Image Quality-aware Deep Networks Ensemble for Efficient Gender Recognition in the Wild

Mohamed Selim[1], Suraj Sundararajan[1], Alain Pagani[2] and Didier Stricker[1,2]
[1]*Augmented Vision Lab, Technical University Kaiserslautern, Kaiserslautern, Germany*
[2]*German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany*

Keywords: Gender, Face, Deep Neural Networks, Quality, In the Wild.

Abstract: Gender recognition is an important task in the field of facial image analysis. Gender can be detected using different visual cues, for example gait, physical appearance, and most importantly, the face. Deep learning has been dominating many classification tasks in the past few years. Gender classification is a binary classification problem, usually addressed using the facial image. In this work, we present a deep and compact CNN (GenderCNN) to estimate the gender from a facial image. We also, tackle the illumination and blurriness that appear in still images and appear more in videos. We use Adaptive Gamma Correction (AGC) to enhance the contrast and thus, get more details from the facial image. We use AGC as a pre-processing step in gender classification in still images. In videos, we propose a pipeline that quantifies the blurriness of an image using a blurriness metric (EMBM), and feeds it to its corresponding GenderCNN that was trained on faces with similar blurriness. We evaluated our proposed methods on challenging, large, and publicly available datasets, CelebA, IMDB-WIKI still images datasets and on McGill, and Point and Shoot Challenging (PaSC) videos datasets. Experiments show that we outperform or in some cases match the state of the art methods.

1 INTRODUCTION

Gender classification is a very important problem in facial analysis. For humans, the face provides one of the most important sources for gender classification. Besides faces; clothes, physical characteristics and gait (Ng et al., 2012) can provide information that identify the gender of a person. While this problem is a routine task for our brain, it is a challenging task for computers. Identifying gender from faces has huge potential in fields like face recognition, biometrics, advertising and surveillance. Millions of images and videos are uploaded every day to the Internet. These images and videos are captured using a variety of devices ranging from mobile phones to DSLR cameras, under varying conditions. These variations in the capture process, result in variations in headpose, illumination, resolution and noise, making gender recognition from faces challenging (Ng et al., 2012). Gender recognition on videos is more challenging, due to the presence of blurriness in the video frames. A gender recognition learning-based method in general involves face detection, feature extraction and finding the gender from the features (Ng et al., 2012).

In the past few years, deep learning has been dominating different classification tasks (Levi and Hassner, 2015). Convolutional Neural Network (CNN) is the most widely used neural network for visual recognition systems and natural language processing. Deep CNNs came into the spotlight in 2012, when a deep CNN called AlexNet (Krizhevsky et al., 2012) outperformed traditional machine learning methods on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al., 2015). ImageNet is a 1000-class classification challenging competition. Deeper CNN architectures like VGG (16 or 19 layers) (Simonyan and Zisserman, 2014) and Residual Networks (152 layers) (He et al., 2015) keep dominating the challenge. Gender recognition is a binary classification problem, however it is very challenging in unconstrained environments, where different variations exist in the images. It even gets more challenging in videos due to blurriness or quality compared to still images.

In this work, we propose a CNN with 3 convolution layers. Some approaches like (Mansanet et al., 2016), uses location of the eyes to transform faces to a canonical pose with eyes located in the same horizontal line. Our CNN accepts non-aligned faces as input to predict gender. We evaluate our net-

Selim, M., Sundararajan, S., Pagani, A. and Stricker, D.
Image Quality-aware Deep Networks Ensemble for Efficient Gender Recognition in the Wild.
DOI: 10.5220/0006626103510358
In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2018) - Volume 5: VISAPP*, pages 351-358
ISBN: 978-989-758-290-5

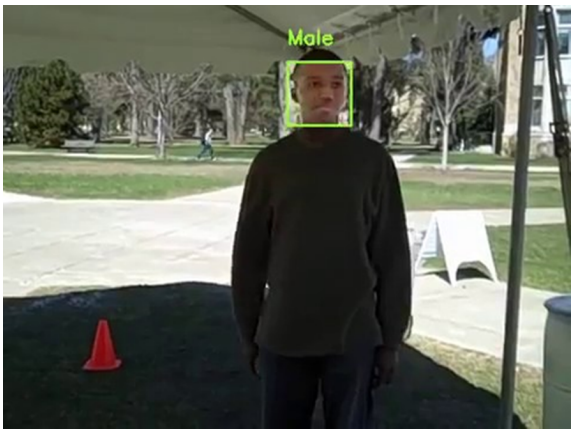


Figure 1: Sample gender detection from our proposed approach using AGC and EMBM GenderCNNs. The video frame illumination is bad, however AGC improves the facial image illumination and shows the details of the face.

