Understanding Neural Responses to Face Verification of Cross-Domain Representations

Maneet Singh[1], Shruti Nagpal[1], Daksha Yadav[2], Naman Kohli[2], Prateekshit Pandey[3], Gokulraj Prabhakaran[4]
[1]*IIIT-Delhi, India;* [2]*West Virginia University, USA;* [3]*University of Pennsylvania, USA;* [4]*Otto-von-Guericke University, Germany*
Email: {maneets, shrutin}@iiitd.ac.in, {dayadav, nakohli}@mix.wvu.edu, prateekshit.pandey@asc.upenn.edu
Richa Singh[5], Mayank Vatsa[5], Afzel Noore[6], Julie Brefczynski-Lewis[2], Harsh Mahajan[7]
[5]*IIT Jodhpur, India;* [6]*Texas A&M, Kingsville, USA;* [7]*Mahajan Labs, India*
Email: {richa, mvatsa}@iitj.ac.in, Afzel.Noore@tamuk.edu, jblewis@hsc.wvu.edu

***Abstract*—Face verification involves identifying whether two faces belong to the same person or not. It relies heavily upon face perception, processing, and the decision making of an individual. This research studies cross-domain face verification, where one face image belongs to a controlled, well-illuminated environment, while the other is of a varying representation having differences in image type or quality. Specifically, two cross-domain face verification tasks are analyzed: controlled-low resolution and controlled-sketch face verification. functional Magnetic Resonance Imaging (fMRI) data has been collected for 23 participants of two ethnic groups while performing face verification. Statistical comparisons were performed with same-domain controlled face verification for both the tasks. Our findings reveal regions of Right Frontal Gyrus, Bilateral Insula, and Right Middle Cingulate Cortex demonstrating higher activation for controlled-sketch face verification, as compared to controlled face verification. Similar analysis were performed for controlled-low resolution face verification, where regions responsible for higher visual load and difficult tasks result in higher activation. Further, stimuli ethnicity differences influence activations for low-resolution face verification but do not affect sketch face verification. Regions of Right Middle Occipital Gyrus and Right Fusiform Gyrus present higher activity, suggesting increased face processing effort for within ethnicity low resolution face verification. We believe the findings of this research will help enable further development in the field of brain-inspired facial recognition algorithms.**

***Index Terms*—Face Recognition, fMRI, Neural Responses**

I. Introduction

It has long been established that humans possess an exceptional ability to perform face processing [1]–[3]. Coupled with decision making, we are also able to carry out face verification with utmost ease and precision. Face verification refers to the task of matching two face images and deciding whether they belong to the same person or not. As shown in Fig. 1(a), the past couple of decades have seen studies aimed at understanding the functioning of the human brain for processing and recognition of objects and faces [4]–[9]. Existing literature has identified regions of Fusiform Face Area (FFA), Occipital Face Area (OFA), and face-selective region in the Superior Temporal Sulcus (fSTS) to have higher activations for face perception and recognition [10]–[15]. Research has also focused on understanding the human brain behavior while performing face recognition under different conditions/covariates such as familiar faces (or kinship) [16], [17], inverted faces [18], [19], emotion [20], [21], self [22], and other-race effect [23], [24].

