Curriculum Learning for Face Recognition

Barış Büyüktaş
Dept. of Electrical and Electronics Eng.
Özyeğin University
Istanbul, Turkey
baris.buyuktas@ozu.edu.tr

Çiğdem Eroğlu Erdem
Dept. of Computer Eng.
Marmara University
Istanbul, Turkey
cigdem.erdem@marmara.edu.tr

Tanju Erdem
Dept. of Electrical and Electronics Eng.
Özyeğin University
Istanbul, Turkey
tanju.erdem@ozyegin.edu.tr

***Abstract*—We present a novel curriculum learning (CL) algorithm for face recognition using convolutional neural networks. Curriculum learning is inspired by the fact that humans learn better, when the presented information is organized in a way that covers the easy concepts first, followed by more complex ones. It has been shown in the literature that that CL is also beneficial for machine learning tasks by enabling convergence to a better local minimum. In the proposed CL algorithm for face recognition, we divide the training set of face images into subsets of increasing difficulty based on the head pose angle obtained from the absolute sum of yaw, pitch and roll angles. These subsets are introduced to the deep CNN in order of increasing difficulty. Experimental results on the large-scale CASIA-WebFace-Sub dataset show that the increase in face recognition accuracy is statistically significant when CL is used, as compared to organizing the training data in random batches.**

***Index Terms*—face recognition, deep learning, curriculum learning**

I. Introduction

Face recognition is an important problem, which has been attracting the attention of researchers for a long time [1], [2]. Face recognition has many application areas including security, education, marketing, advertising, and entertainment [3].

During the past few years, great progress has been achieved for face recognition using deep Convolutional Neural Networks (CNN), which require large annotated datasets for training containing a diverse set of images for each subject [4]–[10]. The diversity of face recognition datasets arise from variations in head pose, illumination, resolution, image quality and accessories [11]–[17]. Diversity of image datasets has been shown to increase the accuracy of deep CNNs for classification tasks [18]. During the training phase, these images are fed to the network in batches, which are selected randomly from the training dataset and hence may contain both easy and more challenging samples of the same person. The CNNs learn the optimal parameters from these randomly formed batches of data.

The term *curriculum* is defined as the educational process with regard to the educator's or school's instructional goals [19]. The educational process mostly starts with easier subjects. As the learning level increases, students can understand more complex subjects readily. This is due to the well-known fact that it is more intuitive for humans to start learning with easy samples at the beginning of learning. Elman [20] demonstrates this fact with children's language learning using training samples presented from easy to complex and in random order.

Curriculum learning has been applied to various machine learning problems. It is used mainly to improve the speed of convergence and achieve better generalization by increasing the complexity of the training data gradually [21]. Pentina et al. [22] use curriculum learning for object categorization. It was shown that learning multiple tasks in an order is more beneficial as compared to learning them jointly. Amramova [23] states that ordering the training samples in ascending or descending order based on their difficulty could affect the convergence of a deep CNN.

Bengio et al. [24] use curriculum learning with recurrent neural networks to help the model explore further during training for the tasks of image captioning, constituency parsing and speech recognition. Guie et al. [25] demonstrate that curriculum learning can be beneficial for facial expression recognition. They split their training data into subsets of increasing complexity levels based on the intensity of the facial expression and the training starts from the easiest samples. As the network becomes more robust, complex samples are added into training set without discarding easy samples.

In this work, we propose a novel curriculum learning algorithm for face recognition. We investigate the idea that head pose can be used as a measure of difficulty, especially for large datasets collected in the wild with a variety of head poses. Hence, we present the training data to the CNN, in order of increasing difficulty using the head pose information obtained from pitch, yaw and roll angles as illustrated in Fig. 1. Experimental results on the large scale CASIA-WebFace dataset [11] indicate statistically significant improvements in face recognition accuracy using the proposed curriculum learning approach as compared to random ordering of training data. To the best of our knowledge, this is the first work, which applies curriculum learning to the task of face recognition.

II. Proposed Method

In this section, we give the details of the proposed curriculum learning based face recognition method. We first describe the head pose estimation method and then give the

This work was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under project 116E088. The Titan V used for this research was donated by the NVIDIA Corporation.

