Attribute Classification in Low resolution

1st Akarsha Sehwag
IAB lab, CSE Dept
IIIT Delhi
Delhi, India
akarsha15010@iiitd.ac.in

2nd Dr. Richa Singh
IAB lab, CSE Dept
IIIT Delhi
Delhi, India
richa@iiitd.ac.in

3rd Dr. Mayank Vatsa
IAB lab, CSE dept
IIIT Delhi
Delhi, India
mayank@iiitd.ac.in

***Abstract*—This report contains a summary of the literature review and implementation of baseline models for the Btech project, titled Attribute Classification in low-resolution.**

***Index Terms*—Attribute Classification , Low Resolution, SVM, Multi-Task Convolution Neural Network, Face recognition**

I. Introduction

Humans focus attention on different face regions to recognize face attributes and most of the algorithms previously used consider the entire face image as an input to the algorithm. Semantic features (a.k.a Attributes) have been actively used in the Biometrics industry for more than ten years for domains ranging from activity recognition to face verification. And for all of these applications, attribute classification needs to be done as the initial step. Metrics show that more than 90% of the images can be correctly classified using the existing algorithms.

But the problem we identified is that these algorithms are not good enough to classify images in low resolution without a huge false positive and false negative rate. For practical applications, eg. Identification of people in real-life datasets like Missing Persons Dataset (published by National Crime Records Beureau) is not easy because most of them are low-quality clicked images where the accuracy is not even close to the accuracies tested on pre-defined databases. In the report I establish a baseline for my Btech Thesis wherein I create a Multi-Task Convolution Neural Network for Attribute classification and then try to improve the accuracies on low-quality input image.

II. Literature Review

Ample of work has been done in Attribute classification on datasets as large as CELEBA. But all these datasets contain high-resolution images which can give us accuracies more than 90% but when it comes to real-life implementation of such algorithms such as surveillance videos or poor quality phone images, their performance goes down and contain multiple false positives or false negatives.

A. Attribute Classification

Kumar et Al [2] proposed using face attributes for face recognition and visual search and infact, was the first one to use attributes as image descriptors for face verification [5] . They downloaded face images from the internet, labeled them with 65 binary attributes such as gender,hair color, visible forehead, nose shape, etc and present a pub-fig dataset with 60,000 images.

Convolution Neural networks have replaced most of the algorithms after 2013-14. And Deep Learning came even later in around 2015. [3] proposed a deep learning framework which cascades two CNNs, LNet and ANet, all differently trained but fine-tuned jointly with attribute tags. In that paper, the authors use LNet for face localization and ANet, which was pre-trained using face identities has been used for attribute prediction. In [6], published in 2017, the authors propose to use a cascade network to learn localization of face region specific to attributes and classify attributes on unaligned images.

[7] attempts to build a regression model for Facial Attribute Classification in 73 Facial Attributes. They use the real-valued scores (predicting Facial Attributes) to determine the probability of a particular attribute in the input image after which two dictionary learning methods are used to learn regression and then further a multi-level feature extraction was proposed for classification. [9] test the robustness of the DNNs and used fast-flipping attribute technique to generate adversarial examples than existing algorithms and test the correlation of facial attributes to find that only related attributes change the classification of others. Until 2017, the attributes were generally considered to be independent [10]. Others like [11] did Pose Aligned Networks for Deep Attributes (PANDA) focusing on gender and age.

B. Low Resolution

As per my knowledge, low-resolution images have only been considered for attribute classification on complete human body and not just faces. I, through my Btech Thesis plan to work on face attribute classification in low resolution. In the previous work, [4] proposes a part-based approach based on poselets to recognize attributes like gender, hair style and types of clothes in unconstrained environments tested under large variation in pose, articulation, viewpoint and occlusion. They are one of the very few people to work in low resolution as well but they don't use face attributes for classification. [8] attempts to detect people attributes like Gender, clothing, race, etc. in the context of dealing with surveillance footages. They use Deep Convolutional Generative Adversarial Networks (DCGAN) and successfully extracted attributes for low quality images or occluded images by combining Generative reconstruction and deep attribute classification network.