work on image and video datasets. Due to the challenges available in data nowadays, new challenging "in the wild datasets" have been introduced in the past few years. In our work, the still images datasets used are IMDB-WIKI (500k images, 2015)(Rothe et al., 2016) and CelebA (200k images, 2015). Video datasets used are McGillFaces (35 videos, 10k video frames, 2015) (Demirkus et al., 2016; Demirkus et al., 2014; Demirkus et al., 2015) and PaSC (2802 videos, 663846 frames, 2013) (Beveridge et al., 2013). The meta-data provided in some datasets used face detectors that existed during the time of publishing the datasets. However, newer face detectors have been introduced to the community although the problem of face detection have been studied for many years. The face detector presented in (Mathias et al., 2014) works in very challenging situations such as dark or blurred faces. We decided to use it in our work.

Poorly illuminated faces or blurred ones exist in still images datasets and exist on a larger scale in videos datasets. In order to tackle these challenges, we propose using an adaptive gamma correction (AGC) (Rahman et al., 2016) method for pre-processing the faces. We evaluate the impact of using AGC faces as input to the network. Due to the movement of the subject or the camera, blurriness appears in videos. Manually filtering the sharp and blurry faces is a time consuming process. Hence an automated method is necessary to separate the faces based on sharpness. One way of achieving this is to use a image blurriness metric. We tackle the challenge of motion blur by using the image blurriness metric described in (Guan et al., 2015) (EMBM), to group together faces that are similar in sharpness. The EMBM values are used to split the pre-processed faces into groups and separate CNNs are trained for each group. Each group has faces that fall within a specific EMBM range.

1.1 Contributions

Our contributions in this work are

- We propose a compact, yet accurate CNN for gender recognition
- We tackle illumination, and quality issues in still images by pre-processing the faces using AGC (Rahman et al., 2016).
- For videos, we propose a blurriness-aware GenderCNNs pipeline to detect gender on both sharp and blurred video frames

We evaluated our approach on challenging, large and publicly available still images and videos datasets McGill (Demirkus et al., 2016) , CelebA (Liu et al., 2015), PaSC (Beveridge et al., 2013) , and IMDB-Wiki (Rothe et al., 2015) datasets. We outperform state of the art methods or we perform second-best in some cases.

2 RELATED WORK

In this section we provide an overview of gender recognition methods. We start by a brief overview of face detection methods and pre-processing techniques which can be used to enhance the input image before feeding it to a gender classifier. Then, gender recognition methods are discussed.

2.1 Face Detection

Face detection has been widely studied in computer vision. It is very common that in any application that involves facial image analysis, it starts with face detection. The Viola and Jones face detector has been widely used (Viola and Jones, 2001). However, it has some limitations as it can be limited to frontal faces, or it can fail in case of occlusions. In 2010, Deformable Parts Model (DPM) gained popularity in face detection(Felzenszwalb et al., 2010). In 2014, Mathias et. al. (Mathias et al., 2014) introduced a DPM-based face detector. It works well in harsh and unconstrained environments. We use their face detector in our approach, as we work with "in the Wild" datasets.

2.2 Pre-processing Techniques

Pre-proccessing an image can refer to either image enhancement or restoration. Image enhancement can

be required to get the most details available in the image. The details can be lost due to capturing conditions like illumination or the quality of the capturing device itself. The capturing conditions are not always optimal for example in images taken by amateurs using smartphones. Image enhancement is carried out to improve the image before analyzing it. Enhancment can be carried out to improve contrast, saturation, or brightness of an image. The illumination conditions affects the image contrast, consequently, important details of an image can be lost(Rahman et al., 2016). Global methods for image enhancement like histogram equalization can result in over or under-enhanced image, which also results in losing details of the image. Local methods use neighbouring pixels to overcome the problems in global methods, however, they are computationally expensive. Hybrid methods use both local and global information form the image to improve the image, like Adaptive Histogram Equalization (AHE) or Contrast Limited Adaptive Histogram Equalization (CLAHE), however they don't perform well on various problems, like dark or bright images.