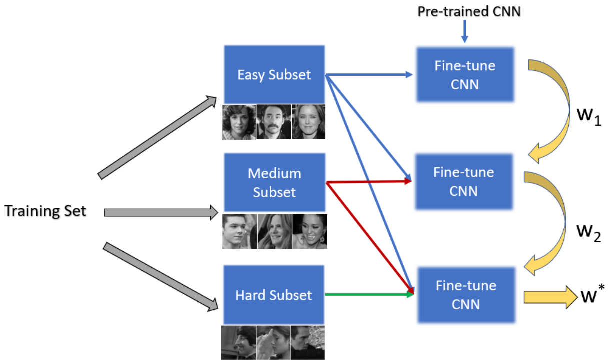

Fig. 1. The framework of the proposed curriculum learning approach for face recognition. The training set is ranked into subsets using difficulty levels based on head pose. First, the easiest subset is used to fine-tune a pre-trained CNN. Then, more difficult subsets are included to the training set for further fine-tuning the CNN.

details of the proposed curriculum learning algorithm for face recognition.

A. Determining Training Data Difficulty Based on Head Pose

The effect of image quality on face recognition performance has been analyzed by Dutta et al. in [26]. Differences in area under the ROC curve were calculated for each image quality parameter, which were based on the head pose, illumination, noise, motion blur, and resolution. According to the experimental results, head pose was found to be the most important factor. Therefore, we decided to specify the complexity of images in our training set using the head pose angles. Since it is intuitive that upright frontal faces are the easiest to recognize, we used the sum of absolute values of yaw, pitch and roll angles of the head pose to rank the difficulty level of the face images in the training dataset.

There are a number of head pose estimation approaches in the literature [27]–[31]. We used OpenFace 2.0, which has been shown to perform well under difficult conditions [32].

We obtained our training and test data from the CASIA-WebFace database which contains 494,414 images from 10,575 people. Although the number of images per person ranged from 3 to 804, in order to reduce the imbalance in this distribution, individuals with fewer than 100 images were not used in the experiments as in [33]. Therefore, the CASIA-WebFace-Sub dataset used in the experiments contained 181,279 images from 923 individuals. These images are grayscale with a size of 100 × 100 pixels.

CASIA-WebFace-Sub dataset contains face images taken with a variety of head poses, several examples of which are shown in Fig.2 along with the estimated head pose angles and difficulty levels. The histograms of the absolute values of yaw, pitch and roll angles of the head pose are given in Fig.3, together with the histogram of their absolute sum. A yaw angle of 90°means, the face is seen from a profile view. The range of values for pitch and roll angles are less than 90°. As can be seen from the histogram of sum of absolute values of angles, most face images have a total angle below 50°, and less than 10% of the images have a total angle above 100°. We used the total angle value to rank the face images into subsets of increasing difficulty levels. Since the distribution of pitch and roll angles have a narrower range, yaw angle is the most effective angle for determining the total head pose angle.

Yaw	8.37°	14.4°	35.10°	49.14°	65.43°	68.22°
Pitch	0.63°	6.12°	23.58°	11.70°	9.45°	19.90°
Roll	1.62°	1.26°	0.81°	0.63°	11.70°	13.50°
Sum	10.62°	21.78°	59.49°	61.47°	86.58°	101.62°
Difficulty	easy	easy	medium	medium	hard	hard

Fig. 2. Absolute values of estimated yaw, pitch, and roll angles of sample images from the CASIA-WebFace-Sub dataset. Difficulty level is determined by the sum of all absolute angles.

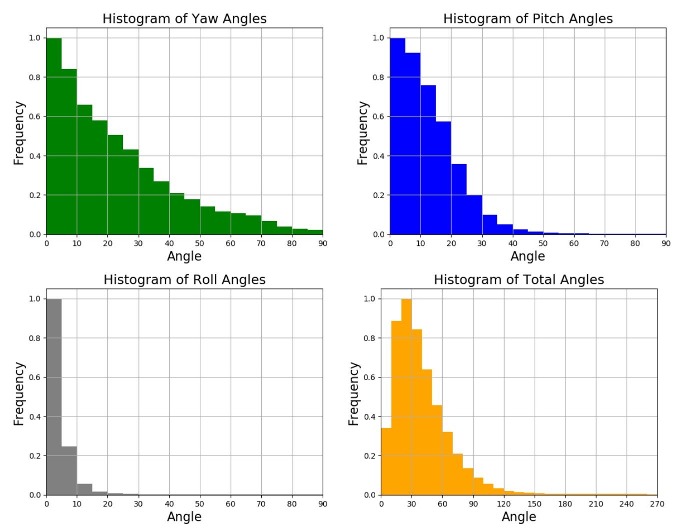

Fig. 3. Histograms of absolute values of estimated yaw, pitch, roll and total angle for the images in CASIA-WebFace-Sub dataset.

