Supplementary Material for Multimodal Prediction and Personalization of Photo Edits with Deep Generative Models

Ardavan Saeedi
CSAIL, MIT

Matthew D. Hoffman
Google Brain

Stephen J. DiVerdi
Adobe Research

Asma Ghandeharioun
Media Lab, MIT

Matthew J. Johnson
Google Brain

Ryan P. Adams
Princeton and Google Brain

1 Details of Experiments

Hyperparameter settings For training all the models, we use two possible learning rates 0.001 and 0.0001. For the MDN, MLP and CGM-VAE, we use 4 hidden layers with the same number of hidden nodes (500 or 1000) in all the layers. For LBN, we use two deterministic hidden layers with linear activation function same as Dauphin and Grangier (2015) and 500 or 1000 nodes; we also use two stochastic layers with the same number of nodes with sigmoid activation functions following Dauphin and Grangier (2015). We try two possible minibatch sizes of 100 and 200. For the models which need the number of mixture components (*i.e.,* CGM-VAE and MDN), we select this number from the set {1, 3, 5, 10}. Finally, for the CGM-VAE model, we choose the dimension of the latent variable from {2, 20}. We choose the best hyperparameter setting based on the variational lower bound of the held-out dataset.

Baselines We choose a set of reasonable baselines that can cover related models in both domains of multimodal prediction and automatic photo enhancement. As mentioned in Section 2, literature on the automatic photo enhancement can be divided into two main categories of models: 1) parametric methods which typically minimize an MSE loss: MLP and MDN baselines capture these methods, and 2) nonparametric methods that are not reasonable baselines for us since their proposed edits are destructive (*e.g.,* Lee et al., 2015) or do not benefit from other users' information (*e.g.,* Koyama et al., 2016). We also add LBN as a strong baseline since it has been shown that it can outperform the MDN and other standard multimodal prediction baselines (Dauphin and Grangier, 2015) .

Splitting the datasets We split each dataset into random subsets of train, validation and test in a way that for all three datasets (*i.e.,* casual, frequent and expert users) we have subsets with reasonable sizes. Larger training sets may result in non-representative validation and test sets for our small dataset (*i.e.,* expert users), and larger test and validation sets may result in non-representative training set for the same dataset. The ratio that we used is just one way for having a reasonable size subsets; however, we showed for these random subsets and across all three datasets and for three different evaluation metrics our approach outperforms all the other baselines significantly.

To apply the P-VAE model to the experts dataset, we split the image-slider combinations from each of the 5 experts into groups of 50 image-sliders and pretend that each group belongs to a different user. This way we can have more users to train the P-VAE model. However, this means the same expert may have some image-sliders in both train and test datasets. The significant advantage gained in the experts dataset might be due in part to this way of splitting the experts. Note that there are still no images shared across train and test sets.

