COLLABORATIVE FACIAL COLOR FEATURE LEARNING OF MULTIPLE COLOR SPACES FOR FACE RECOGNITION

Hyung-Il Kim and Yong Man Ro[†]

Image and Video Systems Lab, School of Electrical Engineering, KAIST, Republic of Korea

ABSTRACT

Facial color is known as playing an important role in face recognition. Color face recognition has been investigated in the last decade. Recently, deep learning has attracted considerable attention due to their high performance in face recognition. The importance of the color in a deep learning framework is not fully investigated yet. In this paper, we have conducted experiments to investigate the effectiveness of facial color in face recognition with deep learning. Through experimental results, we have demonstrated that facial color is helpful for enhancing the recognition performance in deep learning and color space selection is crucial to achieve high performance. Moreover, by fusing features from multiple color spaces, the face recognition accuracy has been considerably improved.

Index Terms— Color face recognition, color spaces, deep learning, feature fusion, complementary effect

1. INTRODUCTION

Facial color is known to be important in face recognition (FR), which provides complementary information resided in different color spaces (Fig. 1)[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. In particular, FR accuracy with facial color information could be significantly improved when face images are degraded by illumination variations and low resolution problems [1, 2]. Many facial color-based FR frameworks have been investigated in the last decade. The authors in [3, 4] extended the existing FR framework for grayscale face to color FR framework (e.g., color texture feature analysis [3, 4]). In [5], to fully utilize complementary information in color face image, local color vector binary patterns (LCVBPs) were proposed to encode discriminating texture patterns derived from spatial interactions among different spectral-band images [5]. To obtain more discriminant color space for FR, color image discriminant (CID) space-based FR framework [8] was designed. Likewise, the authors in [7] suggested a hybrid color space constructed by combining R from RGB color space, Q from YIQ space, and C_r from YC_bC_r space, respectively (so called RQC_r). However, the authors in [2] pointed out that the FR accuracy according to the selected color space could vary depending on the FR application and the environment. To deal with color space selection, color-component feature selection framework was suggested to find the best set of color-component features from various color spaces aiming to achieve the best FR accuracy [2].

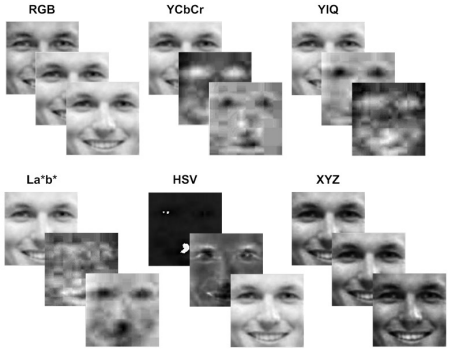

Fig. 1: Example face images in various color spaces.

Recently, deep learning frameworks have attracted research attentions due to their feature representation capability and their outstanding performance in visual classification [12, 13, 14]. By adopting a deep learning framework into FR problem, the FR accuracy has significantly improved [15, 16, 17]. Currently, RGB color images (or grayscale images) are mostly used as an input to deep learning FR methods [13, 15, 16, 17]. To the best of our knowledge, the color utilization in deep learning frameworks has not been fully investigated yet. Despite of well-known advantages of facial color and their effectiveness, the effect of facial color in deep learning is still an open question.

In this paper, we investigate FR with multiple color spaces in deep learning framework. In particular, we propose a collaborative facial color feature learning with multiple color spaces. The contributions of the paper can be summarized in two folds:

- The effectiveness of facial color is investigated through experiments with various color spaces in a deep learning framework. Through the experimental results, we have observed that learned facial color features from multiple color spaces could provide different types of complementary information.
- Motivated by the observed complementary effect of

[†]Corresponding author (ymro@kaist.ac.kr)

multiple color spaces in the deep learning framework, we propose a collaborative facial color feature learning framework. The proposed framework performs effective feature fusion by learning to aggregate the features of deep convolutional neural networks (DCNN) obtained from multiple color spaces. The color feature fusion improves the discriminative power for the learned features, and significantly improves the FR accuracy.