B. Curriculum Learning for Face Recognition

During the training phase of a deep neural network, the weights of the model are learned using iterative numerical methods, since the objective function to be minimized is highly non-convex. Therefore, in curriculum learning, it is important to present the network with easy instances first, so that the network is guided towards a better local minimum. A recent theoretical analysis of curriculum learning shows that CL effectively modifies the optimization landscape [34].

We propose a novel curriculum learning algorithm for face recognition, which uses the intuitive difficulty in face datasets, the head pose. We propose to organize the training dataset into easy, medium and hard subsets using the total head pose angle. Existence of images with larger variations in head pose in the training set increases face recognition accuracy [35], but these difficult images should be introduced to the network after the easier ones.

We begin with the pre-trained AlexNet CNN model, which was trained on the ImageNet dataset for object recognition [36]. Then, we fine-tune the several final layers of the CNN using the easy subset (see Fig.1). After convergence, the medium subset is augmented to the easy subset and the fine-tuning process continues with the augmented training set until convergence. Finally, we augment the dataset with the hard subset and repeat. We summarize the followed steps in Algorithm 1.

Algorithm 1 Curriculum learning for face recognition

Inputs: Pre-ordered input dataset ranked into n subsets of increasing difficulty $X = \{X^i\}_{i=1}^n$, which is determined with the curriculum.
Pre-trained CNN.
Output: Optimal model parameters W*

```
X^train = ∅
for i = 1 : n do
    X^train = X^train ∪ X^i
    while not converged do
        fine-tune (W, X^train)
    end
    choose optimal W*
    decrease learning rate
end
```

III. Experimental Results

In this section, we first give the details of the experimental setup including the five datasets used in the experiments. Then, we compare the face recognition accuracies for three different CNN fine-tuning strategies using random ordering, curriculum-based ordering and inverse-curriculum based ordering of training data.

A. Experimental Setup

We used five-fold cross validation during the experiments. In each fold, 80% of the images (approx. 145.055 images) in CASIA-WebFace-Sub dataset were used for training and 20% of them were used for testing (approx. 36.255 images). Then, we estimated the head pose of all images in the training dataset and formed subsets of increasing difficulty. Head pose of 6464 out of 145.055 images could not be estimated with OpenFace 2.0, since they were blurry, noisy or non-frontal. Hence, these images are directly placed into the most difficult subset. Table I shows the total absolute head pose angles corresponding to the complexity (difficulty) ranking in the ordered training dataset.

We formed five different training image sets, from the whole training set of 145.055 images as shown in Table II, which have increasing number of total samples selected from the whole training dataset. Image sets 1-3 consist of 30.000, 60.000, and 90.000 images, respectively. The goal of constructing image sets 1-3 is to study how the recognition accuracies are affected as the training dataset includes distinct

TABLE I
RANK OF THE ORDERED IMAGES IN THE TRAINING SET FROM EASY TO DIFFICULT AND THE CORRESPONDING TOTAL HEAD POSE ANGLES.

Complexity Order	Angle
0	0.18°
10.000	10.53°
20.000	15.21°
30.000	19.08°
75.000	36.72°
85.000	41.76°
95.000	47.61°
105.000	54.81°
115.000	64.44°
125.000	80.28°
135.000	197.10°

difficulty levels between images in the easy, medium and difficult (hard) subsets.

- **Image set 1**: Consists of image subsets having a rank between 0-10.000 (very easy), 85.000-95.000 (medium), and 135.000-145.000 (very hard) in the ordered list of all images CASIA-WebFace-Sub dataset.
- **Image set 2**: Consists of subsets having a rank between 0-20.000 (easy), 80.000-100.000 (medium), and 125.000-145.000 (hard) in the ordered list of all images CASIA-WebFace-Sub dataset.
- **Image set 3**: Consists of subsets having a rank between 0-30.000 (easy), 75.000-105.000 (medium), and 115.000-145.000 (hard) of all images CASIA-WebFace-Sub dataset.
- **Image set 4**: Consists of all the training images in the CASIA-WebFace-Sub dat set. There are three subsets of increasing difficulty as follows. Images with a total absolute head pose angle below 35° were grouped as easy images, whereas images with angles between 35°-65° and higher than 65° were grouped as medium difficulty and hard images, respectively.
- **Image set 5**: Consists of the same total number of images as image set 4, but it has been split into 4 subsets of increasing difficulty to provide a smoother transition from easy to difficult images. It is separated into subsets using total head pose angles as follows: 0° - 30° (Easy), 30° - 60° (Medium 1), 60° - 70° (Medium 2) and above 70° (Hard).

