Table 5: SNMO structure vs. MNSO comparison

	MNSO	SNMO (ours)	Difference
Average Loss	0.31	0.24	0.7 less
# of Paramters	208,800	26,800	88% less
# of Epoches	50	100	2× more

outputs of the same network. Also, the total number of parameters of the 8 networks (MNSO) is much higher than one network with 8 outputs (SNMO), but it requires half the number of epochs to train for every single network alone.

Perceptual evaluation

To evaluate the performance in perceptual manner, we designed two qualitative experiments. First, we randomly selected 20 images without makeup from our collected dataset and applied the automatic facial attribute classification and obtained the recommended makeup style and synthesized it on the images. We presented two photos of those 20 subjects without and with recommended makeup to 20 persons (10 males and 10 females from different cultures). We asked the participants to give an evaluation for every makeup style as {Very bad, Bad, Fine, Good, Very good} and reported the percentage that every evaluation obtained. The evaluation of females and males are presented separately. Perceptual survey results are given in Fig.4(A) where we obtain the highest score for *fine* evaluation and we obtained *Good and Very good* higher than *Bad and Very bad*. The evaluation is positive and it shows that our recommendation and implementation is good from the view point of human perception.

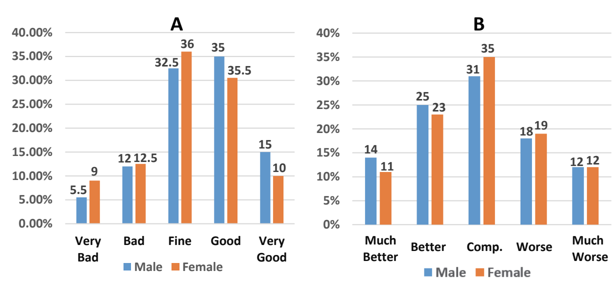

Figure 4: Male-Female qualitative evaluation experiments

In Fig.5(A), there are three samples of the 20 testing images used for the experiment. The face before makeup, with professional makeup, with suggested makeup are presented.

The second experiment is more challenging, where we compared our suggested makeup with professional one. We showed 3-tuple contains: *before, professional, our* makeup for the same participants in last experiment. We exchanged randomly between the position of our and the professional makeup and asked: *Do you think that the left is (Much worse, Worse, Comparable, Better, Much better) than the right makeup?*. The statistical results are presented in Fig.4(B). From the obtained results, we can see that the evaluation *Comparable* has the highest evaluation from males and females, and the evaluations *Much better, Better* got more votes than *Worse, Much worse* too. These two experiments demonstrate the efficiency of our proposed makeup recommendation system versus professional makeup images from the view point of the end user.

Makeup synthesis results

To demonstrate the efficiency of our makeup synthesis implementation as illustrated in Fig.5(B), we compare our synthesis results with two main makeup synthesis websites *TAAZ* and *DailyMakever*. From this figure, we can see that in TAAZ, it is not possible to work on the eye brows, and the eye lashes effect is not natural. In *Dailymakeover*, the ability to control the effect intensity is limited, and the lips shape detection is not accurate. We can see here the positive effect of using different blending types and combining two types in some cases in makeup implementation. For example, we have a natural effect for foundation and blush that requires homogeneous blending with the skin, and have elegant eye shadows and eye lashes effect where the contrast with the nearby facial area need to be preserved. Besides, our synthesis is fully automatic where these two websites require manual intervention. Our makeup synthesis system accuracy is higher in spite of it is fully automatic where TAAZ and Daily Makeover requires user intervention.

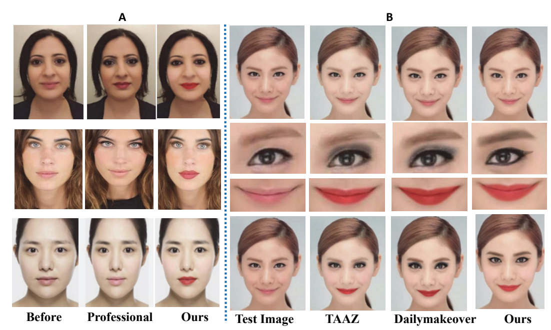

Figure 5: A) Three samples show the face before, after professional and after our makeup. B) Comparison of our synthesis results with TAAZ and DailyMakeover. From top to down, foundation, eye shadow, lipstick, blush, overall effect.

Conclusion and Future Work

In this paper, a deep neural network based makeup recommendation model is trained from examples and knowledge base rules jointly. We demonstrated its ability to recommend homogeneous makeup style that fits face according to its automatically classified facial traits. The recommended makeup style can be synthesized efficiently as well. Another contribution of this work is the Before-After makeup database. This system can be improved by several aspects which we consider them as a future work like extending the database for more robust learning and evaluation, make the recommendation flexible for new trends, and able to recommend different hairstyles and accessories, enriching the makeup synthesis system by adding more trends, templates and colors to have richer suggestions. Also generalizing the proposed approach beyond the makeup recommendation.