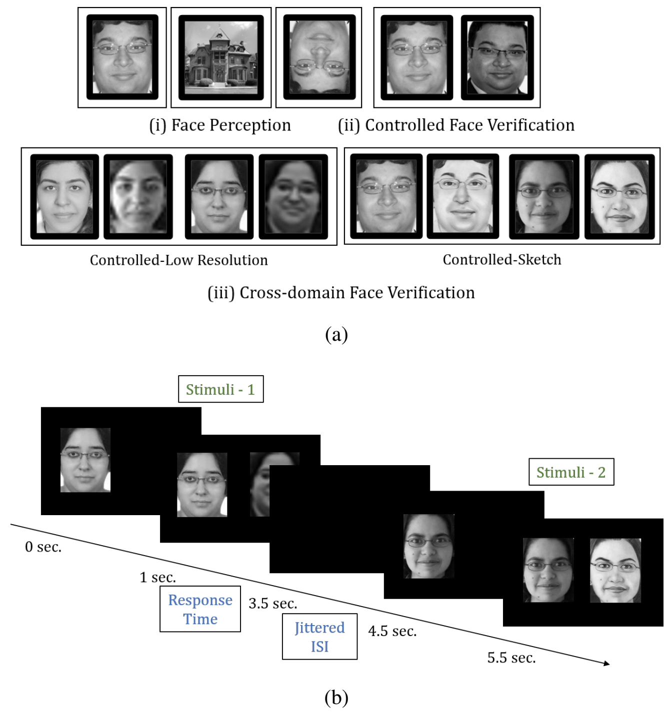


Fig. 1: (a) Research has focused on understanding how humans perform (i) face perception, and (ii) controlled face verification. This research analyzes how humans perform (iii) cross-domain face verification in comparison with (ii) controlled face verification. (b) Timeline of two sample stimuli presented to the participants. The face images in the figure are representative of the stimuli used in the experiment. The images belong to the co-authors of this paper who have given informed consent to use these images for publication purposes.

While research has extensively focused on understanding the neural correlates behind face perception or recognition, face verification has primarily been explored via behavioral studies [25]–[27]. For the task of face verification, the existing neuroimaging studies focus on *controlled* and *same-domain* face images only [28]–[30]. Here, *controlled* refers to regular high quality face images matched in low-level stimulus features such as luminance and contrast, captured in good illumination, often in frontal pose. *Same-domain* refers to the scenario where both the images are captured in similar settings, resulting in similar face representations. In real world scenarios, two face images are rarely captured in a controlled environment, thereby resulting in a pair of face images having different representations from two domains. Such situations result in a need for *cross-domain face verification*, where one domain contains face images captured in controlled environment, while the other domain contains face images of a varying representation. For example, law enforcement personnel often require matching low resolution face images (captured in surveillance scenarios) with high quality mugshot images. In another scenario, the sketch face image of a wanted person (generated based on an eyewitness description) is often required to be matched against the high quality mugshot images.

Both the situations require humans to perform cross-domain face verification in order to identify possible suspects. Such scenarios are further characterized by the same-ethnicity or different-ethnicity pairing of the investigating officer and the person under investigation. In the literature, researchers have performed behavioral studies on cross-domain face verification [31]–[33] or behavioral studies on varying representations [34], however, to the best of our knowledge, no existing research attempts to understand the functioning of the human brain for the said task. Research has also focused on the own-race effect for face recognition [23], however, no studies exist for understanding the human behavior for cross-domain face verification with respect to the ethnicity pairings. This leads to the main contribution of this study: understanding the functioning of the human brain while performing cross-domain face verification. Matching controlled face images with low resolution face images and face sketches can be more challenging than matching with other controlled face images. While low resolution images lack high frequency components and facial contours, sketch face images require thinking about abstract facial features to match them with real face images. Thus, while low resolution face verification and sketch verification are both challenging, they are likely to elicit different cognitive processing on top of face image processing. Specifically, we expect respondents to engage in more high level visual processing of the images for low resolution face verification, and in more high level cognition for sketch face verification. In this exploratory study, we hypothesize that the differences in the underlying psychological processes would reflect in the brain activity in response to different face verification tasks. Specifically, we test our hypothesis that the human brain performs cross-domain face verification differently from same-domain controlled face verification. Further, we hypothesize that the neural activity for face verification varies with varying face representations, and are affected by the ethnicity pairing of the self and the stimuli.

It has been observed that face verification relies less on the memory of the participant and more on the matching ability [35]. It mimics the real world scenarios where security personnel are often required to match an individual with his/her photograph on their identity proof document. In this study, participants were required to verify whether two given images belong to the same individual or not, while their brain activity was simultaneously being recorded using functional MRI. The face images constituting the stimuli belong to different face representations, namely *low-resolution* and *sketch* face images. As a part of this work, we analyzed how processing of cross-domain images differs from processing of same-domain face images of good quality, while performing the task of verification. It is observed that as compared to good quality face images captured in controlled, well-illuminated environment (termed as *controlled*, hereon), sketch image verification results in activation of regions associated with object processing tasks. Face sketch images are processed similar to high frequency information, whereas no such observation can be made for low resolution face images. Both the tasks are attributed as complex tasks. The effect of stimuli-participant ethnicity pairing is also analyzed for the two cross-domain face verification tasks, where a difference in processing is observed for low resolution face images, however, the verification of sketch face images remains unaffected. The current research extends the current pool of literature by providing novel insights into the functioning of the human brain while performing cross-domain face verification. Further, the results obtained as part of this research may also enable the development of novel human brain-inspired facial recognition algorithms.