TABLE II
IMAGE SETS OBTAINED USING THE CASIA-WEBFACE-SUB DATASET, AND THE NUMBER OF IMAGES IN EACH CATEGORY.

	Total	Easy	Medium		Hard
Image Set 1	30.000	10.000	10.000		10.000
Image Set 2	60.000	20.000	20.000		20.000
Image Set 3	90.000	30.000	30.000		30.000
Image Set 4	145.055	71.150	44.355		29.550
	Total	**Easy**	**Medium 1**	**Medium 2**	**Hard**
Image Set 5	145.055	59.165	31.467	28.643	25.780

When we use curriculum learning, training starts with the easiest subset, and the next subset is augmented to the previous one after convergence. After each data augmentation step, the network is trained with a lower learning rate. Early stopping is used, when the accuracy of the network does not increase after 30.000 iterations with a test interval of 10.000 iterations. CNN is updated using the stochastic gradient descent with momentum 0.9, learning rate 0.001, and weight decay 0.0001.

B. Face Recognition Results

We first compared the performances of two AlexNet models without using curriculum learning: i) trained from scratch and ii) after fine-tuning the last two fully connected layers of the pre-trained AlexNet model using all 145.055 images in the training dataset. The accuracy after training from scratch was 77.28%, and the accuracy after fine-tuning the last 2 fully connected layers was 80.93%. Hence, we observed that fine-tuning increased the accuracy by 3.65%.

We conducted experiments on the five training sets described above using three different approaches: with curriculum learning, without CL (i.e.random ordering) and with inverse CL (from hard to easy). We also investigated the effect of fine-tuning the last, last two and last three fully connected layers (with the parameters of the previous layers frozen), which are given in Table III, Table IV, and Table V, respectively.

TABLE III
ACCURACIES WITHOUT CL (W/O CL), WITH CL (W/ CL), AND WITH INVERSE CL LEARNING, WHEN FINE-TUNING THE LAST FULLY CONNECTED LAYER.

Methods / Image sets	w/o CL	w/ CL	Increase	Inverse CL
Image set 1	0.4892	0.5232	3.40%	0.3631
Image set 2	0.6359	0.6894	5.35%	0.5838
Image set 3	0.7030	0.7482	4.52%	0.6743
Image set 4	0.7981	0.8106	1.25%	0.7513
Image set 5	0.7981	0.8135	1.54%	0.7682

TABLE IV
ACCURACIES WITHOUT CL (W/O CL), WITH CL (W/ CL), AND WITH INVERSE CL LEARNING, WHEN FINE-TUNING THE LAST TWO FULLY CONNECTED LAYERS.

Methods / Image sets	w/o CL	w/ CL	Increase	Inverse CL
Image set 1	0.5093	0.5396	3.03%	0.3876
Image set 2	0.6493	0.6971	4.78%	0.5942
Image set 3	0.7054	0.7490	4.36%	0.6727
Image set 4	0.8093	0.8223	1.30%	0.7613
Image set 5	0.8093	**0.8240**	1.47%	0.7787

TABLE V
ACCURACIES WITHOUT CL (W/O CL), WITH CL (W/ CL), AND WITH INVERSE CL LEARNING, WHEN FINE-TUNING THE LAST THREE FULLY CONNECTED LAYERS.

Methods / Image sets	w/o CL	w/ CL	Increase	Inverse CL
Image set 1	0.4826	0.5238	4.12%	0.3894
Image set 2	0.6514	0.6963	5.35%	0.6013
Image set 3	0.7049	0.7488	4.49%	0.6784
Image set 4	0.8030	0.8160	1.30%	0.7523
Image set 5	0.8030	0.8181	1.51%	0.7785

According to the results given in Table III, Table IV, and Table V, the proposed CL algorithm consistently gives higher results as compared to random ordering. The highest accuracies are achieved when the last two fully connected layers are fine-tuned as can be seen in Table IV. Total accuracy increase on the five image sets are 3.03%, 4.78%, 4.36%, 1.30%, and 1.47%, respectively. The highest accuracy achieved is 0.8240 with CL on image set 5, which contains image subsets of 4 difficulty levels.

