Visual Siamese Clustering for Cosmetic Product Recommendation

Christopher J. Holder[1], Boguslaw Obara[1], Stephen Ricketts[2]

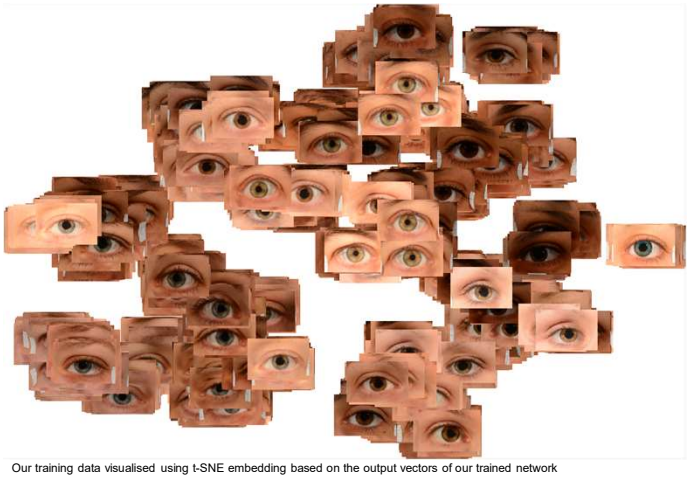
[1]Department of Computer Science, Durham University, UK
[2]Walgreens Boots Alliance

Example images of subjects not in our training dataset (top) along with images of the training subjects determined by our model to be the closest matches (bottom)

Motivation

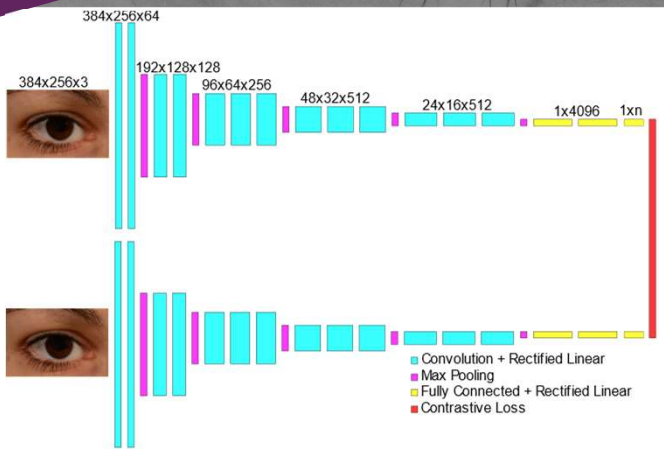
- Recommender systems typically require large quantities of data on a customer's past habits
- For certain products where aesthetics are important, visual information may be a useful substitute
- We propose a visual recommender system for cosmetics retail that makes mascara recommendations based on the preferences of others who are visually similar

Approach



Our training data visualised using t-SNE embedding based on the output vectors of our trained network

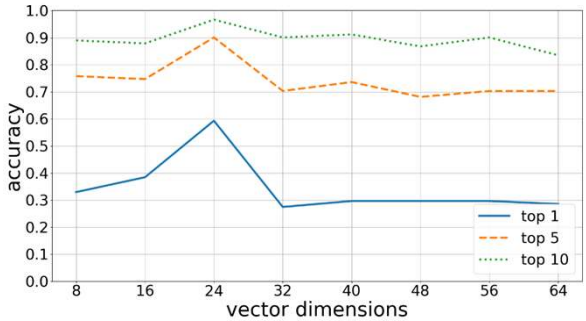
- We train a Siamese convolutional neural network to create a multi-dimensional embedding of subjects based on eye image features
- New subjects can be matched to existing subjects who are visually similar based on their closeness within this embedding space
- Network model is based on VGG16, using a contrastive loss function to optimise such that images of eyes belonging to the same person will output similar feature vectors



The network model used is VGG16 in a Siamese configuration with a contrastive loss function, with the aim being to optimise such that two input images of the same subject will output vectors that are close together, while images of different subjects will output vectors that are further apart.

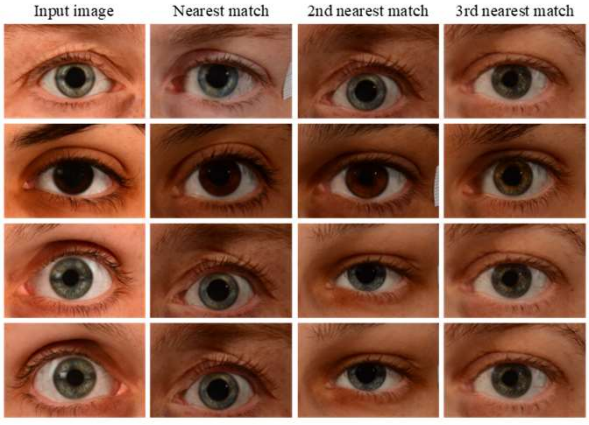
Evaluation

- We train using multiple images of the eyes of 91 women
- 7 images of each subject are used for training, quantitative evaluation is based on a model's ability to match an 8th unseen image to the correct subject
- Models with output dimensionality between 8 and 64 are evaluated
- We record each model's ability to identify the correct subject within top 1, top 5 and top 10 closest matches



We evaluate models with different output dimensionality based on their ability to match images to the correct subjects

- Qualitative evaluation is conducted using images of unseen subjects



The top 3 closest matches suggested by our model for 4 images of unseen subjects.