Visualizing and Quantifying Discriminative Features for Face Recognition - Supplemental Material

Gregory Castanon and Jeffrey Byrne
Systems & Technology Research, Woburn MA

I. Introduction

In this supplemental material, we provide additional visualizations of contrastive network attention.

In section II-A, we provide examples of contrastive excitation backprop under pose, illumination, expression conditions. This provides additional visualizations across multiple examples of a single subject to characterize stability of this attention map for a single subject. We also provide cEBP examples for a random set of face images.

In section II-B, we explore the question, "Do two subjects share contrastive features in common"? We provide a tool based on "split screens" to visualize the contrastive signal for a specific subject on a composite image of two subjects, which allows for visualization of contrastive features common to two subjects.

In section II-C, we provide examples of a common phenomenon for contrastive EBP which shows shirt collars are often visualized as contrastive. This suggests that collared shirts on men are a training set bias for these subject classifiers, and provides a test to confirm whether this bias has been corrected following training set refinement.

In section II-D, we provide additional visualizations of contrastive EBP under occlusion conditions. This shows that when an image is partially occluded, that regions such as the hair or beard are used for contrastive network attention.

II. Experiments

A. Contrastive Excitation Backprop Examples

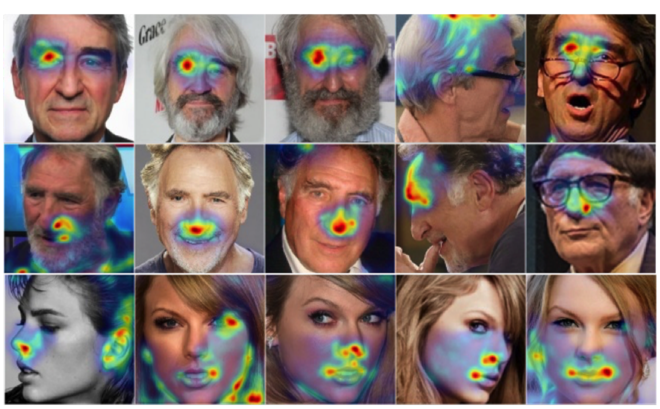

Fig. 1. Contrastive EBP subject gallery. The contrastive EBP signal for three subjects in a range of pose, illumination and expression conditions.

Figure 1 shows additional examples of cEBP for three subjects under varying pose, illumination and expression conditions. The contrastive attention map produced by cEBP tends to highlight Judd Hirsch's large nose and Sam Waterston's bushy eyebrows, but the signal is not perfectly stable. When the face is seen in profile and in some cases when the face is perfectly aligned, the signal move around across different images of a subject. While this signal is not as effective as EBP at identifying the regions that are most important to the network (as measured by the hiding game), this can offer insight as to what is unique about a given subject.

Figure 5 shows a mosaic of additional examples of cEBP for a random selection of subjects. Observe that this result shows the instability of the cEBP signal as discussed in the main body, which resulted in the proposed truncated cEBP.

B. Split Screens

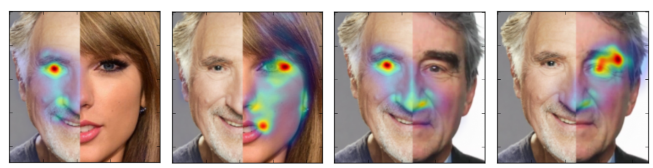

Fig. 2. Split screen experiment. By back-projecting different cEBP classes onto the same image, we confirm that the contrastive signal is not universal, which enables visualization of contrastive features in common between two subjects.

Do two subjects share contrastive features in common? Recall that the contrastive network attention map localizes regions that are slightly more excitory for a subject, and slightly less excitory for the average subject in the training set. However, given two subjects, do they share similar contrastive features? In other words, can we show that Sam Waterston's nose is notably different than Judd Hirsch's or Taylor Swift's? To explore this question, we form a set of hybrid images by roughly-aligning near frontal images of faces from two different subjects, forming "split screen" images, and computing the contrastive network attention with respect to a particular subject.

