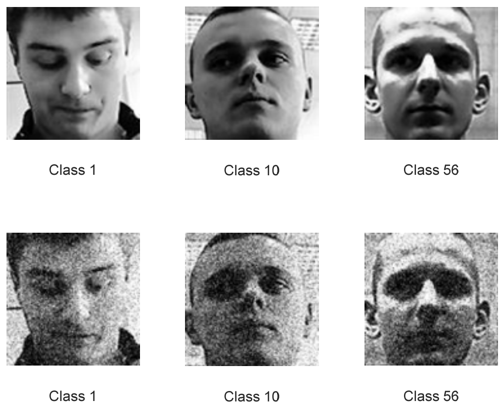
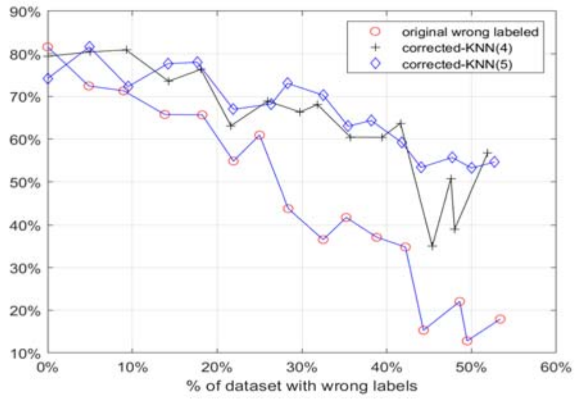
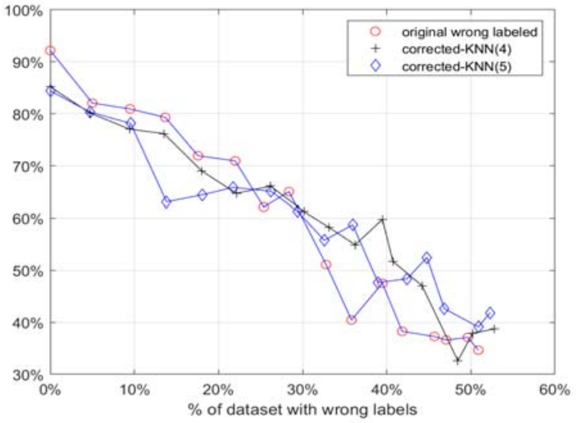


Fig. 6 Examples of face images with and without noise.

In Figure 7a we can observe a general small decrease in accuracy compared to the original (without noise) samples equal approximately to 7%. This behavior is expected for samples including noise. However, the proposed in this paper approach to using the KNN procedure for correcting labels shows acceptable resilience to the noise, as the trends of the accuracy in Figure 7a are the same as in the original (without noise) samples in Figure 5. Thus, for the ratio of wrong labels greater than 10% we can observe a significant improvement in the accuracy.



a)



b)

Fig. 7 The comparison of the accuracy of class recognition for the data set containing the wrong labeling of samples versus the ratio of wrong labeling. Fig. 6a corresponds to the case when the noise is added to the whole database, and 6b to the presence of noise only in wrong labeled samples.

It can also be noted that the parameter k in the correction procedure plays an important role in cases with wrong label ratios greater than 40% (see Figure 7a). However, in the second case when the noise was added to only the samples with the wrong labels the proposed procedure shows no improvement. This is in the authors' opinion the expected observation and is because of the distance calculation of the KNN algorithm between the features of the CNN layer. It appears that the distances between the noised suspicious samples are not close to the clean data samples with the good labels, and they can not be corrected.

Conclusions

The paper has presented a method based on the KNN principle for improving the recognition accuracy of the classification system when some input data samples are wrongly labeled. The proposed solution was tested on the dataset of images used for face recognition applying the CNN classifier.

The key point in this approach is the application of the KNN classifier. The original images are first converted to the features with help of the pretrained CNN layers. The KNN is applied to these features, looking for the nearest neighbors representing the majority class. This class is used for relabeling the questioned input image.

The results of experiments performed on the non-trivial database containing 68 classes have shown significant improvement in the recognition accuracy of not perfectly labeled images. Additionally, the experiments have shown the resilience of the proposed correction approach to noise included in all images. However the proposed method is not suggested to be used for datasets with expected high quality of labelling (below 5% of wrongly labeled samples).

The source code for the experiments carried out during the preparation of the paper is publicly available under the link: https://github.com/szmurlor/cpee_2023_relabelling.

***Authors**: prof. dr hab. inż. Stanisław Osowski, Politechnika Warszawska, Instytut Elektrotechniki Teoretycznej i Systemów Informacyjno-Pomiarowych, ul. Koszykowa 75, 00-661 Warszawa, E-mail: stanislaw.osowski@pw.edu.pl; dr inż. Robert Szmurło, Politechnika Warszawska, Instytut Elektrotechniki Teoretycznej i Systemów Informacyjno-Pomiarowych, ul. Koszykowa 75, 00-661 Warszawa, E-mail: robert.szmurlo@pw.edu.pl.*

REFERENCES

[1] B. Frenay, M. Verleysen, Classification in the presence of label noise: a survey, IEEE Trans. on NN and Learn. Sys., 25, 845–869 (2014)
[2] T. Cannings, Y. fan, R. J. Samworth, Classification with imperfect training labels, arXiv:1805.11505 [math ST], https://doi.org/10.48550/arXiv.1805.11505 (2019)
[3] J. Bootkrajang, A generalised label noise model for classification in the presence of annotation errors. Neurocomputing, 192, 61–71 (2016)
[4] N. Natarajan, I. S. Dhillon, P. K. Ravikumar, A.Tewari, Learning with noisy labels. NIPS, 26, 1196–1204 (2013)
[5] Rolnick, D., Veit, A., Belongie, S. & Shavit, N. Deep learning is robust to massive label noise. ArXiv e-prints, 1705.10694 (2017).
[6] P. Hall, B. U. Park, R. J. Samworth, Choice of neighbour order in nearest-neighbour classification. Ann. Statist., 36, 2135–2152 (2008)
[7] Patrini, G., Rozza, A., Menon, A. K., Nock, R. & Qu, L. Making deep neural networks robust to label noise: a loss correction approach. IEEE Conf. Comp. Vis. & Patt. Recog.,1944–1952 (2017).
[8] K. Siwek, S. Osowski, Deep neural networks and classical approach to face recognition – comparative analysis, Przegląd Elektrotechniczny, 94(4), 1-4 (2018).
[9] I. Goodfellow, Y. Bengio and A Courville, Deep learning, MIT Press, 2016.
[10] Krizhevsky, I. Sutskever, G.E. Hinton, “ImageNet classification with deep convolutional neural networks” Communications of the ACM, 60(6) 84–90 (2014).