An Adaptive Gamma Correction (AGC) method is proposed in (Rahman et al., 2016), and it is able to enhance dark or bright images. In short, the AGC method classifies the image first as high or low contrast, and then classifies it as dark or bright. AGC can enhance the images that need enhancement without over or under-enhancing them. For details of the AGC method please refer to (Rahman et al., 2016). A comparison of different image enhancement techniques is shown in figure 2. HE over enhanced the input image. Contrast stretching didn't improve the face, it is still dark. CLAHE over enhanced some face regions, however AGC improved the image without over or under-enhancing it. Facial details are more visible. We decided to investigate the use of AGC in our approach as a pre-processing step.

Figure 2: various contrast enhancing techniques. Over or under-enhancement in b,c,d. We use AGC (Rahman et al., 2016) in our approach.

2.3 Gender Recognition

Face has been widely used for gender detection as shown in the survey (Ng et al., 2012). Earlier works were constrained to frontal faces. Recent methods handels varying head pose, illumination, and capturing location. Recently deep learning have been used in many applications, when the ILSVRC (Russakovsky et al., 2015) was won by Deep Convolutional Neural Networks (Krizhevsky et al., 2012).

Recent works that use CNNs to solve the problem of gender classification along with other facial attributes prediction can be found in (Liu et al., 2015) and (Ranjan et al., 2016). In (Liu et al., 2015), authors introduced using two known CNNs architectures, LNet and ANet. Both networks were pretrained separately and jointly fine tuned. The LNet was used to localize the face and ANet was used to extract features for attribute prediction. One attribute was gender. They reported accuracy of 98% for CelebA and 94% for LFWA datasets.

Later the work in (Ranjan et al., 2016) was introduced. It uses AlexNet initialized with ImageNet (Russakovsky et al., 2015) weights. They used intermediate layers of the network for feature extraction. The features were classified using SVM. They were able to reach comparable results with (Liu et al., 2015) using much less data for training the network. Another work that uses a deep CNN and SVM for age and gender recognition, is described in (Levi and Hassner, 2015). The CNN used has three convolution layers, 3 fully connected layers and 2 drop layers (for training). The dataset used for evaluation was Adience (Eidinger et al., 2014). Adience dataset contains 26,000 images of 2284 subjects. The dataset is constrained with limited pose variations and no changes in location. Approaches with and without oversampling are described in the work. The method using oversampling achieved a gender recognition accuracy of 86.8% on Adience.

A deep CNN architecture was proposed in (Simonyan and Zisserman, 2014). The study indicates that the accuracy of image recognition tasks improved when the depth of the network was increased to upto 19 layers. Two deep architectures called VGG-16 and VGG-19 were proposed in the work. A face descriptor based on the VGG-16 architecture, was proposed in (Parkhi et al., 2015), for face recognition. The use of the descriptor can be investigated in gender recognition.

Gender recognition from still images gathered more attention than that in videos. In (Wang et al., 2015), a temporal coherent face desriptor was introduced. Faces from the video frames are used to create one descriptor, and that descriptor is used to identify gender using a SVM. The datasets used in this work were McGillFaces (Demirkus et al., 2016) and NCKU-driver (Demirkus et al., 2016). The datasets had variations in illumination, back-

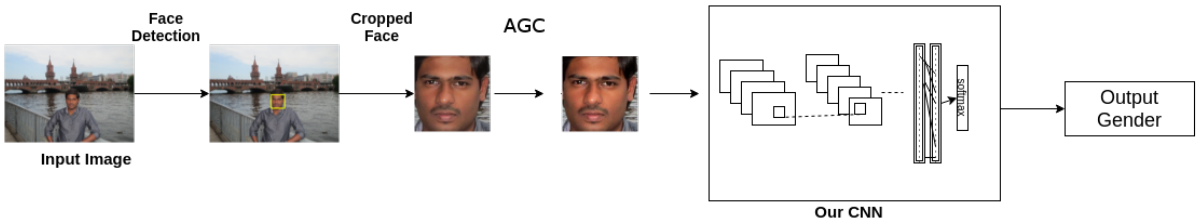


Figure 3: Proposed pipeline for gender detection for still images.

Table 1: Details of the Gender CNN.