II. Experimental Design

A. Subjects

23 participants (11 Caucasian and 12 Indian) took part in this study (12 females and 11 males) within the age range of 18-40 years having normal or corrected-to-normal vision. All participants gave an informed written consent for this study, which has been approved by the IIIT-Delhi Ethics board and the West Virginia University Institutional Review Board.

B. Task and Stimuli

Each participant was shown a pair of face images (stimulus) and was asked to perform the task of face verification. Using a response button, the participant had to respond whether the two images presented belonged to the same person (genuine) or different persons (impostor). The entire experiment consisted of three runs, where each run consisted of 60 stimuli. Each run begins with the user fixating on the screen, which is followed by the presentation of the first image for the first stimulus, for 1 second. Following this, the second image is simultaneously displayed beside the first image for 2.5 seconds. The participant is expected to give their response (genuine or impostor) during this time via the controller. As soon as the participant responds, the screen turns blank for the remaining time of 2.5 seconds. This is followed by a blank screen for another second, thereby resulting in jittered inter-stimuli-interval. Fig. 1(b) presents the timeline of one stimulus. This cycle is repeated 60 times for a run. The face images in Fig. 1(b) are representative of the stimuli used in the experiment. These images belong to the co-authors of this paper who have given informed consent to use these images for publication purposes. There are three categories of face image pairs shown to a participant, two for cross-domain face verification and one for same-domain verification: (1) *controlled-low resolution*: a regular face image captured in a controlled setting and a low resolution face representation, (2) *controlled-sketch*: a regular face image captured in a controlled setting and a sketch face representation, and (3) *controlled-controlled*: two regular face images captured in controlled settings. Each participant was shown 60 pairs for each category. For the face verification task, the pair of stimulus face images can belong to the same person (genuine) or different persons (impostor).

C. Stimuli Creation and Pre-processing

Most of the stimulus face images are collected from publicly available face datasets [36], [37], while some images are collected from the Internet as well. Each participant's familiarity response to the stimuli shown was also recorded, in terms of the participant being familiar with the face stimulus. Since only Indian and Caucasian subjects participated in this study, the stimuli also had only Caucasian and Indian ethnicities, with a 70-30% split. The CSU toolbox [38] is used for pre-processing the face images to remove extra artifacts by covering the background, such that only the face region is visible to the participants. The interface for stimuli presentation and response collection was created using the PsychToolbox (http://psychtoolbox.org/) and MATLAB platform.

D. fMRI Data Acquisition

Structural and functional MRI scans were collected using a 3 Tesla Siemens and 3 Tesla General Electric MRI scanner (for Caucasian and Indian participants, respectively) with a 32 channel head coil. The functional scans (T2*-weighted MRI) were collected using a gradient echo sequence, from the negative to positive direction, with 35 axial slices (slice thickness = 3.5mm, slice spacing = 0.0mm, repetition time (TR) = 2.0 seconds, echo time (TE) = 30ms, flip angle = 65^0, field of view = 224mm, matrix = 64×64). 128 volumes were captured per run, with no time gap between two volumes. High resolution anatomical scans were also collected for each subject. The T1-weighted sequences consist of 172 sagittal slices in interleaved sequence (slice thickness = 1mm, TR = 600ms, flip angle = 10^0, field of view = 224mm, matrix = 256×256).

E. Data Analysis

Pre-processing and voxel-based analysis of the fMRI data has been performed using the Statistical Parametric Mapping toolbox v12 [39]. As part of pre-processing, slice time correction and realignment is performed on all the scans such that each subject's scans are spatially aligned with their first scan. This is followed by normalization and co-registration, where all subjects' scans are mapped onto a common T2* template, in order to eliminate the structural variations between subjects. Finally, spatial smoothing is applied by using a Gaussian kernel having a width of 8mm at Full Width at Half Maximum (FWHM) [40], [41]. A General Linear Model (GLM) is fitted on the pre-processed data for all subjects with the six motion parameters regressed out and the predicted time series model was convolved with a canonical Haemodynamic Response Function. This is followed by group-level random effects analysis for the given experiment.

