Detecting Emotion from People and Scenery

Logan Preston, Neal Desai, Nikhil Nagam

Introduction

- Goal: Help computers identify emotions from images
  - Detecting emotion can help guide computers to socially acceptable expectations
  - For example, being able to detect frustration from a user's face could inform an a system to suggest help more or less often
  - In general, we expect being able to identify emotions accurately will enhance interactions between AI-infused systems and their users
- This is important because AI-infused applications are increasing in use, and Human-AI interactions often mention AI should have "socially appropriate behaviors" [1,2]

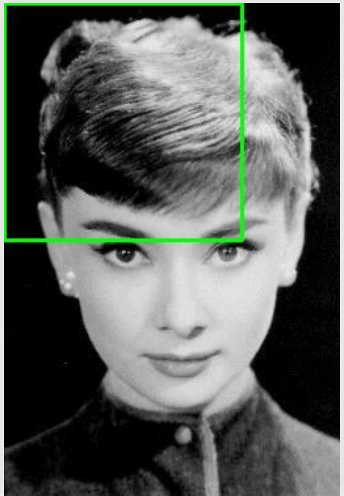
- Our project extends recent research on detecting emotions from scene and body to include face information. With this in mind, we also wanted to tell how much each factor (face, body, scene) contributes to the overall estimated emotion

Related Work

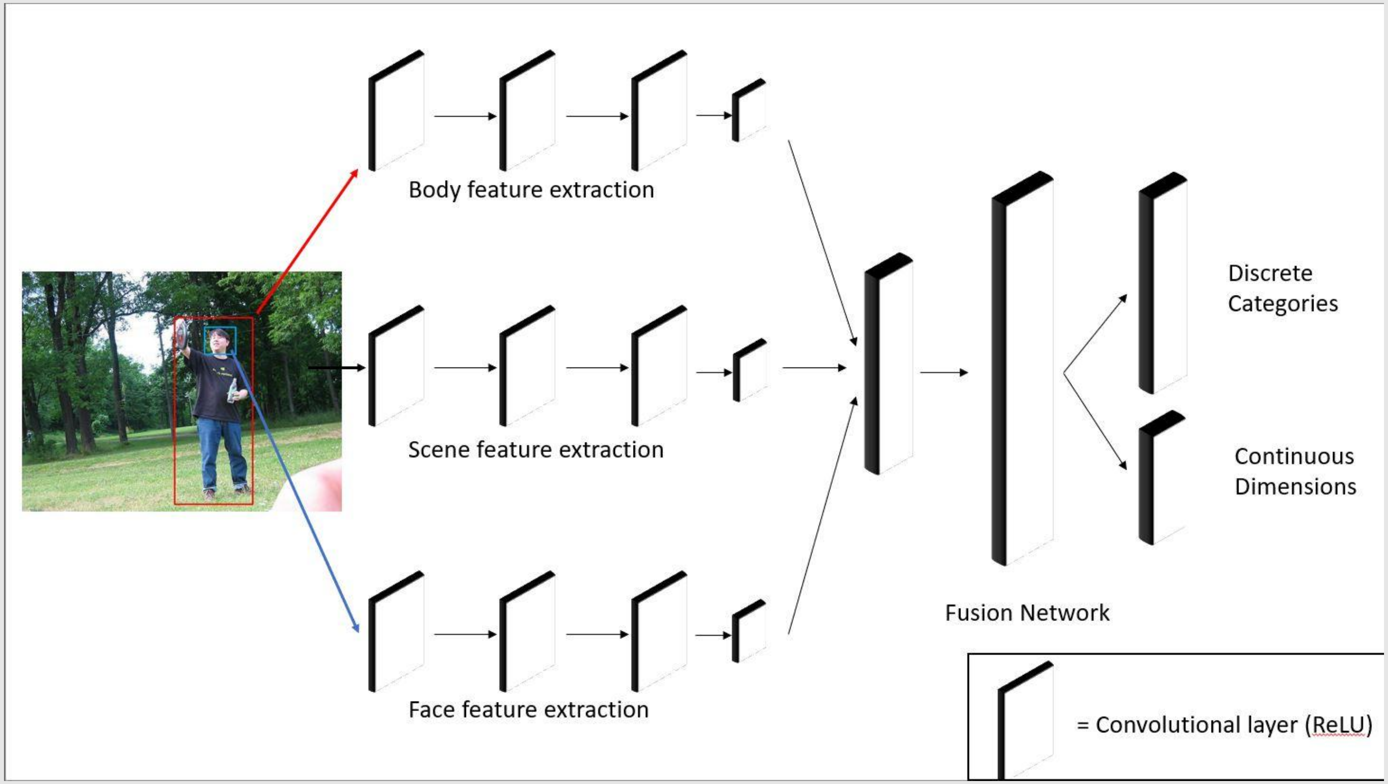
- Classic emotion recognition is usually done with face detection and drawing emotion from the face. Deep Learning models estimate emotions quite well now [3]
  - However, there is rich information in the pose of the subject and the scene around them that could change the result
- Recent work has also investigated how to learn emotion from the scene and body poses [4,5]
  - This work does not consider face information, the model was trained using the entire person’s body.
  - This work demonstrated that learning from the scene and the body together provided better results than either of them individually