Layer	Parameters
conv1	num_output: 96, kernel_size: 18, stride: 4
pool1, pool2, pool3	pooling type: MAX, kernel_size: 3, stride: 2
conv2	num_output: 256, pad: 2, kernel_size: 5, group: 2
conv3	num_output: 512, pad: 2, kernel_size: 3, group: 2
fc1	num_output: 6000
fc2	num_output: 2
norm1, norm2 (LRN layers)	local_size: 5, alpha: 0.0001, beta: 0.75

ground, head pose, facial movements and expressions. The method showcased an accuracy of 94.12% and 96.67%for McGillFaces and NCKU respectively.

3 DEEP LEARNING FOR GENDER RECOGNITION

3.1 Proposed Network

Deep and well known networks like AlexNet, LeNet, and others were designed for 1000-class classification problem in ImageNet (Large Scale Visual Recognition Challenge (Russakovsky et al., 2015)). The problem in hand is a binary classification problem, where it is required to know from the facial image, the gender of the subject. It is not a trivial task, however we work in the context of faces only. We propose to use a compact network, that consists of only 3 convolution layers followed by two fully connected layers. In the evaluation section, we show that we do not need a network with many convolution layers to recognize gender. We show the details of our proposed network in table 1.

The network takes facial images of size 200 x 200 pixels. The first convolution layer conv1 has 96 filters of size 18 x 18 pixels. The second convolution layer conv2 has 256 filters of size 5 x 5 pixels. The last convolution layer conv3 has 512 filters of size 3 x 3 pixels. The first fully connected layer has size of 6000, followed by the second layer fc2 of size 2. It represents male and females, and is followed by a SoftMax layer to predict the gender of the input image.

3.2 Gender Recognition in Still Images

In order to perform gender classification on still images, we use our proposed network model with Adaptive Gamma Correction as a pre-processing step. In figure 3, we show an overview of the pipeline. First, we detect the face in the input image using the face detector in (Mathias et al., 2014). The detected face is cropped, and pre-processed using AGC (Rahman et al., 2016). The pre-processed face is fed to the network and the gender is predicted. Using AGC helps in enhancing the image, in case that is required. If the image is already good enough, the AGC will not distort it with over or under-enhancing. More challenges exist in video frames and we discuss gender detection in video frames in the next subsection.

3.3 Gender Recognition in Videos

Videos captured using mobile phones, low resolution cameras or Point and Shoot cameras are challenging for gender recognition, due to the presence of varying illumination and motion blur. These challenges are not usual in still images. The presence of poorly illuminated frames, results in dark faces being fed to the machine learning algorithm. We tackle the illumination problem using the AGC, as we do in the still images pipeline. Motion blur is a common problem in videos captured by in an unstable manner, for example, handheld cameras, or where the subject is moving fast and the capturing device is not fast enough.

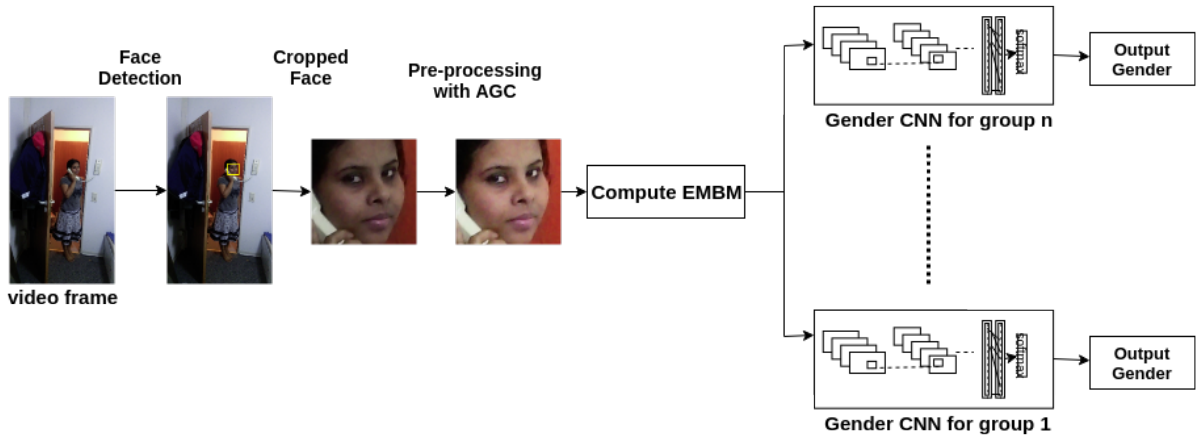


Figure 4: Gender detection for video frames, considering detected face EMBM value.