2 Sample edits from CGM-VAE and sample user categorization from P-VAE

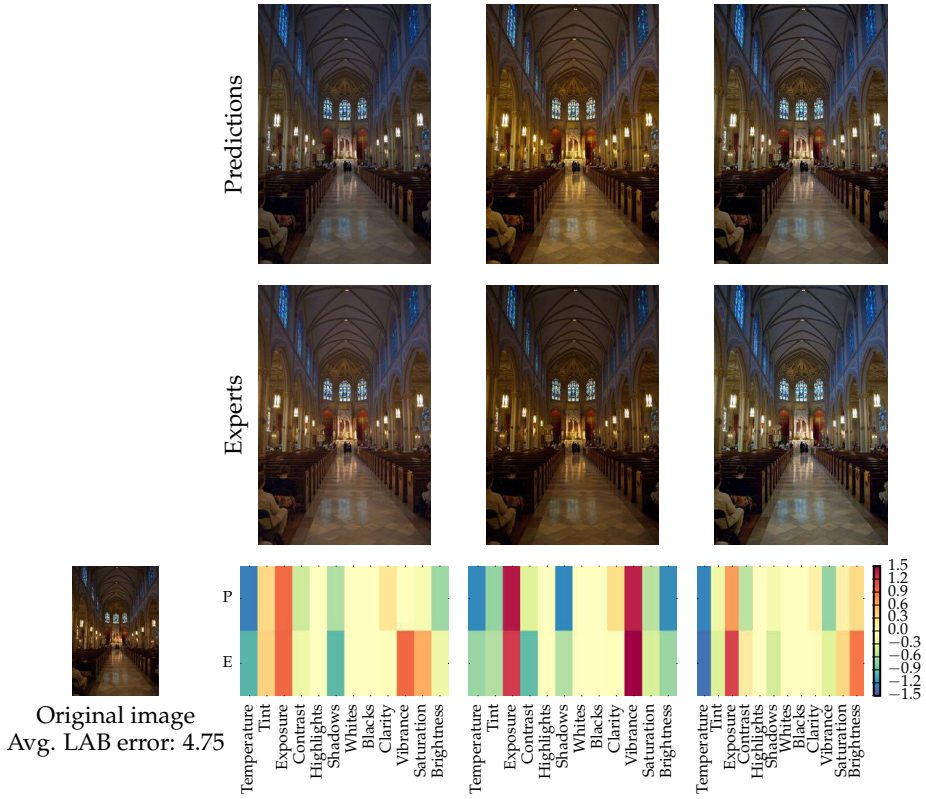

Figure 1: Image 4876 from Adobe-MIT5k dataset