Consider the four hybrid images shown in Figure 2. In the first two images, we alternately create a cEBP attention map for the Judd Hirsch class and the Taylor Swift class onto the hybrid image. Note that we expect that these two classes have very little in common. While we would not expect center-oriented features like nose and mouth to trigger, because of the discontinuous nature of the image, we still see peaks at the corners of Judd Hirsch's nose and Taylor Swift's lips. In the second two images, we show an example of two faces that are somewhat similar in age and gender. In this example, we see common contrastive features between Judd Hirsch and Sam Waterston. While Judd Hirsch's eyebrows are not as bushy as Sam Waterston's, their noses are similar. This provides another tool for an analyst to explore the common contrastive features for subjects.

C. Training Set Bias

Fig. 3. Contrastive EBP examples exhibiting dataset bias. Due to dataset bias, shirt collars are a discriminating feature for male celebrities.

Despite its noise, the contrastive network attention map suggests that it can be used to identify dataset biases. Figure 3 shows the contrastive class signal as applied to a number of male celebrities, and shows that the contrastive signal significantly highlights shirt collars in a number of them. The training set for this network comprised images entirely of celebrities, and men are often photographed in suits with white collars at awards shows and premiers, which suggests that for some celebrities on this dataset, collars are a discriminating feature. In general, collars are not discriminative as shirt type is independent of identity. However, in a dataset of celebrities, this bias may exist.

This also suggests a test for confirming that dataset bias has been removed. The cEBP signal can be used to identify those subjects with shirt collars as contrastive features. These subjects can be augmented in the training set to include additional examples of this subject without a shirt collar. Then, the network can be retrained using this augmented dataset, and the cEBP signal can be regenerated to confirm that the short collars are no longer contrastive. This could be used as a general techniquie to identify training set biases, however we leave such an investigation to future work.

D. Occlusions

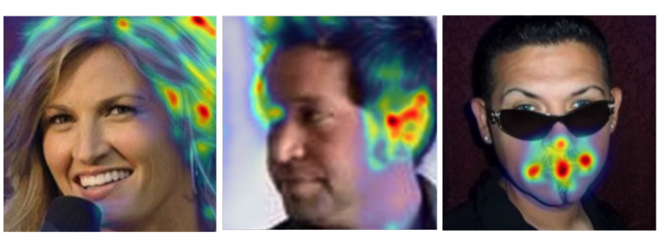

Fig. 4. Contrastive EBP examples exhibiting partial occlusion.

Figure 4 shows examples of the cEBP signal under occlusion. We can see these signals change when parts of the image are naturally obscured. In Figure 4(a), contrastive neural attention highlights features that might be expected to be prominent - Sam Waterston's Eyebrows. However, when a subject's features are obscured by sunglasses, as in Figure 4 (b) and (c), it can highlight other prominent features which might draw neural attention more than average, such as a unique mustache or a prominent lower lip.

E. Truncated cEBP

In the main submission, we described an extension of the contrastive excitation backprop to address an observed instability. In this section, we show the same montage of 84 subjects that we have used as a running example for EBP and cEBP to show the benefit of truncated cEBP.

Figure 6 shows this montage. When compared to the cEBP montage for the same images in figure 5, we see that the majority of the contrastive attention maps are nearly identical. However, there are a few notable examples with large differences. For example, Row 3, Column 7 shows Janeane Garafalo where the cEBP signal shows attention in the image corners, whereas the truncated cEBP shows attention within the facial region. Similarly, Row 4, column 5 shows contrastive EBP on the clothes, whereas truncated cEBP is on the lips. Row 1, column 6 shows contrastive EBP in the bottom corner whereas truncated cEBP is primarily on the nose. These are the examples which highlight the robustness of the truncated contrastive EBP contribution, without affecting the cases where the cEBP is already stable.