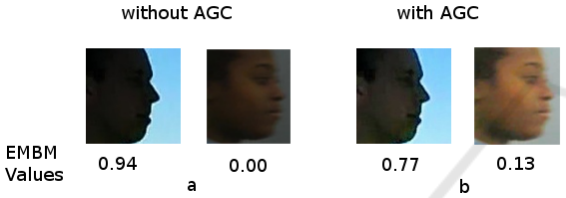


Figure 5: Effect of AGC on dark faces. **a** Original images and their EMBM (Guan et al., 2015) values. **b** AGC-enhanced (Rahman et al., 2016) images and their EMBM values. Faces extracted from PaSC(Beveridge et al., 2013).

We can clearly see these effects in the PaSC dataset, especially the handheld videos. In order estimate gender in bad quality or in blurred video frames, we use a blurriness metric to quantify the amount of motion blur. We use EMBM metric proposed in (Guan et al., 2015), to separate faces into groups according to the EMBM value. However, in case of dark faces, the EMBM value is mistakenly estimated to be very blurry, which is not true. The EMBM value can be also wrong in case of dark face with bright background, as the EMBM value estimates the sharpeness between the face and the background. We apply the AGC to enhance the facial image, and then evaluate EMBM. This gives a more accurate and reliable value of the EMBM as the details of the face are restored by AGC. Example images of evaluating EMBM with and without applying AGC are shown in figure 5. We can see in the figure that the AGC-enhanced images gives more reliable EMBM values.

After computing the EMBM value, we split the images into groups, and for each group, we use our proposed network model to estimate the gender for the face under its corresponding blurriness group. Based on our experimental results, we split the images into two groups based on EMBM threshold. The threshold value can differ based on the data used.

4 EVALUATION

In this section, we discuss the evaluation of our proposed method for gender classification. The next two subsections discuss the datasets used in our evaluation, and shows the results on these datasets. In our results we compare to state of the art methods, and we show that we either outperform other methods or we reach comparably close accuracy. Our experiment setup is as follows, we use 80% of the dataset for training and 20% for testing. We make sure that there is no overlap between the training and testing data. In order to have a fair comparison with other learning-based state of the art methods, we consider fine tuning their models similarly as we do for our model. However, our CNN model is trained from scratch.

4.1 Still Images

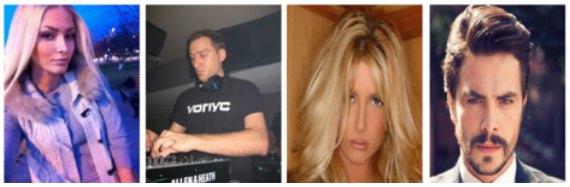
In 2015, Liu et. al. (Liu et al., 2015) introduced the CelebA dataset. The dataset consists of 202,599 images of 10,177 celebrities collected from the internet along with 40 facial attributes for each image. One of the attributes is gender.

Another still images dataset we work with is the IMDB-Wiki(Rothe et al., 2015) dataset. It was also introduced in 2015. It consists of 460,723 images from IMDB, and 62,328 from Wikipedia. The dataset is mainly intended for apparent age estimation, however, it also has the gender ground truth information. Sample Images from the datasets are shown in figure 7. We consider the images from IMDb and Wikipedia as separate datasets in our evaluations.

Experimental results are shown in table 3. We evaluated our proposed method on the following still images datasets, IMDB-Wiki (Rothe et al., 2015) and CelebA (Liu et al., 2015). We compare our results to

Table 2: Results and comparisons on still images datasets. Best result is shown in **Bold**, second best is shown in ***Bold Italic***.