Fig. 1. Large-scale CelebFaces Attributes (CelebA) Dataset: 10,177 identities, 202,599 face images, 5 landmark locations and 40 binary attributes annotations per image

Fig. 2. Labeled Faces in the Wild dataset : 13233 images of 5749 people and 76 attribute scores

III. Datasets

We use the two most challenging datasets available: Labeled Faces in the Wild and CelebA

IV. Implemeted Algorithms

Two algorithms have been implemented till now to create a baseline for the Btech Thesis.

V. SVM based Classifier

[2] presented two novel methods in 2009 for attribute classification and face verification. The attribute classifier would get binary output which shows the absence or presence of the the visual attribute. They tested their algorithm on LFW which was the most challenging dataset then with manifold variations in the dataset images. They try to learn the high level visual features/traits which are insensitive to pose, illumination, expression etc.

A. Architecture

A binary classifier has been trained to recognize the presence and absence of a particular attribute eg. gender, hair color, age, race etc. which is insensitive to environment changes. This has been compared with human performance. We start with extracting Low level features from the image. First, we detect fiducial point locations using OpenCV. Then pick a face region (manually labeled) and extract a feature-type and normalize it by either subtracting the mean and dividing it my std. deviation or perhaps by just dividing by the mean. Finally, aggregate or concatenate these values to construct the low level feature vector. Spatial regions are manually labelled different part of face images such as Eyes, nose, forehead etc. Thus, several attribute classifiers are trained with the positive and negative set of attributes as shown in fig.2. It then uses a simplified AdaBoost algorithm to learn the set of low level features from the given set using forward feature selection in which we keep on appending one feature vector and notice the change in the error rate. We repeat this until our change in error rate is less than a given threshold. I used SVM with RBF kernel trained using sklearn SVC. Table 2 shows the results of the SVM classifier in the given paper vs my results.

VI. CNN based Classifier

Here, an architecture similar to the Multi-task Convolution Neural Network [1] has been implemented to obtain an architecture for baseline since the original implementation was not available opensource. The architecture is training just one single attribute network and shares information throughout the network. It learns the relationship for 40 defined attributes.

A. Architecture

Multi-task learning is a subield of ML where multiple learning tasks are solved at the same time. It aims to improve the performance of multiple classification tasks by learning them jointly. Multi-task learning has been successful in pose estimation, face detection and multiple other domains. We're applying MTL to attribute prediction since the relationship between the facial attributes is strong.

The MCNN architecture as given in Fig 4 consists of three convolution layers followed by two fully connected layers resulting in 40 seperate attribute scores which is further thresholded to get binary outputs. In pre-processing of the images and the attributes, the image pixel value has been normalized from 0 to 1 from 0 to 255 and similarly, the attribute values are normalized from 0 to 1. Apart from this, data needs to be annotated 5 times, I am using annotations like Flip, Blur, Jittering etc. The architecture is explained in the serialized form:

1. Convolution Layer 1 - 75 convolution filters (7x7)
- applied ReLU, 3x3 MaxPooling
2. Convolution Layer 2 - 200 convolution filters (5x5)
- applied Relu, 3x3 MaxPooling
Groupings into 6 are done before applying convolution layer
3. Convolution 3 layers - one for each group i.e. Gender,

TABLE I
MCNN RESULTS ON LFWA DATASET

No. of Epochs	Output	
	Average Accuracy	*Avg. Loss*
5	58.35%	0.02[a]
10	61.97%	0.018[a]
20	63.66%	0.01[a]

[a]The results are for unaugmented data.

Nose, Mouth, Eyes, Face and Others - this layer has 300 convolution filters (3x3)
- applied ReLU, 5x5 MaxPooling
4. Each Conv3 layer is connected to a Fully Connected Layer (FC-1) with 512 units - Conv3 layer for the group, Others, is connected to 4 FC-1s i.e one for each AroundHead, Facial Hair, Cheeks and Fat.
- applied DropConnect with keep probability as 0.5
5. All 9 FC-1's are connected to another fully connected layer (FC-2) which is then fully connected to the binary output nodes.
- applied DropConnect with the probability as 0.5.