Through comparative experiments, we show that facial color in deep learning framework is helpful for enhancing FR accuracy. Especially the selection of color space in DCNN is important to achieve a high performance.

The remainder of the paper is organized as follows: Section 2 introduces a baseline DCNN for a color space for FR. In Section 3, the proposed collaborative facial color feature learning framework is described. Section 4 presents the comparative experimental results with analysis in terms of FR accuracy. Finally, conclusions are drawn in Section 5.

2. BASELINE DEEP CONVOLUTIONAL NEURAL NETWORKS FOR A COLOR SPACE

To investigate latent features of color face images with multiple color spaces, we start with a baseline DCNN as shown in Fig. 2. The baseline DCNN is comprised of 4 convolutional layers (C1, C2, C3, and C4), 2 subsampling (i.e., max-pooling) layers (S1 and S2), and 2 fully connected layers (F1 and output layer with softmax function), i.e., a total of 8 layers (L =8). For 3 color components of the input face image with size of 32×32 pixels, 32 feature maps are obtained by convolving 32 filter kernels with size of 3×3 (denoted as 32@3×3) in the first convolutional layer. The convolution operation is conducted as follows:

$$\mathbf{h}_j^{(l)} = \sigma\left(\sum_{i=1}^{n_l} \mathbf{h}_i^{(l-1)} * \mathbf{w}_{ij}^{(l)} + \mathbf{b}_j^{(l-1)}\right),\ l = 1, \cdots, L, \tag{1}$$

where $\mathbf{h}_j^{(l)}$ denotes the j-th feature map in the l-th layer, and n_l is the number of feature maps in the l-th layer. The DCNN is parameterized by weights $\mathbf{w}_{ij}^{(l)}$ and biases $\mathbf{b}_j^{(l)}$ to be learned. Note that the convolution in the first convolutional layer for a color face image is computed as

$$\mathbf{h}_j^{(1)} = \sigma\left(\mathbf{C}_k^1 * \mathbf{w}_{1j}^{(1)} + \mathbf{C}_k^2 * \mathbf{w}_{2j}^{(1)} + \mathbf{C}_k^3 * \mathbf{w}_{3j}^{(1)} + \mathbf{b}_j^{(1)}\right), \tag{2}$$

where $\mathbf{C}_k = \left\{\mathbf{C}_k^1, \mathbf{C}_k^2, \mathbf{C}_k^3\right\}$, which has three color components, e.g., RGB image, $\mathbf{C}_k^1 = R, \mathbf{C}_k^2 = G$ and $\mathbf{C}_k^3 = B$. In other words, the convolution in the first convolutional layer with color face image is the weighted summation of color components with weight parameters of DCNN.

The weights to be learned in DCNN are expected to have different shapes depending on the color space of input face image. Figure 3 shows examples of the 3×3 filter kernels (weights of DCNN) of the first convolutional layer learned with different color spaces, i.e., RGB, La^*b^*, YC_bC_r, and HSV. As shown in Fig. 3, weights have different shapes, thus the corresponding activation maps for input color face image have different responses (for details, please refer to Fig. 6 and 7 in the experiments). The learned features in different color spaces (i.e., the output of hidden layers) are represented in different feature subspace as well, which could lead to complementary effects. The DCNN learned with a color space $\mathbf{C}_k$ is denoted by "DCNN-C_k" throughout the rest of the paper.

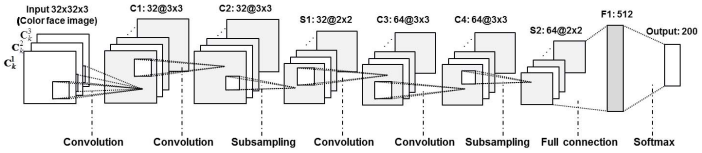

Fig. 2: Baseline DCNN structure for a color space.