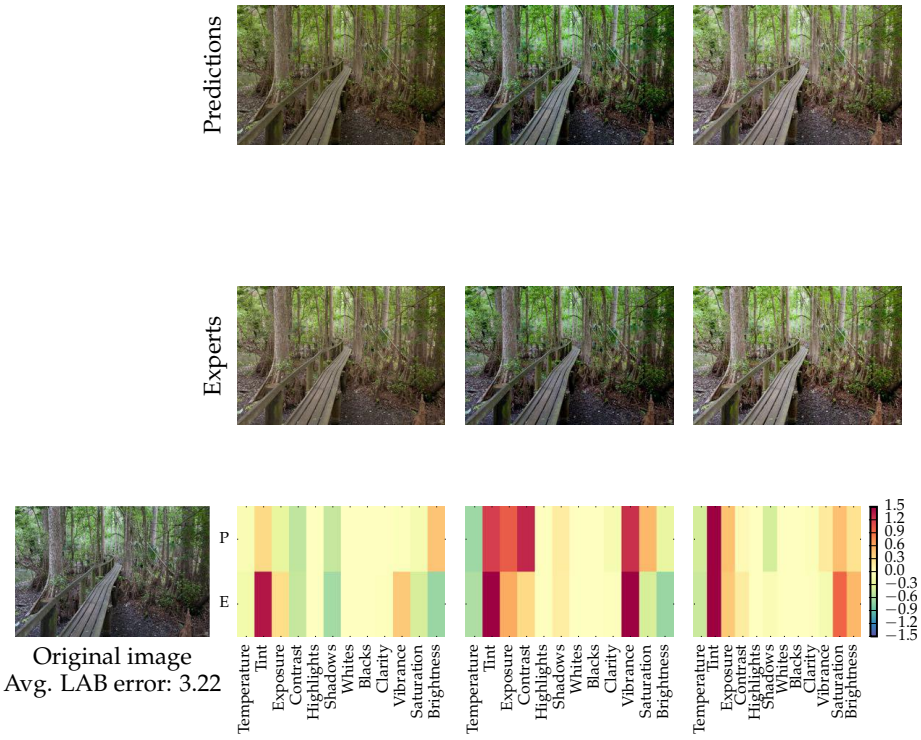

Figure 2: Image 4855 from Adobe-MIT5k dataset

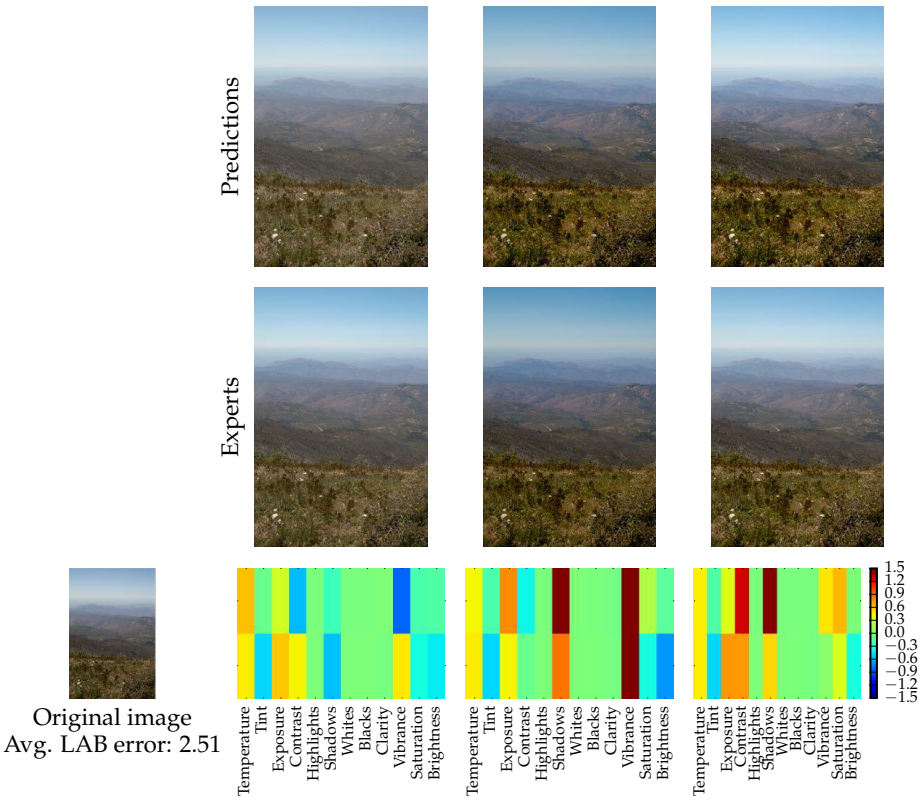

Figure 3: Image 4889 from Adobe-MIT5k dataset

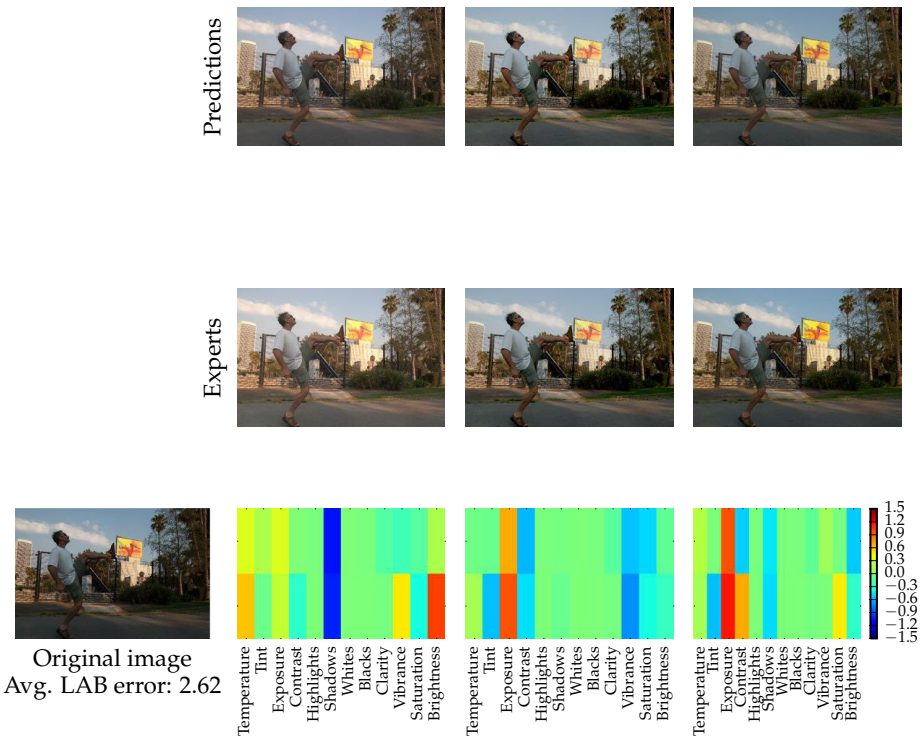

