Table 3: Facial attributes classification Accuracy (%), where # denotes number of classes, ROI is facial region used.

Face att.,#	ROI	Descriptor	Accuracy
Skin color,4	R1	RGB-Hist, HOG	87.20
Face shape,3	R1	HOG	89.95
Eye Shape,6	R2	HOG,LBP	61.05
Lips Shape,3	R4	HOG,LBP	79.41
Eye Color,5	R3	RGB-Hist	81.00
Ethnicity,4	R1	RGB-Hist, LBP	87.42

matically. To this end, 83 facial landmarks are detected on the face using face++ framework[2] and different regions of interest are extracted for different facial attributes as illustrated in Fig.3. We can see that the whole facial region called R_1 is used for skin color, face shape and ethnicity classification, the R_2 region for eye shape, R_3 for eye color and R_4 for lips shape. After cropping the region of interest for a certain attribute, a combination of color and shape descriptors like RGB-Histogram of 8 bins, HOG (Dalal and Triggs 2005) and LBP (Ojala, Pietikainen, and Maenpaa 2002) are selected empirically to extract the best feature vector for every attribute. After data reduction and noise removal the resulting feature vector is passed to multi-class SVM classifier (Chang and Lin 2011). These experiments are conducted on 900 before makeup facial images following 10-fold cross validation. Table 3 summarizes the facial attribute, and the facial region of interest cropped, descriptors, and the average classification accuracy. We obtained a good classification rate for most of the attributes as presented in Table 3. For the eye shape, it is 61.05%, there are 6 different classes and it is a challenging task even for people.

Statistical evaluation

Experimental settings: the statistical evaluation of the proposed system is conducted on 961 pairs of images from our collected database. 80% pairs of images (examples) are used for training, 10% for validation and 10% for testing in 9-fold cross validation. Mini-batch gradient descent algorithm (Vincent et al. 2010) is used for more robust gradient descent performance with min-batch size 10. Number of epoches in the training is 100 and learning ration $\beta = 0.1$ selected empirically. The network has one input layer, 3 hidden layers each has 100 hidden units, learning rate: $\eta = 10^{-4}$, and one output layer with 8 different outputs (*softmax*). The class number for every facial trait is repeated 10 times to make feature vector of size 10 and the six concatenated in one feature vector V of size 60 to serve as input for the model.

Experiment: to validate the merit of combining rules and examples together in training, we trained our system with examples alone and we applied the same loss function on the suggested makeup from the rule-based recommendation system to compare both of them statistically with Example-Rules guided system. In training, we used the labeled values for the facial traits and for testing the automatic facial traits classification is applied. To compare with state of the art, we compared with distance-based similarity makeup recommendation approaches followed in (Liu et al. 2016), (Scherbaum et al. 2011) where *Deep* features and *Eigen* features are used respectively to compute similarity metric between the test face and available images in the dataset. We repeated these two methods on every 100 testing images (without makeup) and computed the lose between the closest face makeup style and the makeup style of the testing image using the same loss functions used in our deep learning model. The statistical loss values are reported for in Table 4.

Table 4: Statistical results of the loss values for each makeup element. **Eigen:(Scherbaum et al. 2011)**, **Rule:** Rule-based recommendation, **Exp:** Examples trained network, **Deep:(Liu et al. 2016)** and **Exp-Rul:** Examples-Rules Guided network. # indicates to the number of the classes of the makeup element. In makeup elements, *C* denotes to Color.

Makeup,#	Eigen	Rule	Exp	Deep	Exp-Rul
Foundation, 4	0.37	0.55	0.42	0.23	**0.1**
Lipstick C, 5	0.50	0.62	0.45	0.47	**0.31**
Lip liner, 2	0.40	0.32	0.23	0.35	**0.20**
Blush, 3	0.27	0.02	0.02	0.19	**0.01**
Blush C, 8	0.55	0.52	0.47	0.56	**0.34**
Eyeshad, 6	0.53	0.65	0.57	0.60	**0.32**
Eyeshad C, 5	0.70	0.59	0.45	0.67	**0.36**
Eyeliner, 3	0.37	0.48	0.39	0.32	**0.27**
Average	0.46	0.43	0.38	0.32	**0.24**

We can see from these statistical results that the combination of rules and examples gives the lowest loss values for every makeup element and it is less than the two state of the art methods **Deep** and **Eigen**. Also, it is less than Rule-based recommendation and Examples-alone trained system. These results approve our hypothesis about the advantage of combining the rules and examples to learn the model parameters for makeup recommendation and shows the superiority of this method over the state of the art similarity based makeup recommendation methods. Also, the makeup elements which are related strongly to the facial traits such as foundation tone, lip liner and blush style loss are less than other makeup elements such as lipstick color, eye shadow colors which may have more than one good choice.

Model parameters analysis

To investigate the effect of the homogeneity term in the cost function, we repeated the same statistical analysis experiments with and without this term and the results showed that loss is less for every makeup and by 0.9 in average with this term by comparison with the case without it. This demonstrates the importance of enforcing homogeneity between different makeup style elements. Also, we compared the proposed network structure single network multiple outputs (SNMO) against using multiple networks single output (MNSO) for every makeup style independently. From the obtained results in Table 5, we can see that the average loss for MNSO is higher than SNMO adopted structures since we lose the ability to enforce homogeneity among multiple

[2] www.faceplusplus .com