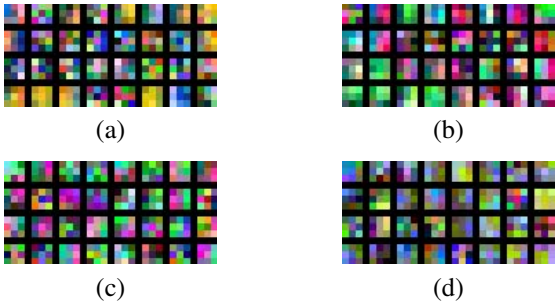

Fig. 3: Example of learned filter kernels for different color spaces. (a) RGB, (b) La^*b^*, (c) YC_bC_r, and (d) HSV.

3. COLLABORATIVE FACIAL COLOR FEATURE LEARNING WITH MULTIPLE COLOR SPACES

This section describes the collaborative facial color feature learning framework with multiple color spaces. The proposed framework is motivated by the observation of the complementary effects of multiple color spaces in deep learning frameworks. Figure 4 shows the proposed collaborative facial color feature learning framework, which is mainly comprised of: 1) multiple color feature extraction by the DCNNs for multiple color spaces and 2) multiple color feature fusion using deep neural networks (DNN). Given an input RGB face image, multiple color spaces are generated. By learning a DCNN with each color space (DCNN-C_k), feature representation for each color spaces is obtained. Then, multiple color feature fusion is applied to aggregate the learned features from multiple color spaces. Here, features can be simply combined by concatenating all feature vectors from multiple color spaces to achieve high FR accuracy. However, the discrimination capabilities of various feature vectors are different. So, maximum performance cannot be guaranteed by simply concatenating

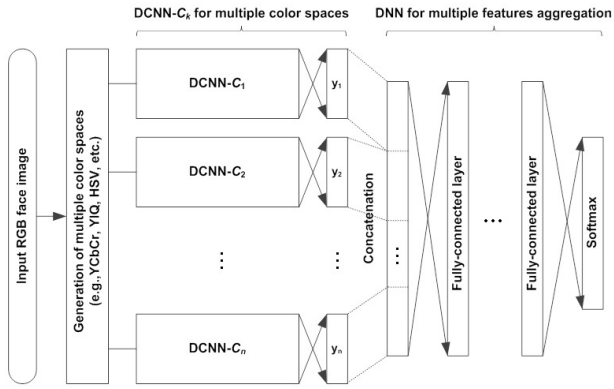

Fig. 4: Collaborative facial color feature extraction framework with the learned DCNNs and DNN to aggregate multiple features obtained from multiple color spaces.

all feature vectors. Furthermore, the curse-of-dimensionality with the increased dimension cannot be avoided. The proposed feature fusion is conducted by learning a DNN comprised of several fully-connected layers. By learning and aggregating the feature vectors learned from DCNN-C_k in multiple color spaces, the final output feature vector is a compact representation, which enhances the discriminative power of the learned features. Both feature selection and aggregation could be latently learned by learning the outputs of DCNN-C_k.

In Fig. 4, a color input (RGB) face image is transformed to multiple color spaces (e.g., YC_bC_r, YIQ, HSV, etc.). Then, from the learned DCNNs in multiple color spaces (denoted as DCNN-C_1, DCNN-C_2,$\cdots$,DCNN-C_k for multiple color space), multiple features $\mathbf{y}_{\mathbf{C}_1}$, $\mathbf{y}_{\mathbf{C}_2}$, $\cdots$, $\mathbf{y}_{\mathbf{C}_n}$ with d-dimension are extracted. By concatenating all features, a $(n \times d)$-dimensional feature vector is formed, i.e., $\mathbf{y} = \left(\mathbf{y}_{C_1}^\top, \mathbf{y}_{C_2}^\top, \cdots, \mathbf{y}_{C_n}^\top\right)^\top \in \mathbb{R}^{(n \times d)}$. For the given $\mathbf{y}$, combined feature is extracted by the DNN parameterized by $\mathbf{Q}^{(m)}$ and $\mathbf{b}^{(m)}$.

$$\mathbf{f}^{(m)} = \sigma\left(\mathbf{Q}^{(m)^\top}\mathbf{f}^{(m-1)} + \mathbf{b}^{(m)}\right),\ m = 1, \cdots, M, \quad (3)$$

where $\mathbf{Q}^{(m)}$ and $\mathbf{b}^{(m)}$ denotes weights and bias of the m-th layer for the DNN, respectively. M is the number of layers in the DNN. $\mathbf{f}^{(m)}$ is aggregated feature representation in the m-th layer, where $\mathbf{f}^{(0)} = \mathbf{y}$. By the DNN parameters, the concatenated features are mapped to a low-dimensional aggregated feature space. Aggregating the color features from multiple color spaces results in features that have a compact representation. Furthermore, discriminating features could be emphasized by the learned weights of DNN.