Figure 4: Image 4910 from Adobe-MIT5k dataset

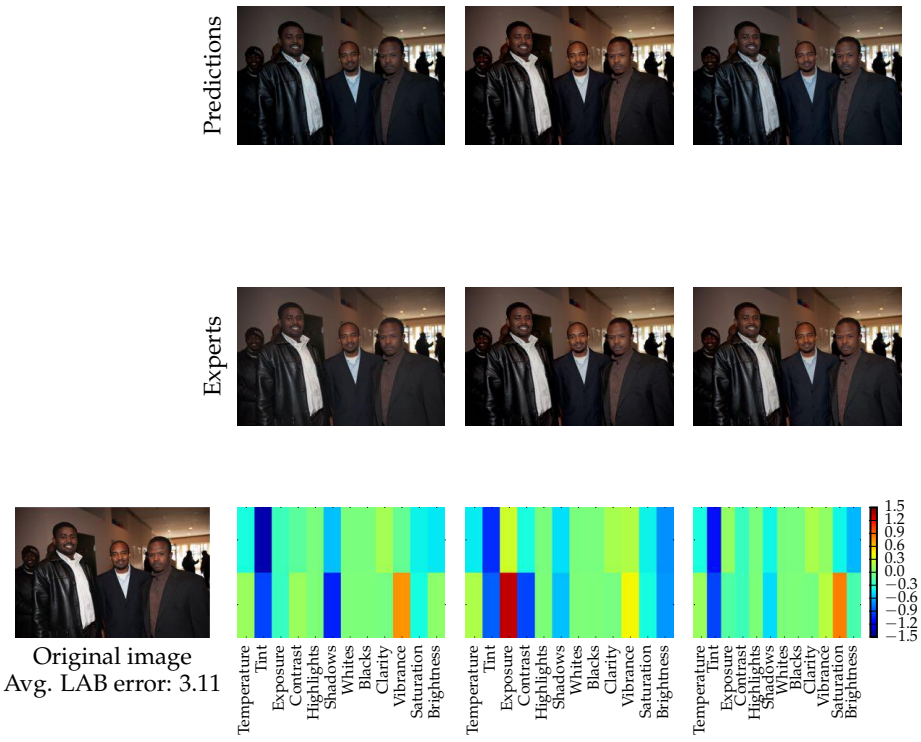

Figure 5: Image 4902 from Adobe-MIT5k dataset

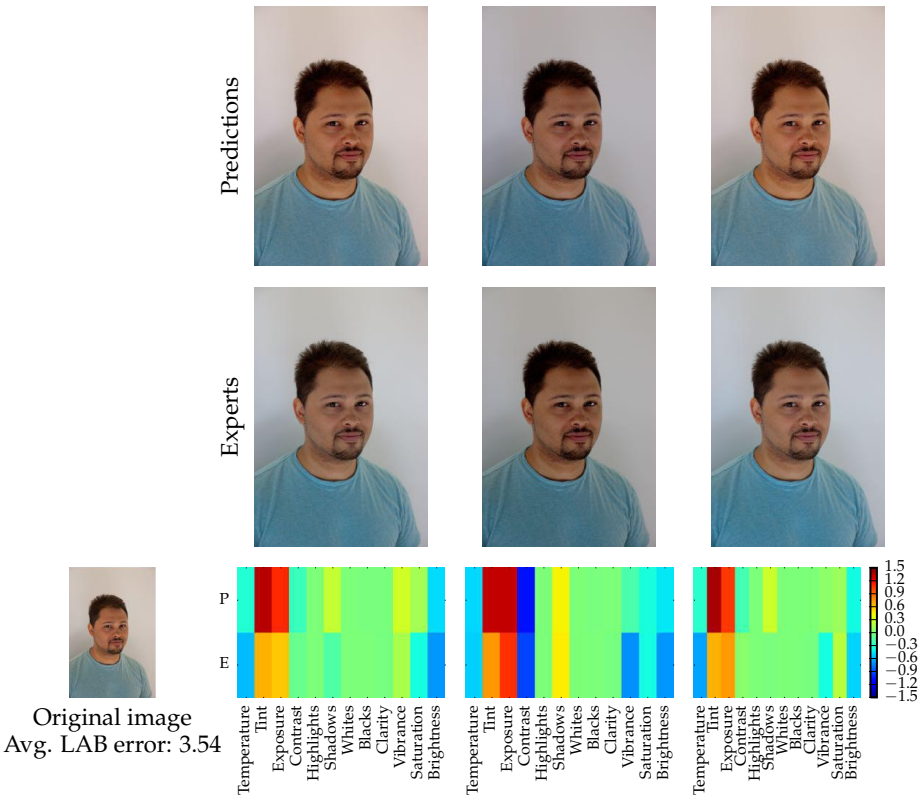

