A Suspect Identification Framework using Contrastive Relevance Feedback (Supplementary Material)

1. Additional Studies

We mention some additional studies that we performed which provide the process which was involved behind obtaining the $SCLoss$ and the formulation of the algorithm. Moreover, these studies also give an insight on certain alternatives that could have been used for this algorithm but contained caveats which were revealed during experiments and the user study.

1.1. Pretraining

The parameters of the projection network can also be initialized randomly at the start of the search or obtained through pretraining the projection network. The projection network can be pretrained in the same way as the training is done to learn representations in [1] with the parameters of the encoder being frozen and the projection network being highly regularized. Pretraining the projection network did not show much improvement in the performance on both the simulation and the user study as compared to a projection network which was randomly initialized at the start of the search. We show experiments comparing both the pretrained ($\texttt{FaIRCoP}-\texttt{P}$) and the non-pretrained ($\texttt{FaIRCoP}$) variants of our algorithm in the supplementary. We present the user study results and simulation results comparing the pretrained and non-pretrained versions of FaIRCoP as given in Table 2 and Table 1. We can see that there are no significant differences in both the methods.

1.2. Alternate Loss Function

An alternative of the $SCLoss$ and the scoring function introduced in the main paper can be written as Equation 1 and Equation 2 respectively.

$$\begin{aligned} L_s(S,D) &= \sum_{x\in S}\sum_{y\in S-\{x\}} l_D(x,y) \\ L_d(S,D) &= \sum_{x\in D}\sum_{y\in D-\{x\}} l_S(x,y) \\ \texttt{SCLoss}_{alt}(S,D) &= \frac{L_s(S,D)}{2|S|(|S|-1)} + \frac{L_d(S,D)}{2|D|(|D|-1)} \end{aligned} \tag{1}$$

In this function, the definition of l remains the same as defined originally in the main paper. This function establishes an explicit dichotomy between the similar images and dissimilar images and forces the projection network to cluster them separately as compared to the original loss function which relaxed the clustering constraint on the dissimilar set of images. Consequently, this change would accordingly modify the scoring function which would also be characterized by its distance from the cluster containing the projected embeddings of dissimilar images selected by the user as given in Equation 2.

$$score_{alt}(u) = sim(u, \frac{1}{|S_a|}\sum_{x\in S_a} x) - sim(u, \frac{1}{|D_a|}\sum_{x\in D_a} x) \tag{2}$$

This loss function gave very similar results on the simulations but proved to be inefficient when it was employed for the user study. We speculated that this problem arises due to the practical problem of encountering dissimilar images of a higher number in the initial iterations as compared to similar images. Thus, constraining the dissimilar images to a single cluster leads to learning degenerate solutions where the latent space is mostly covered by the dissimilar images, hence, not leading to the desired separation between both the clusters. Moreover, our problem hypothesizes that all the similar images have certain common features which cause the users to select an image as similar to their mental image, but with clustering all the dissimilar images, we essentially force the projection network to project large feature differences nearer to each other, which may cause the projection network to learn less meaningful projections, thus, leading to delayed convergence in real time. With relaxing the dissimilarity constraint, we obtained better results with the user study.

2. Distribution Similarity Interpretation

Figure 1 highlight the data distribution of different sensitive features for the full dataset as well as the dissimilar images obtained from the user simulator logs. As depicted in these plots, both the distribution are similar which supports the claim that the framework employed works in fair manner and is not biased towards specific classes within each tangible factor of variation.

Representation			ACI		AR	
Criminal Dataset						
FaceNet	MIX	HOG	FaIRCoP	FaIRCoP − P	FaIRCoP	FaIRCoP − P
✓	✓	✓	**57.25**	87.20	0.82	**0.83**
✓	✓		**68.33**	73.10	**0.83**	0.82
	✓	✓	**41.66**	74.90	0.79	**0.88**
✓		✓	**98.33**	100.75	**0.79**	0.64
	✓		89.00	**74.40**	**0.88**	**0.88**
CelebA Dataset						
FaceNet	MIX	HOG	FaIRCoP	FaIRCoP − P	FaIRCoP	FaIRCoP − P
✓	✓	✓	**40.5**	50.4	0.61	**0.68**
✓	✓		**27.4**	76.6	**0.70**	0.61
	✓	✓	**50.0**	58.0	**0.87**	0.83
✓		✓	98.2	**96.6**	0.54	**0.55**
	✓		**20.2**	44.0	0.82	**0.87**

Table 1: Quantitative metrics obtained from user simulation using different methods on the Criminal and CelebA dataset.