4. EXPERIMENTS

4.1. Experimental Settings

In order to verify the proposed method, 9 different color spaces were obtained from the original RGB face image, which are YC_bC_r, YIQ, XYZ, La^*b^*, HSV, RQC_r [7],RIQ, YQC_r, and Grayscale. For the purpose of FR, a subset of Multi-PIE dataset [18] under illumination variations was used. The dataset has 14,320 training images from 200 subjects, 5,863 probe face images, and 137 gallery face images from 137 subjects (i.e., the total number of subjects was 337). Every face image was resized to have 32×32 pixels. Note that subjects in the training and testing were mutually exclusive. The FR accuracy was evaluated via a 1-nearest neighbor classification method by comparing the feature vector from the gallery face image to the feature vector from the probe face image.

In our DCNN structure, the learning rate was set to 0.01 with the exponential decay [19]. The probability of the DropOut [20] was set to 0.5 for the softmax layer, and 0.25 after two subsampling layers. For the DNN structure, two fully-connected layers were constructed with 2,048 units and 512 units, and softmax layer. The learning rate in the DNN was set to 0.001 with DropOut probability 0.5. And, in the DCNN-C_k, a rectified linear unit (ReLU) [21] was used for activation function, i.e., $\sigma(x) = \max(0, x)$. In the DNN, hyperbolic tangent activation function was used, $\sigma(x) = (e^x - e^{-x})/(e^x + e^{-x})$. In order to learn both DCNN and DNN, cross-entropy loss [22] function-based mini-batch gradient descent was adopted, where the batch size was set to 128.

4.2. Experimental Results

Table 1 shows the FR accuracy for the DCNN-C_k learned with various color spaces. Overall, FR accuracies for color face images were higher than that of grayscale face image. For the La^*b^* face image, FR accuracy was significantly increasing, i.e., 17.05% gain. In addition, 8.27% improvement was achieved compared to the FR accuracy for the RGB color image. This means that facial identity information is learned differently in DCNN-C_k according to the color space. For example, Fig. 6 and 7 show feature maps for a test face image after the first and the second convolutional layers. Facial identity feature maps in different color spaces appeared differently by DCNN-Ck. Moreover, as shown in Fig. 7, the distance between "Class1 (RGB)" and "Class1 (YC_bC_r)" in the feature subspace was further apart from each other even though they represented the same class. Added to that, it surpassed the between-class distance in feature subspace. To analyze the complementary effect resided in multiple features from multiple color space, we selected a subset of features from the concatenated 4,608 (512 dim × 9 color spaces) dimensional features (outputs of DCNN-C_k) based on Fisher's criterion [23]. When 3,000 features from 4,608 dimensional features were selected, FR accuracy was about 87%. In 3,000 features, 338, 326, 342, 326, 341, 333, 327, 357, 310 features were from RGB, YC_bC_r, YIQ, XYZ, HSV, RIQ, RQC_r, YQC_r, and La^*b^* color spaces. Features from each color

Table 1: FR accuracy for the DCNN-C_k learned with k color space with Multi-PIE dataset.