Figure 6: Image 4873 from Adobe-MIT5k dataset

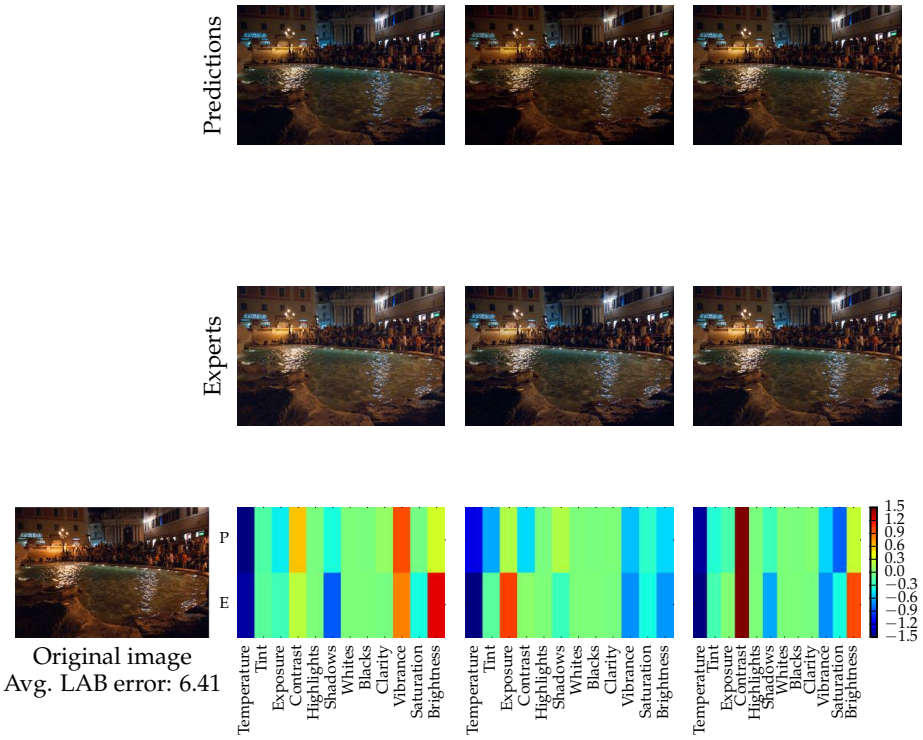

Figure 7: Image 4882 from Adobe-MIT5k dataset

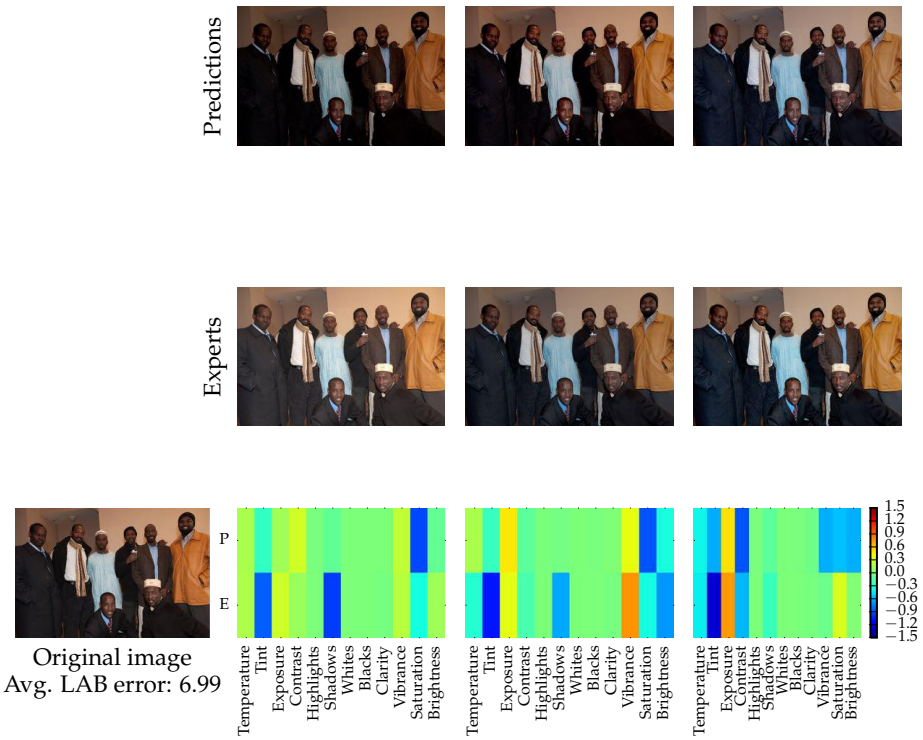