Algorithm	PREF	REL	RESP	CONV
Criminal Dataset				
FaIRCoP	0.70	0.72	0.81	0.44
FaIRCoP − P	0.68	0.71	0.80	0.42
CelebA Dataset				
FaIRCoP	0.63	0.71	0.67	0.36
FaIRCoP − P	0.62	0.71	0.68	0.35

Table 2: Cumulative metrics obtained from the User Study conducted on Criminal and CelebA dataset.

3. Algorithms Discussed in Paper

We present the algorithms for suspect retrieval using relevance feedback and user simulator in algorithms 1 and 2 respectively below that we discussed in the main paper.

4. Web Interface

FaIRCoP is built on highly scalable and dynamic open-source frameworks: Next.js, Redux, Geist UI (frontend) and Django (backend). Apart from providing an overall theme to the system, advanced techniques such as Dynamic Import, Static HTML Export and Internationalisation have been used to ensure fast loading times and universal deployability. The entire data from visual elements to the results take shape as per variables configured on the backend, which ensures true CMS like behaviour, thus, full control over the frontend. To address security concerns, the app is hosted with it's static HTML export which leaves only a single entry point into the system. After a successful login, a user session is authorized every minute to never leave a user with stale data.

A video demonstration of the working system can be viewed at https://ijcai-faircop-default.layer0-limelight.link/0

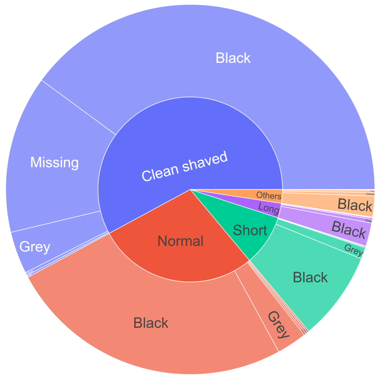
5. Future Work

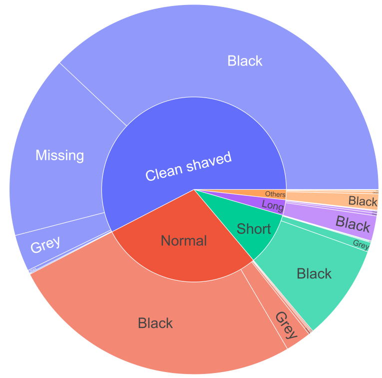
The system currently starts by showing the user a completely random set of images based on certain attributes explicitly selected by the user. We intend on making this initialization more robust by incorporating components of natural language to improve the system metrics. Furthermore, a metric can be added to this system which determines when to explore the database or when to exploit the previous data, and suggest those images in addition to the base set of recommended images. We also plan to improve the user simulator by using the concept of Eigen faces [2] and designing a way to determine visual similarity using its geometric properties.

References

[1] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning*, Proceedings of Machine Learning Research, pages 1597–1607. PMLR, 13–18 Jul 2020.

[2] M.A. Turk and A.P. Pentland. Face recognition using eigenfaces. In *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 586–591, 1991.

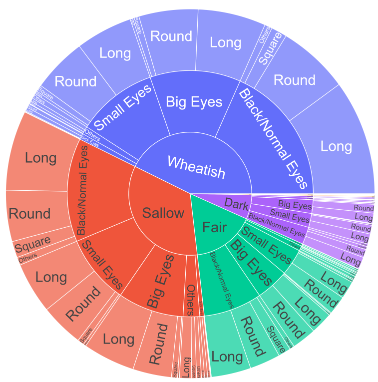


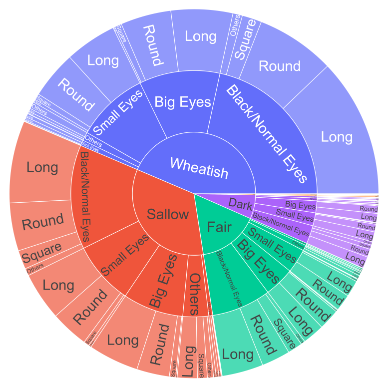


Full Dataset

Selected Dissimilar Images

(a) Beard

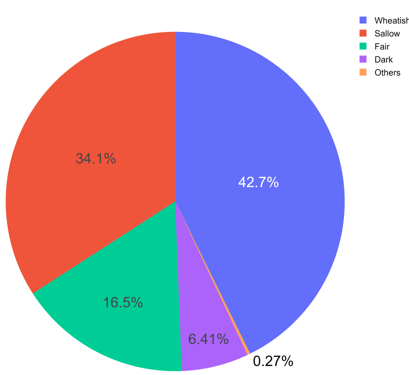


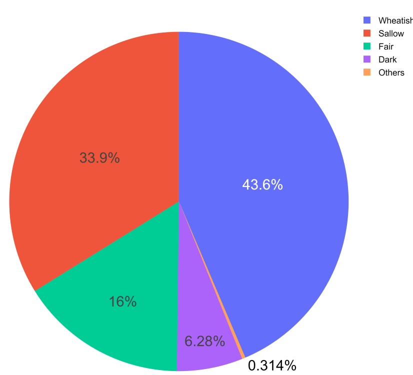


Full Dataset

Selected Dissimilar Images

(b) Facial Characteristics





Full Dataset

Selected Dissimilar Images

(c) Complexion

Figure 1: Data distribution between full dataset and dissimilar images generated from the simulation logs with respective to different features