The attributes for each group with the corresponding names are given below:

- **Gender** - Male
- **Nose**: Big Nose, Pointy Nose
- **Mouth**: Big Lips, Smiling, Lipstick, Mouth Slightly Open
- **Eyes**: Narrow Eyes, Arched Eyebrows, Bushy Eyebrows, Eyeglasses, Bags Under Eyes
- **Face**: Attractive(Man & Woman), Oval Face, Young, Pale Skin, Blurry, Heavy Makeup
- **AroundHead**: Blond Hair, Brown Hair, Black Hair, Gray Hair, Straight Hair, Wavy Hair, Necklace, Necktie, Earrings, Balding, Receding Hairline, Hat, Bangs
- **FacialHair**: Mustache, Sideburns, No Beard, Goatee, 5 oclock Shadow
- **Cheeks**: High Cheekbones, Rosy Cheeks
- **Fat**: Double Chin, Chubby

The groups were based on spatial datapoints on the face image. I used Pytorch for my implementation which is a python package providing two high level features: Tensor computation with GPU acceleration and it provides maximum flexibility and speed. Training same CNNs independently would result in 40 trained CNNs which would require high computation power and days of training time.

VII. RESULTS

Results for SVM based classification are pretty close to the baseline model. The difference arise since they used commercial systems to obtain the fiducial points which might be more accurate than an opensource model. And for the MCNN model, the results are yet not comparable since we need to augment the data to avoid overfitting. Apart from this, I am yet to test it with 50 epochs to make the results comparable to the paper.

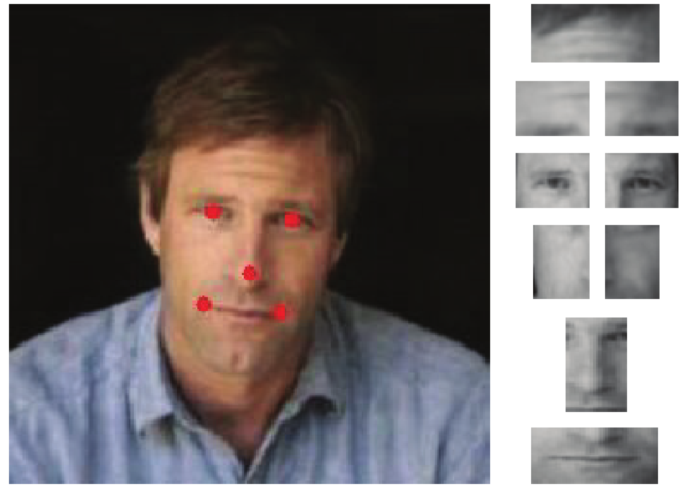

Fig. 3. Fiducial points and facial region extraction for training an SVM

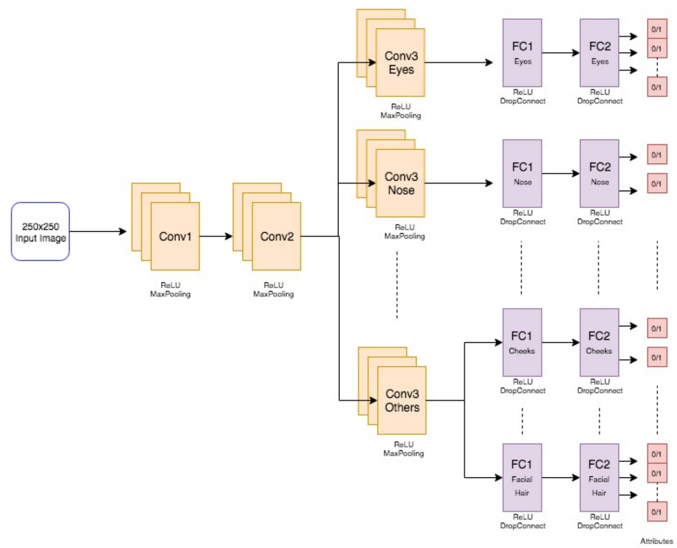

Fig. 4. Multi-Task CNN Architecture

CONCLUSION

In my first semester of Btech Project, I have done the following tasks:
- Deep understanding of the problem statement.
- Identify Challenges for the project.
- Thorough Literature Study.
- Implemented Baseline Models using SVM and MCNN.