	IMDB	Wiki	CelebA
Age & Gender CNN (Levi and Hassner, 2015)	87.2%	94.44%	97.27%
VGG face desc (Simonyan and Zisserman, 2014) & SVM	57.05%	81.37%	91.99%
LNet+ANet (Liu et al., 2015)	-	-	**98%**
GenderCNN wo AGC (Ours)	***87.34%***	***94.75%***	97.35%
GenderCNN w AGC (Ours)	**87.46%**	**94.76%**	***97.38%***



(a) CelebA (Liu et al., 2015)

(b) Wiki (Rothe et al., 2015)

(c) IMDB (Rothe et al., 2015)

Figure 6: Still images datasets samples.

other state of the art methods. We outperform other methods on the IMDB-Wiki dataset. On the CelebA dataset, the work in (Liu et al., 2015) still has the best accuracy value of 98%, however we score the second best accuracy in the table of value 97.38%. Another important point is the use of nearly 30k images less for training. In our work, we use a network that has 3 convolution layers, however, in (Liu et al., 2015) they use two deep CNNs to perform facial attribute predictions. We get the score of LNet and ANet (Liu et al., 2015) method from their paper (Liu et al., 2015). Preprocessing the images with AGC shows slight improvement in still images datasets. The next subsection discusses the videos datasets used and our evaluation results on them. The temporal coherent face descriptor cannot be applied on still images datasets, as it works only with videos.

4.2 Videos

The McGill (Demirkus et al., 2016; Demirkus et al., 2014; Demirkus et al., 2015) dataset, created in 2015, has a total of 60 videos of 31 female and 29 male subjects. Videos have 300 frames each with a resolution of 640 x 480, and varies in terms of location (indoor or outdoor), illumination conditions, head pose, occlusions and background clutter. Only a subset of 35 videos (10308 frames) have been shared with us. Figure 7(b) shows sample video frames from two separate videos.

The Point and Shoot Face Recognition Challenge (PaSC) (Beveridge et al., 2013) dataset, created in 2013, includes still images and videos. We perform our evaluations only on the videos from this dataset. The dataset has a total of 2802 videos of 265 different people shot at six different locations. There are equal number of control videos and handheld videos. The control videos were shot using a high quality camera at a resolution of 1920x1080 pixels (Full HD) with a tripod. The handheld videos were shot using various devices with resolution ranging from 640x480 to 1280x720. The dataset did not contain gender information needed for our evaluations. We were able to add the gender information manually. Of the 265 people, 143 were male and 122 were female. The varying locations (indoor and outdoor), resolution, illuminations, head poses, makes the PaSC (Beveridge et al., 2013) dataset ideal for evaluating gender recognition.

The evaluation results are shown in table 3. We evaluated the GenderCNN with and without AGC, along with AGC and EMBM GenderCNNs on the McGill and PaSC datasets. On the two datasets, we outperform the Temporal coherent face desriptor used by (Wang et al., 2015) for gender detection by a big margin. On the McGill dataset, using VGG face descriptor and a SVM (Simonyan and Zisserman, 2014) gives the best accuracy, however, we score the second best accuracy value. On the PaSC dataset, we outperform the state of the art methods we compared to, our proposed methods give the best and second best scores. Using GenderCNN with AGC pre-processing gave the best results on the control subset of the PaSC dataset. The experimental results show that on the challenging subset, the handheld, using the blurriness metric EMBM to train different CNNs gave the best result. Figure 1 shows gender detection on a sample frame from a handheld video with bright background.

Table 3: Results and comparisons on videos datasets. Best result is shown in **Bold**, second best is shown in ***Bold Italic***.

	McGill	PaSC Control	PaSC Handheld
Age & Gender CNN (Levi and Hassner, 2015)	84.4%	92.41%	92.62%
VGG face desc (Simonyan and Zisserman, 2014) & SVM	**92.13%**	90.45%	91.26%
Temporal gender detector (Wang et al., 2015)	71.42%	86.45%	88.04%
GenderCNN wo AGC (Ours)	90.14%	92.1%	92.84%
GenderCNN w AGC (Ours)	***90.4%***	**93.6%**	***94%***
AGC +EMBM GenderCNNs (Ours)	-	***93.31%***	**94.2%**

(a) McGill (Demirkus et al., 2016)

(a) PaSC (Beveridge et al., 2013)

Figure 7: Videos datasets sample frames.

The effect of AGC in improving the contrast of the face is visible.

5 CONCLUSION

In this paper a deep and compact CNN, with only 3 convolution layers was proposed for gender recognition from non aligned faces. We evaluated our GenderCNN on recent and challenging images and videos datasets. The still images used in our evaluation were IMDB-Wiki(Rothe et al., 2015) dataset and CelebA (Liu et al., 2015). We outperform the state of the art or match their accuracy values with small difference. In order to have a fair comparison, we trained the learning-based state of the art methods using the different datasets. We proposed using Adaptive Gamma Correction, AGC (Rahman et al., 2016) in our gender classification pipeline as a pre-processing step. It improved the results on still images datasets.