Our Approach

- Our implementation builds upon prior work to estimate emotion based on the scene, body pose, and face individually, then combine them with a fusion network at the end
- Our data set provided scenes with people, and had been labeled with emotions.
  - Bounding boxes provide a reference to the person being evaluated
- We use Haar features to detect faces in the images for our face model
  - One difficulty we had with this was detecting faces appropriately, to improve our face recognition we only accepted “faces” that were in the top half of a bounding box around people. This was a good heuristic to ensure we have actual faces detected



Model Architecture



Results

- From the results below, you can see that the face, body, and scene information has the highest mean average error and the second highest precision. The scene and body gets very similar results, while the other two pairings get similar but slightly worse results. Each of the components alone have the lowest precision

Components	Validation Mean Avg Error	Testing Mean Avg Error	Validation Mean Avg Precision	Testing Mean Avg Precision
Face, Body, Scene	0.9067	1.0597	0.3183	0.2111
Scene and Body (baseline)	0.8924	1.0381	0.3133	0.2102
Scene and Face	0.8878	1.0332	0.3057	0.2156
Body and Face	0.9093	1.0158	0.3002	0.2004
Body	0.8570	0.9976	0.2802	0.1962
Face	0.8587	1.0183	0.2698	0.1862
Scene	0.8547	1.0052	0.2572	0.1878

Discussion

- We learned how to more reliably detect faces using Haar features, and using heuristics to increase detection rate
- We also learned where the current state of emotion detection is, common data sets and methods for robust detection, and what contributes to emotions
- It may be apparent already, but having more specialized systems (e.g. one explicitly for faces, and another for body) tends to lead to better results
    - We are curious how isolating other expressive parts of the body, such as hands, could improve detection accuracy
- This could lead to more robust emotion detection, especially in scenes where context matters in addition to the face

Questions?

References

1. S. Amershi, D. Weld, M. Vorvoreanu, A. Fourney,B. Nushi, P. Collisson, J. Suh, S. Iqbal, P. Bennett,K. Inkpen, J. Teevan, R. Kikin-Gil, and E. Horvitz. Guidelines for human-ai interaction. In CHI 2019.ACM, May 2019. CHI 2019 Honorable Mention Award.
2. E. Horvitz. Principles of mixed-initiative user interfaces. In Proceedings of CHI '99, ACM SIG CHI Conference on Human Factors in Computing Systems, Pittsburgh, PA, ACM Press., pages 159-166,May 1999.
3. B. C. Ko. A brief review of facial emotion recognition based on visual information. volume 18, 2018.
4. R. Kosti, J. Alvarez, A. Recasens, and A. Lapedriza.Context based emotion recognition using emotic dataset. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019.
5. R. Kosti, J. M. Alvarez, A. Recasens, and A. Lapedriza. Emotion recognition in context. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
6. J. Lee, S. Kim, S. Kim, J. Park, and K. Sohn.Context-aware emotion recognition networks. In Proceedings of the IEEE/CVF International Confer-ence on Computer Vision (ICCV), October 2019.