جامعة نيويورك أبوظبي
NYU ABU DHABI

Image Classification

Dataset

- Two datasets were selected: a fashion photographer dataset from Instagram and a wedding couple photos dataset from Google Images.
- The fashion photographer dataset included images from Brandon Woelfel, Theo Wenner, Petra Collins and Faye, known for their focus on manipulating light.
- The wedding couple photos dataset consisted of 10 different colors (pink, white, blue, green, black, orange, red, yellow, light green and purple) outfits.

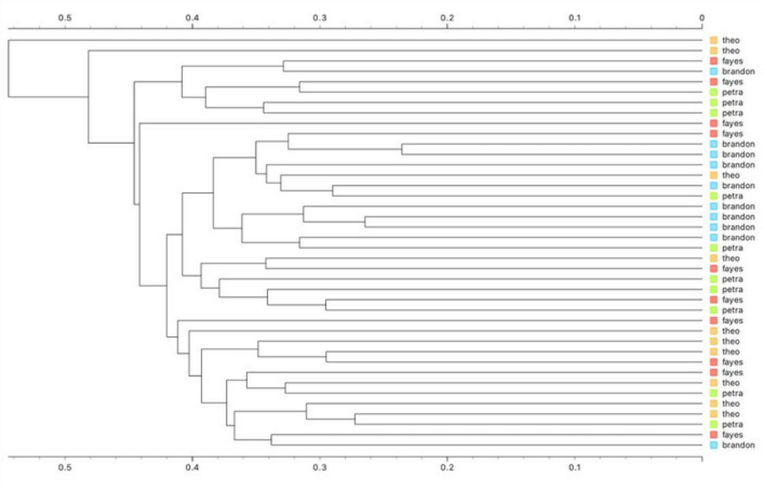
Analysis

- The clustering method demonstrated in the tutorial did not effectively cluster images from the fashion photographer dataset.
- Similar compositions caused images from different photographers to be clustered together, despite differences in lighting.
- The algorithm appeared to identify features of the subjects, such as hair and eye color, in addition to photographic style.
- Different image embedding options were explored, with Inception and VGG-16 settings yielding the best results.

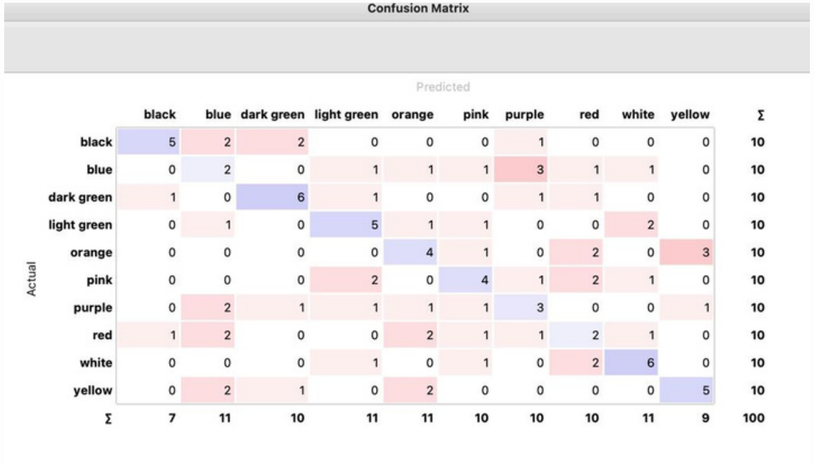
Conclusion

- The clustering process revealed challenges in accurately grouping images based on composition and lighting. This highlights the complexity of image analysis and the need for further refinement in algorithmic techniques.
- Experimenting with different image embedding options, specifically Inception and VGG-16 settings, proved instrumental in achieving better results for our datasets. The choice of embedding settings should be tailored to the specific characteristics of the data being analyzed.
- Ethical considerations surrounding data usage and sharing, particularly with photos, require attention to avoid biases and potential harm.
- The concept of "distant viewing" aligns with the study's use of Orange's hierarchical clustering algorithm.
- This approach enables the extraction of metadata and identification of patterns and relationships within the data.

Visualisation of Datasets



Hierarchical Clustering of Photographers Dataset



Confusion Matrix of Wedding Couple Photos Dataset