Color space (k)	Grayscale	XYZ	RGB	HSV	YIQ	YQC_r	RQC_r	RIQ	YC_bC_r	La^*b^*
FR accuracy	62.72%	69.44%	71.50%	72.83%	74.47%	74.55%	75.03%	76.34%	78.17%	79.77%

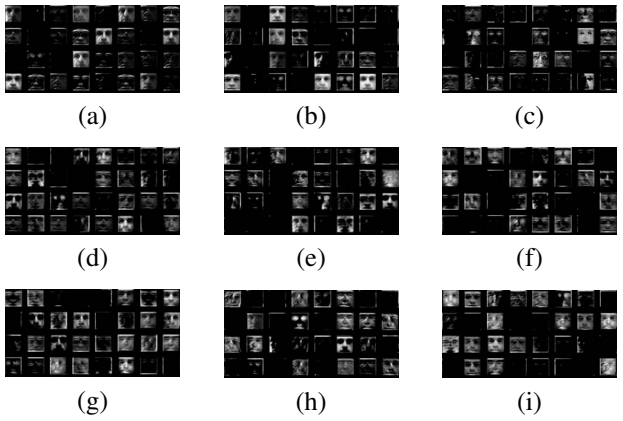

Fig. 5: The feature maps after the first convolutional layer (C1) for face images from (a) XYZ, (b) RGB, (c) HSV, (d) YIQ, (e) YQC_r, (f) RQC_r, (g) RIQ, (h) YC_bC_r, and (i) La^*b^* color spaces.

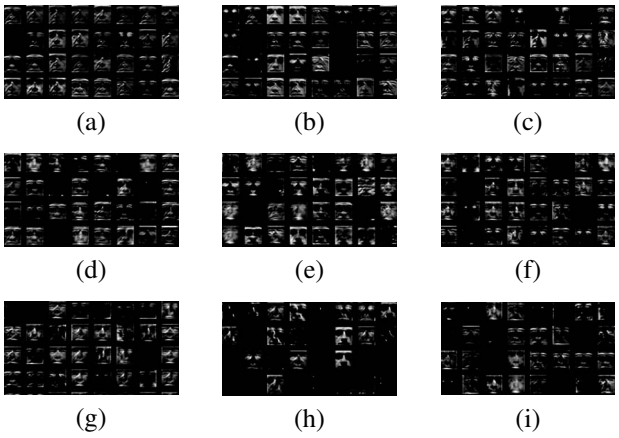

Fig. 6: The feature maps after the second convolutional layer (C2) for face images from (a) XYZ, (b) RGB, (c) HSV, (d) YIQ, (e) YQC_r, (f) RQC_r, (g) RIQ, (h) YC_bC_r, and (i) La^*b^* color spaces.

Table 2: FR accuracy for the feature fusion.

	FR accuracy
Feature-level concatenation (4,608 dim)	87.34%
Collaborative feature learning (512 dim)	**90.81%**

space contributed with about 10% contributions equivalently. From the result and FR performance with single DCNN-C_k shown in Table 1, we observed that learned features from various color spaces are complementary in FR performance.

Table 2 compares the FR performance by the proposed method to that of the feature-level concatenation with Multi-PIE dataset [18]. As aforementioned, for the concatenated features obtained from 9 color spaces, the FR accuracy was significantly improved to 87.34%. By simply fusing multiple features, we could obtain highly discriminating complementary effect from different color spaces. When the proposed feature fusion method by the DNN was used, FR accuracy was further improved with a lower feature dimension. The DNN for the feature fusion supposed to be learned by fully utilizing complementary information from output of multiple DCNN-C_k while eliminating insignificant features.

5. CONCLUSION

In this paper, we investigated the effects of color spaces on face images in DCNN framework-based FR. Experimental result showed that FR accuracy with color face image was

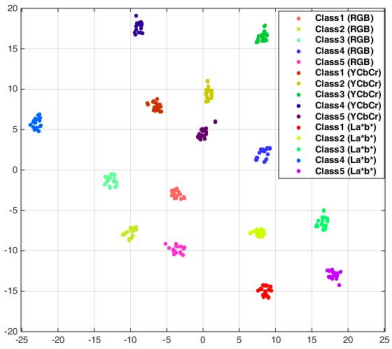

Fig. 7: 2-dimensional feature subspace reduced by t-SNE [24] for test face images from 5 classes with RGB, YC_bC_r, and La^*b^* color spaces.

significantly higher than that of grayscale face image. FR accuracy could be considerably improved according to the aggregation and selection of the learned features from multiple color spaces. In particular, by employing DNN-based feature fusion, the FR accuracy has increased by fully utilizing the complementary information resided in multiple color spaces. 28.09% FR accuracy gain was achieved compared with grayscale FR.

