1 Summary

The paper talks about the problem of Domain Transfer where, given a source domain S and a target domain T, we want to learn a generative mapping G that maps an input sample from S to the domain T. Authors propose deep neural networks of specific structure in which G is composition of an encoder **f** that identifies a common feature space between two domain and a decoder **g** that generates samples in target domain. The authors implement their model for the problem of emoji generation for a given facial image and produce emojis that are visually appealing.

2 Strengths

- The representation of G as composition of g and f is novel. The advantage is that it reduces the learning effort for g, which helps in lowering the reconstruction loss. Low reconstruction loss is pretty important for good model performance as seen in their ablation study.
- For domain transfer, this method does not require paired training images from the two domains.

3 Weaknesses

- The architecture provided is asymmetrical, in the sense that it can only transfer images from source domain to target domain and not vice versa.
- The function f needs to be pre-trained on a different dataset on some task that is helpful in encoding information for the domain transfer task. This seems a painful process and it would have been better if an end-to-end training process for the specific transfer domain task could be fabricated.

4 Critique of Experiments

- The authors first tackle the problem of domain transfer from SVHN to MNIST dataset. Neural network representing function f is first trained for classification task on SVHN dataset and is incorporated in DTN architecture. A classifier trained on MNIST dataset is used to classify samples transferred by DTN from SVHN to MNIST which gives an accuracy of 90.66%. Despite not being accurate, DTN performs visually appealing domain transfer.
- For the second task, the authors perform photos to emoji domain transfer. The representation layer of DeepFace network is taken as function f. Original celebA images are compared with human generated emoji and DTN generated emoji. It can be visually verified that DTN generated emojis tend to be more informative and personalized.
- Retrieval accuracy out of 10^5 distractors is compared for manually created emojis as well as by DTN network. Emojis by DTN obtain a median rank of 16 compared to 16311 for manually created emojis.

5 Follow Ups/Extensions

- The model can be made symmetric using cyclic consistent losses so that images from target domain can also be mapped to source domain.
- As an extension to experiments, it would be interesting to see how the model performs on tasks such as street scene image translation. It also feels like the experiments given in the paper are run only on small resolution images. Quality of transferred images to target image for high resolution images using this model would be interesting to see.