As mentioned previously, this study attempts to understand the human brain behavior for cross-domain face verification when controlled face images are compared with low resolution or sketch based face representation. Two sample t-tests are performed on the imaging data to draw comparisons between brain activity during cross-domain and controlled (same-domain) face verification tasks, along with between the two cross-domain tasks. Furthermore, conjunction analysis is performed to explore brain regions which show activations during both cross-domain and controlled face verification tasks, as well as during both cross-domain tasks.

Analysis is also performed to understand the effect of ethnicity for the two cross-domain face verification scenarios. For a Caucasian participant, a stimulus containing face images of Caucasian ethnicity creates a same ethnicity stimulus-participant pair, while a stimulus containing face images of another ethnicity results in a different ethnicity stimulus-participant pair. Similarly, same and different ethnicity stimulus-participant pairs are created for the Indian ethnicity participants. For each kind of face representation (sketch and low resolution) two contrasts are analyzed, same-ethnicity > different-ethnicity, and different-ethnicity > same-ethnicity. Behavioral analysis has also been performed for all the experiments to evaluate the participants' performance. Processed data used in this study can be obtained from the corresponding author upon request.

III. Behavioral Results

The mean verification accuracies obtained for controlled-controlled, controlled-sketch, and controlled-low resolution face verification are 77.20%, 63.86%, and 59.54%, respectively. This showcases the challenging nature of cross-domain face verification, wherein human performance for low resolution and sketch face image verification is lower as compared to same-domain controlled face verification. In order to evaluate the statistical difference of the results, two-sample t-test was performed. As compared to controlled-controlled face verification, results of both cross-domain face verification were statistically different (p-value < 0.001). While comparing the results of controlled-sketch and controlled-low resolution face verification, a p-value of 0.01 was obtained, which corresponds to statistical difference at a confidence level of 95%. To understand the effect of familiarity, the percentage of familiar stimuli across participants is calculated using the familiarity

TABLE I: Areas of activation observed for (a-c) controlled-sketch face verification, in comparison to controlled-controlled verification, and (d-f) controlled-low resolution face verification, as compared to controlled-controlled verification. Different contrasts are analyzed to understand regions of activation for specific tasks only. Peak cluster co-ordinates have been reported.

Cortical Areas	MNI Coordinates x	y	z	T Value	Cluster Size
(a) Controlled-sketch $>$ Controlled-controlled					
R Superior Frontal Gyrus	28	22	18	5.08	588
R Middle Frontal Gyrus					
R Inf. Frontal Gyrus, Tri.					
R Middle Cingulate					
L Insula	-28	34	10	4.50	22
R Thalamus	-4	-2	16	4.08	66
R Insula	30	14	-12	3.8	59
R Putamen					
(b) Controlled-sketch $<$ Controlled-controlled					
R Inferior Occipital Gyrus	28	-88	0	3.61	33
R Fusiform Gyrus					
R Calcarine Gyrus					
R Middle Occipital Gyrus					
(c) Controlled-sketch AND Controlled-controlled					
L Inferior Temporal Gyrus	-40	-66	-16	5.91	1823
L Fusiform Gyrus					
L Superior Occipital Gyrus					
L Cuneus					
L Calcarine Sulcus					
L Inferior Occipital Gyrus					
L Lingual Gyrus					
(d) Controlled-low resolution $>$ Controlled-controlled					
R Insula	30	14	-12	4.05	53
R Middle Frontal Gyrus	26	38	18	3.82	398
R Superior Frontal Gyrus					
R Sup. Frontal Gyrus, Med.					
R Thalamus	14	-32	14	3.22	
R Anterior Cingulate	8	48	12	2.92	13
(e) Controlled-low resolution $<$ Controlled-controlled					
L Hippocampus	-28	-14	-8	3.55	13
R Inferior Occipital Gyrus	28	-88	0	3.33	14
R Middle Occipital Gyrus					
R Calcarine Sulcus					
L Inf. Frontal Gyrus, Tri.	-46	32	0	2.94	8
(f) Controlled-low resolution AND Controlled-controlled					
L Inferior Temporal Gyrus	-40	-66	-16	6.48	1707
L Fusiform Gyrus					
L Superior Occipital Gyrus					
L Lingual Gyrus					
L Calcarine Sulcus					
L Middle Occipital Gyrus					
L Inferior Occipital Gyrus					