Figure 8: Image 4872 from Adobe-MIT5k dataset

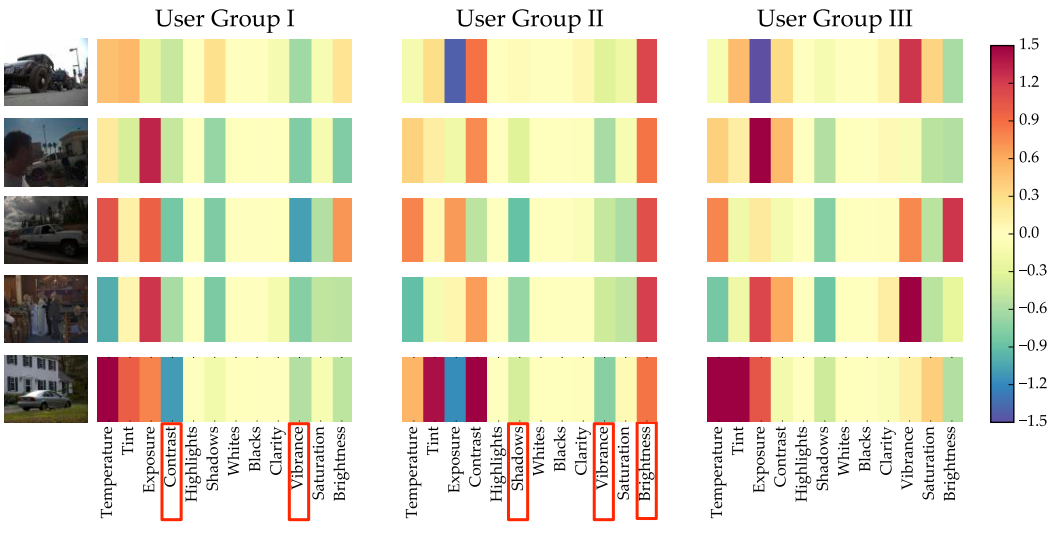

Figure 9: A sample user categorization from the P-VAE model (car image).