Algorithm 1 Relevance Feedback Algorithm

$embed \leftarrow$ Encoder for extracting base embeddings for Images
Require: `Dossier`, k, u, $prev\_samp$, $epochs$, $embed$
  $S_{all}, D_{all} \leftarrow \{\}$
  $R \leftarrow$Sample k random images uniformly from sensitive attributes such as gender and complexion from `Dossier`
  $Rem \leftarrow$ `Dossier`
  $f \leftarrow$ Initialize(f)
  $iter \leftarrow 0$
  while Mental Image $\notin R$ **do**
    $S \leftarrow$ User Selects images from R
    $D \leftarrow R - S$
    $Rem \leftarrow Rem - R$
    if $iter\%2 == 0$ **then**
      $S_B \leftarrow sample(S_{all}, min(prev\_samp, |S_{all}|)) \cup S$
      $D_B \leftarrow sample(D_{all}, min(prev\_samp, |D_{all}|)) \cup D$
      $i \leftarrow 0$
      for $i < epoch$ **do**
        $S_{proj} \leftarrow f(embed(S_B))$
        $D_{proj} \leftarrow f(embed(D_B))$
        $L \leftarrow$`SCLoss`(S_{proj}, D_{proj})
        Backpropagate and Update the projection network f to minimize L
        $i \leftarrow i + 1$
      end for
    end if
    $S_{all} \leftarrow S_{all} \cup S$
    $D_{all} \leftarrow D_{all} \cup D$
    for $v \in Rem$ **do**
      $sc_v \leftarrow score(v, f(embed(S_{all})), f(embed(D_{all})))$
    end for
    $R \leftarrow$ Images corresponding to the top-k scores
    if $iter\%3 == 0$ **then**
      $R \leftarrow R \cup sample(S_{all} \cup D_{all}, u)$
    else
      $R \leftarrow R \cup sample(Rem, u)$
    end if
    $iter \leftarrow iter + 1$
  end while

Algorithm 2 User Simulator

Require: `Dossier`, `HOG`, `FaceNet`, `MIX`, `RF_Algo`, `RF_Algo_Init`, $k \leq |$`Dossier`$|$
  $Target \leftarrow$ A random image sampled from `Dossier`
  $H_T, F_T, M_T \leftarrow$ `HOG`, `FaceNet`, `MIX` (Target)
  $SampledImgs \leftarrow$ Sample k images from `Dossier`
  $thr \leftarrow 0$
  for $s \in SampledImgs$ **do**
    $H_s, F_s, M_s \leftarrow$ `HOG`(s), `FaceNet`(s), `MIX`(s)
    $thr \leftarrow thr + \frac{sim(H_T,H_s)+sim(F_T,F_s)+sim(M_T,M_s)}{3}$
  end for
  $thr \leftarrow \frac{thr}{|SampledImgs|}$
  $S \leftarrow$ `RF_Algo_Init`()
  $iter, STemp \leftarrow 0, \{\}$
  while Target $\notin S$ **do**
    $Sim, DisSim \leftarrow \{\}$
    for $s \in S$ **do**
      $H_s, F_s, M_s \leftarrow$ `HOG`(s), `FaceNet`(s), `MIX`(s)
      $SimVal \leftarrow \frac{sim(H_T,H_s)+sim(F_T,F_s)+sim(M_T,M_s)}{3}$
      if $SimVal > thr$ **then**
        $Sim \leftarrow Sim \cup \{s\}$
      else
        $DisSim \leftarrow DisSim \cup \{s\}$
      end if
    end for
    $STemp \leftarrow STemp \cup Sim$
    $S \leftarrow$ `RF_Algo`(Sim, DisSim)
    $iter \leftarrow iter + 1$
    if $iter\%15 == 0$ **then**
      $u \leftarrow 0$
      for $s \in STemp$ **do**
        $H_s, F_s, M_s \leftarrow$ `HOG`(s), `FaceNet`(s), `MIX`(s)
        $u \leftarrow u + \frac{sim(H_T,H_s)+sim(F_T,F_s)+sim(M_T,M_s)}{3}$
      end for
      $u \leftarrow \frac{u}{|STemp|}$
      $thr \leftarrow 0.95thr + 0.05u$
      $STemp \leftarrow \{\}$
    end if
  end while