REFERENCES

[1] Emily M. Hand and Rama Chellappa: Attributes for Improved Attributes: A Multi-Task Network for Attribute Classification
[2] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and Simile Classifiers for Face Verification. In ICCV, 2009
[3] Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. ICCV (2015)
[4] Lubomir Bourdev, Subhransu Maji and Jitendra Malik: Describing People: A Poselet-Based Approach to Attribute Classification. ICCV (2011)

TABLE II
ATTRIBUTE CLASSIFICATION USING SVM

Attributes	**Paper**	**MyResults**	**Attributes**	**Paper**	**MyResults**
Outdoor	84.8	64.38	Baby	90.45	99.54
White	91.48	89.59	Child	83.58	95.36
Senior	88.74	82.06	Middle Aged	78.39	90.07
No Beard	89.53	70.95	Round Jaw	66.99	86.98
Teeth Not Visible	91.64	77.38	Mouth Wide Open	89.63	90.52
No Eyewear	93.55	79.42	Wearing Hat	85.97	86.25
Eyeglasses	91.56	82.33	Male	81.22	76.96
Smiling	95.33	88.7	Brown Hair	72.42	62.2
Frowning	95.47	75.84	Mustache	91.88	87.16
Receding Hairline	84.15	75.79	Asian	92.32	92.17
Bushy Eyebrows	93.4	53.47	Chubby	77.24	63.84
Big Nose	87.5	68.94	Blurry	92.12	84.43
Mouth Closed	89.27	62.39	Harsh Lighting	78.74	67.94
Mouth Slightly Open	85.13	66.85	Curly Hair	68.88	59.83
Black Hair	80.32	86.34	Wavy Hair	64.49	57.38
Blond Hair	78.05	95.53	Straight Hair	76.81	63.21
Bald	83.22	89.61	Bangs	88.7	83.6
Sunglasses	94.91	98.63	Sideburns	71.07	69.03
Flash	72.33	81.15	Visble Forehead	77.02	68.4
Soft Lighting	67.81	70.12	Arched Eyebrows	80.9	73.96
Part.Visible Forehead	77.02	94.17	Eyes Open	92.52	86.25
Obstructed Forehead	79.11	85.98	Pointy Nose	86.87	69.77
Black	88.65	95.45	Big Lips		63.11
Goatee	80.35	75.44	Youth	85.79	82.79

[5] Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K.: Describable visual attributes for face verification and image search. PAMI (2011)

[6] Hui Ding, Hao Zhou, Shaohua Kevin Zhou, Rama Chellappa : A Deep Cascade Network for Unaligned Face Attribute Classification

[7] Ali Moeini, Hossein Moeini, A.M.Safai,Karim Faez : Regression Facial Attribute Classification via simultaneous dictionary learning, Elsevier,2017

[8] Matteo Fabbri,Simone Calderara, Rita Cucchiara : Generative adversarial models for people attribute recognition in surveillance, IEEE, 2017

[9] Andras Rozsa, Manuel Gunther , Ethan.M.Rudd, Terrance.E.Boul : Facial Attributes: Accuracy and Adversarial Robustness, Pattern Recognition Letters, 2017

[10] Zhang, N., Paluri, M., Ranzato, M.A., Darrell, T., Bourdev, L.: Panda: Pose aligned networks for deep attribute modeling. CVPR (2014)

[11] Levi, G., Hassner, T.: Age and gender classification using convolutional neural networks. CVPR (2015)

[12] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. ACM Comput. Surv., 35(4):399458, 2003.

[13] A. OToole, P. Phillips, F. Jiang, J. Ayyad, N. Penard, and H. Abdi. Face recognition algorithms surpass humans matching faces over changes in illumination. PAMI, 29(9):16421646, Sept. 2007

[14] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski. SexNet: A neural network identifies sex from human faces. In NIPS, pages 572577, 1990