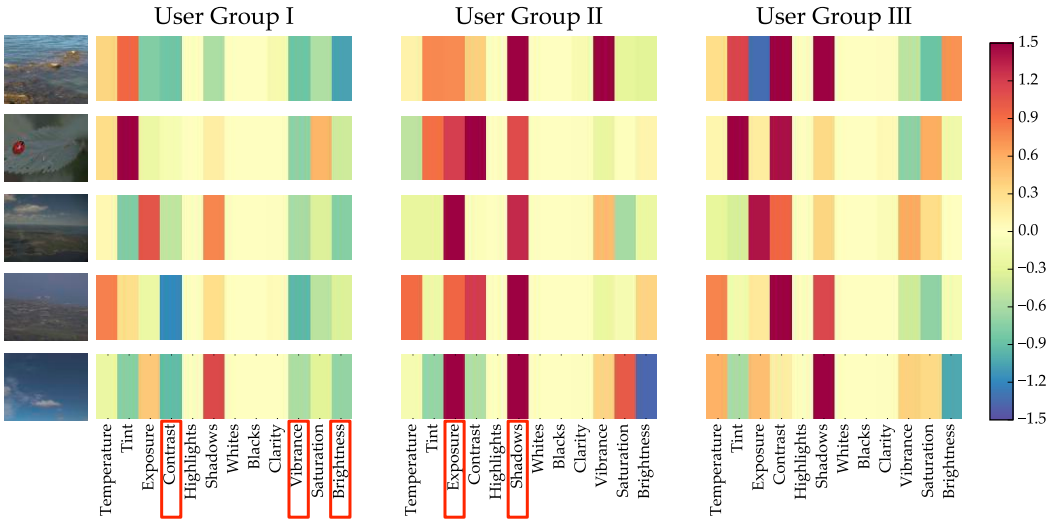

Figure 10: A sample user categorization from the P-VAE model (similar photos with dominant blue colors).

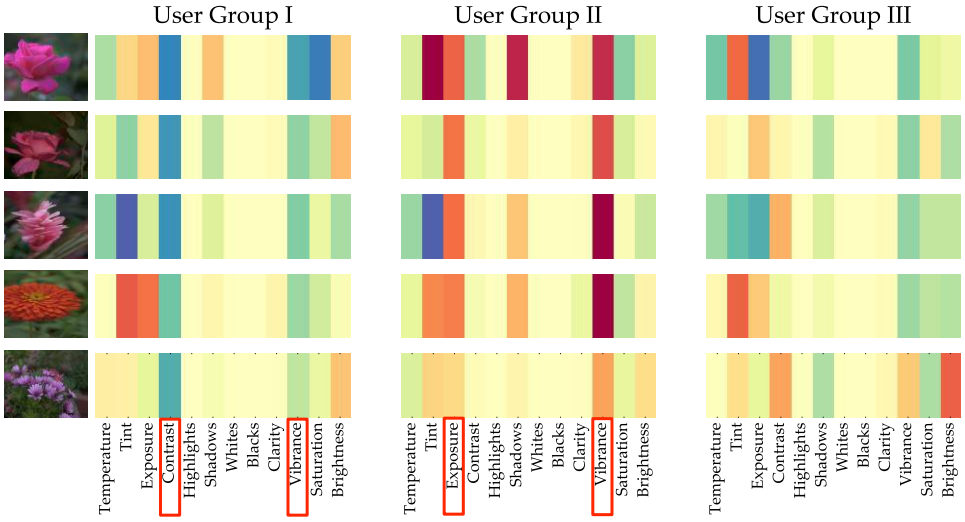

Figure 11: A sample user categorization from the P-VAE model (flowers).

References

Y. N. Dauphin and D. Grangier. Predicting distributions with linearizing belief networks. *arXiv preprint arXiv:1511.05622*, 2015.

Y. Koyama, D. Sakamoto, and T. Igarashi. Selph: Progressive learning and support of manual photo color enhancement. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 2520–2532. ACM, 2016.

J.-Y. Lee, K. Sunkavalli, Z. Lin, X. Shen, and I. S. Kweon. Automatic content-aware color and tone stylization. *arXiv preprint arXiv:1511.03748*, 2015.