On the videos datasets, we proposed a different pipeline that tackles challenges found in videos. One main challenge was the image quality degradation due to blurriness. We propose quantifying the blurriness using EMBM (Guan et al., 2015) and training different CNNs based on this metric. We observed that poorly illuminated faces in video frames give misleading EMBM values, we tackled that using the pre-processing using AGC. Our proposed pipeline for gender detection in videos showed improvement on the challenging handheld videos of the PaSC dataset. The work presented in this paper opens the door for future ideas to investigate methods to tackle the blurriness found in the "in the wild" videos. One possible idea would be investigating the usefulness of deblurring techniques and their impact on gender classification of facial images analysis.

ACKNOWLEDGMENT

This work has been partially funded by the University project Zentrums für Nutzfahrzeugtechnologie (ZNT).

REFERENCES

Beveridge, J. R., Phillips, P. J., Bolme, D. S., Draper, B. A., Givens, G. H., Lui, Y. M., Teli, M. N., Zhang, H., Scruggs, W. T., Bowyer, K. W., Flynn, P. J., and Cheng, S. (2013). The challenge of face recognition from digital point-and-shoot cameras. In *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–8.

Demirkus, M., Precup, D., Clark, J. J., and Arbel, T. (2014). *Probabilistic Temporal Head Pose Estimation*

Using a Hierarchical Graphical Model, pages 328–344. Springer International Publishing, Cham.

Demirkus, M., Precup, D., Clark, J. J., and Arbel, T. (2015). Hierarchical temporal graphical model for head pose estimation and subsequent attribute classification in real-world videos. *Computer Vision and Image Understanding*, 136:128 – 145. Generative Models in Computer Vision and Medical Imaging.

Demirkus, M., Precup, D., Clark, J. J., and Arbel, T. (2016). Hierarchical spatio-temporal probabilistic graphical model with multiple feature fusion for binary facial attribute classification in real-world face videos. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 38(6):1185–1203.

Eidinger, E., Enbar, R., and Hassner, T. (2014). Age and gender estimation of unfiltered faces. *Trans. Info. For. Sec.*, 9(12):2170–2179.

Felzenszwalb, P. F., Girshick, R. B., McAllester, D., and Ramanan, D. (2010). Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(9):1627–1645.

Guan, J., Zhang, W., Gu, J. J., and Ren, H. (2015). No-reference blur assessment based on edge modeling. *J. Visual Communication and Image Representation*, 29:1–7.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition. *CoRR*, abs/1512.03385.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

Levi, G. and Hassner, T. (2015). Age and gender classification using convolutional neural networks. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) workshops*.

Liu, Z., Luo, P., Wang, X., and Tang, X. (2015). Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*.

Mansanet, J., Albiol, A., and Paredes, R. (2016). Local deep neural networks for gender recognition. *Pattern Recogn. Lett.*, 70(C):80–86.

Mathias, M., Benenson, R., Pedersoli, M., and Van Gool, L. (2014). Face detection without bells and whistles. In *ECCV*.

Ng, C. B., Tay, Y. H., and Goi, B.-M. (2012). Vision-based human gender recognition: A survey. *CoRR*, abs/1204.1611.

Parkhi, O. M., Vedaldi, A., and Zisserman, A. (2015). Deep face recognition. In *British Machine Vision Conference*.

Rahman, S., Rahman, M. M., Abdullah-Al-Wadud, M., Al-Quaderi, G. D., and Shoyaib, M. (2016). An adaptive gamma correction for image enhancement. *EURASIP Journal on Image and Video Processing*, 2016(1):35.

Ranjan, R., Patel, V. M., and Chellappa, R. (2016). Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *CoRR*, abs/1603.01249.

Rothe, R., Timofte, R., and Gool, L. V. (2015). Dex: Deep expectation of apparent age from a single image. In *IEEE International Conference on Computer Vision Workshops (ICCVW)*.

Rothe, R., Timofte, R., and Gool, L. V. (2016). Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision (IJCV)*.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, pages 511–518.

Wang, W. C., Hsu, R. Y., Huang, C. R., and Syu, L. Y. (2015). Video gender recognition using temporal coherent face descriptor. In *2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, pages 1–6.