The inverse CL results are worse than random ordering of training data, which further justifies that ordering the training data from easy to difficult is indeed helpful for convergence to a better local minimum. Results on image set 5 show that using four subsets of difficulty is slightly better than using three subsets. First 3 image sets include the easiest and the most difficult samples. Therefore, the increase in accuracies when CL is used is higher for these datasets as compared to image sets 4 and 5.

We also used McNemar's Test [37] to investigate whether the increase in the accuracies are statistically significant. We used the number of misclassified examples with CL and random ordering using image set 5. Chi-square statistic with one degree of freedom was used. We obtained a score of 95.70, which indicates statistical significance since it is above 3.84.

We also observed that the increase in accuracies with CL are due to the correction of the labels of images with higher difficulty levels. In Fig.4, histograms of the total head pose angles of all the test images and images whose labels have been corrected with curriculum learning are given. The average total head pose angle of all test images is 48°and the average total head pose angle of corrected images with CL is 58°. Therefore, we can conclude that using curriculum learning enables the CNN to recognize more difficult images better.

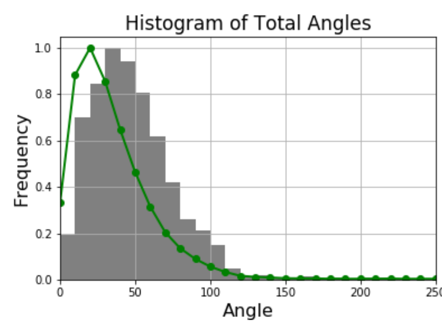

Fig. 4. Histogram of total head pose angles for all test images (green curve) and images whose labels have been corrected with curriculum learning (gray bars).

In order to compare our results with the state-of-the-art, We used the same training parameters as [33] on the CASIA-WebFace-Sub dataset and followed a similar experimental setup. However, we don't restrict the total number of iterations during CNN fine-tuning. Instead, we waited until the accuracy doesn't change for 30.000 iterations. In [33], using all of the 145.055 training images the accuracy was approximately 0.77. Our curriculum learning based face recognition method gives an accuracy of 0.8240, which is significantly higher. Summary of our experimental results are given in Table VI.

Experiments were run on a workstation with an i7 3.4 GHz CPU, NVIDIA Titan V and NVIDIA GTX 1080 Ti GPUs.

TABLE VI
SUMMARY OF THE FACE RECOGNITION RESULTS AND COMPARISON WITH THE STATE-OF-THE-ART.

Our Work (one layer)	0.8135
Our Work (two layers)	0.8240
Our Work (three layers)	0.8181
Lin [33]	≈ 0.77

IV. CONCLUSION

In this paper, we presented a novel curriculum learning algorithm for face recognition. In curriculum learning, the trained data is ranked into subsets of increasing difficulty, which was achieved using the total head pose angle. Experimental results on the large-scale CASIA-WebFace-Sub dataset shows that the recognition accuracy increases in a statistically significant way when the proposed CL algorithm is used as compared to random ordering of data when training a CNN. The increased recognition accuracy is especially evident in difficult images.

As for future work, attributes other than the head pose can be used to assess the difficulty of face images since frontal images can be noisy or blurry. In such cases, images that contain blur and noise could be considered as difficult images.

REFERENCES

[1] T. Kanade, *Computer Recognition of Human Faces*. Basel und Stuttgart: Birkhauser Verlag, 1977.

[2] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[3] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2d and 3d face recognition: A survey," *Pattern Recognition Letters*, vol. 28, no. 14, pp. 1885–1906, 2007.

[4] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1701–1708, 2014.

[5] O. M. Parkhi, A. Vedaldi, A. Zisserman, *et al.*, "Deep face recognition," in *British Machine Vision Conference*, 2015.

[6] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 815–823, 2015.

[7] R. Ranjan, C. D. Castillo, and R. Chellappa, "L2-constrained softmax loss for discriminative face verification," *arXiv preprint arXiv:1703.09507*, 2017.