6. ACKNOWLEDGEMENT

This work was partially supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2015R1A2A2A01005724).

7. REFERENCES

[1] J. Y. Choi, Y. M. Ro, and K. N Plataniotis, "Color face recognition for degraded face images," *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 39, no. 5, pp. 1217–1230, 2009.

[2] J. Y. Choi, Y. M. Ro, and K. N Plataniotis, "Boosting color feature selection for color face recognition," *IEEE Trans. Image Processing*, vol. 20, no. 5, pp. 1425–1434, 2011.

[3] J. Y. Choi, Y. M. Ro, and K. N Plataniotis, "Color local texture features for color face recognition," *IEEE Trans. Image Processing*, vol. 21, no. 3, pp. 1366–1380, 2012.

[4] C. Jones III et al., "Color face recognition by hypercomplex gabor analysis," in *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition (FG)*, 2006, pp. 6–pp.

[5] S. H. Lee, J. Y. Choi, Y. M. Ro, and K. N Plataniotis, "Local color vector binary patterns from multichannel face images for face recognition," *IEEE Trans. Image Processing*, vol. 21, no. 4, pp. 2347–2353, 2012.

[6] S. H. Lee, H. Kim, Y. M. Ro, and K. N Plataniotis, "Using color texture sparsity for facial expression recognition," in *Proc. IEEE Int'l Conf. on Automatic Face and Gesture Recognition (FG)*, 2013, pp. 1–6.

[7] Z. Liu and C. Liu, "Fusion of color, local spatial and global frequency information for face recognition," *Pattern Recognition*, vol. 43, no. 8, pp. 2882–2890, 2010.

[8] Z. Liu, J. Yang, and C. Liu, "Extracting multiple features in the cid color space for face recognition," *IEEE Trans. Image Processing*, vol. 19, no. 9, pp. 2502–2509, 2010.

[9] L. Torres, J. Y. Reutter, and L. Lorente, "The importance of the color information in face recognition," in *Proc. IEEE Int'l Conf. Image Processing (ICIP)*, 1999, pp. 627–631.

[10] J. Yang and C. Liu, "Color image discriminant models and algorithms for face recognition," *IEEE Trans. Neural Networks*, vol. 19, no. 12, pp. 2088–2098, 2008.

[11] J. Yang, C. Liu, and J. Y. Yang, "What kind of color spaces is suitable for color face recognition?," *Neurocomputing*, vol. 73, no. 10, pp. 2140–2146, 2010.

[12] L. Deng and D. Yu, "Deep learning: methods and applications," *Foundations and Trends in Signal Processing*, vol. 7, no. 3–4, pp. 197–387, 2014.

[13] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems (NIPS)*, 2012, pp. 1097–1105.

[14] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[15] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1701–1708.

[16] Z. Zhu, P. Luo, X. Wang, and X. Tang, "Deep learning identity-preserving face space," in *Proc. IEEE Int'l Conf. Computer Vision (ICCV)*, 2013, pp. 113–120.

[17] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Prof. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1891–1898.

[18] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010.

[19] A. Senior, G. Heigold, M. Ranzato, and K. Yang, "An empirical study of learning rates in deep neural networks for speech recognition," in *Proc. IEEE Int'l Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 6724–6728.

[20] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[21] V. Nair and G. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. Int'l Conf. Machine Learning (ICML)*, 2010, pp. 807–814.

[22] P. Golik, P. Doetsch, and H. Ney, "Cross-entropy vs. squared error training: a theoretical and experimental comparison.," in *Proc. Interspeech*, 2013, pp. 1756–1760.

[23] Richard O Duda, Peter E Hart, and David G Stork, *Pattern classification*, John Wiley & Sons, 2012.

[24] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, no. 2579-2605, pp. 85, 2008.