[8] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *IEEE Conf. on Automatic Face and Gesture Recognition*, 2018.

[9] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4690–4699, 2019.

[10] C. Huang, Y. Li, C. L. Chen, and X. Tang, "Deep imbalanced learning for face recognition and attribute prediction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[11] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," *arXiv preprint arXiv:1411.7923*, 2014.

[12] 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. Best of Automatic Face and Gesture Recognition 2008.

[13] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard, "The megaface benchmark: 1 million faces for recognition at scale," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4873–4882, 2016.

[14] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition," in *European conference on computer vision*, pp. 87–102, Springer, 2016.

[15] A. Bansal, A. Nanduri, C. D. Castillo, R. Ranjan, and R. Chellappa, "Umdfaces: An annotated face dataset for training deep networks," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 464–473, IEEE, 2017.

[16] B. Maze, J. Adams, J. A. Duncan, N. Kalka, T. Miller, C. Otto, A. K. Jain, W. T. Niggel, J. Anderson, J. Cheney, *et al.*, "Iarpa janus benchmark-c: Face dataset and protocol," in *2018 International Conference on Biometrics (ICB)*, pp. 158–165, IEEE, 2018.

[17] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *IEEE Int. Conf. on Automatic Face & Gesture Recognition (FG)*, pp. 67–74, IEEE, 2018.

[18] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep deatures for scene recognition using places database," in *Advances in Neural Information Processing Systems*, pp. 487–495, 2014.

[19] A. V. Kelly, *The Curriculum: Theory and Practice*. Sage, 2009.

[20] J. L. Elman, "Learning and development in neural networks: The importance of starting small," *Cognition*, vol. 48, no. 1, pp. 71–99, 1993.

[21] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Int. Conf. on Machine Learning*, pp. 41–48, ACM, 2009.

[22] A. Pentina, V. Sharmanska, and C. H. Lampert, "Curriculum learning of multiple tasks," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 5492–5500, 2015.

[23] V. Avramova, "Curriculum learning with deep convolutional neural networks," Master's thesis, KTH Royal Institute of Technology, 2015.

[24] S. Bengio, O. Vinyals, N. Jaitly, and N. Shazeer, "Scheduled sampling for sequence prediction with recurrent neural networks," in *Advances in Neural Information Processing Systems*, pp. 1171–1179, 2015.

[25] L. Gui, T. Baltrušaitis, and L.-P. Morency, "Curriculum learning for facial expression recognition," in *IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pp. 505–511, IEEE, 2017.

[26] A. Dutta, R. Veldhuis, and L. Spreeuwers, "The impact of image quality on the performance of face recognition," in *33rd WIC Symposium on Information Theory in the Benelux*, (Boekelo, The Netherlands), pp. 141–148, 2012.

[27] L.-P. Morency, J. Whitehill, and J. Movellan, "Generalized adaptive view-based appearance model: Integrated framework for monocular head pose estimation," in *IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 1–8, IEEE, 2008.

[28] X. Liu, H. Lu, and H. Luo, "A new representation method of head images for head pose estimation," in *IEEE International Conference on Image Processing (ICIP)*, pp. 3585–3588, IEEE, 2009.

[29] G. Fanelli, J. Gall, and L. Van Gool, "Real time head pose estimation with random regression forests," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 617–624, IEEE, 2011.

[30] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7291–7299, 2017.

[31] T. Baltrušaitis, P. Robinson, and L.-P. Morency, "Openface: An open source facial behavior analysis toolkit," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1–10, IEEE, 2016.

[32] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *IEEE Int. Conf. on Automatic Face & Gesture Recognition (FG 2018)*, pp. 59–66, 2018.

[33] L. Lin, K. Wang, D. Meng, W. Zuo, and L. Zhang, "Active self-paced learning for cost-effective and progressive face identification," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 40, no. 1, pp. 7–19, 2018.

[34] G. Hacohen and D. Weinshall, "On the power of curriculum learning in training deep networks," in *Int. Conf. on Machine Learning*, (Long Beach, California), 2019.

[35] L. Wiskott, J.-M. Fellous, N. Krüger, and C. Von Der Malsburg, "Face recognition by elastic bunch graph matching," in *Int. Conf. on Computer Analysis of Images and Patterns*, pp. 456–463, Springer, 1997.

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

[37] E. Alpaydin, *Introduction to Machine Learning*. MIT press, 2009.