responses. It is observed that the participants best identified familiarity with stimuli of controlled-controlled face verification (35.37%), followed by controlled-sketch verification (26.81%) and finally controlled-low resolution face verification (4.54%). In order to eliminate the bias incurred due to the familiarity of stimuli, behavioral analysis was also performed on the unfamiliar stimuli only. A similar trend in accuracies was observed, where humans performed best on controlled faces, followed by sketch faces, and low resolution face images (73.91%, 63.71%, 60.25%), demonstrating statistical difference at 95% confidence level (p-value < 0.05).

Difference was also observed in the verification accuracy when participants belonged to the same-ethnicity versus to a different-ethnicity from the stimulus, while performing low resolution face verification. For the same-ethnicity stimuli, the participants obtained a verification accuracy of 67.91%, whereas different-ethnicity stimuli resulted in a verification accuracy of 58.00% (p-value < 0.001 for two-sample t-test). However, a similar difference was not observed for controlled-sketch face verification, where the same-ethnicity verification resulted in an accuracy of 68.43%, while the different-ethnicity verification yielded an accuracy of 69.41% (p-value = 0.9665). This may be attributed to the fact that sketch face images primarily contain edge information, therefore encoding little or no ethnicity information.

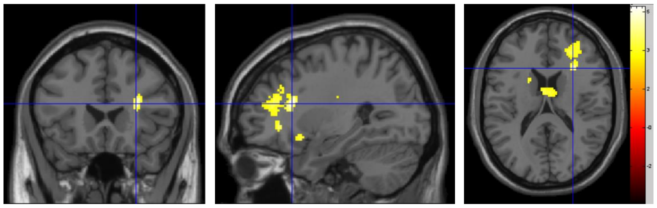
IV. Neuroimaging Results

A. *Verification of Sketch Face Images*

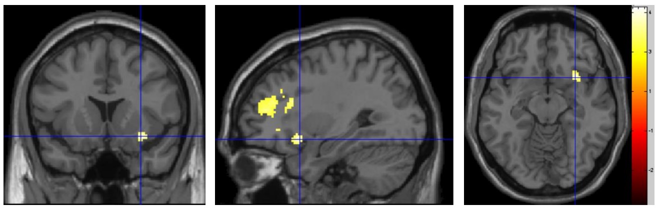
In order to understand how humans perform cross-domain sketch verification differently from controlled face verification, analysis was performed for different contrasts of controlled-sketch and controlled-controlled face verification. Table I(a-c) presents the key areas of activation identified using the AAL atlas [42] at a p-value < 0.005, uncorrected, labeled with clusters of brain regions (clusters having less than 5 voxels were ignored). For the contrast [controlled-sketch $>$ controlled-controlled image verification], larger response was observed in regions of Right Frontal Gyrus (Superior, Middle, and Inferior), Right Middle Cingulate Cortex, Bilateral Insula, Right Thalamus, and Right Putamen (Fig. 2). The opposite contrast ([controlled-sketch $<$ controlled-controlled face verification]) revealed greater response in regions of Right Occipital Gyrus (Inferior, Middle), Right Fusiform Gyrus, and Right Calcarine Gyrus. In order to further understand the human behavior while performing face verification, conjunction analysis was also performed. Table I(c) presents the regions obtained from the conjunction analysis (contrast: [controlled-controlled AND controlled-sketch face verification]).

B. *Verification of Low Resolution Face Images*

Experiments are performed in order to understand the differences in brain activity while performing face verification of low resolution images, as compared to controlled face verification. Table I(d-f) presents regions of activations obtained under different contrasts of controlled-controlled face verification and controlled-low resolution face verification for p-value < 0.005, uncorrected. As before, AAL atlas [42] was used for performing cluster labeling. It was observed that the areas of Right Insula, Right Frontal Gyrus (Middle, Superior), Right Thalamus, and regions of Right Cingulate Cortex displayed higher activations for low resolution face verification, as compared to controlled face verification (contrast: [controlled-low resolution $>$ controlled-controlled face verification]). Table I(e) also presents regions of increased activation for the reverse contrast ([controlled-low resolution face verification $<$ controlled-controlled face verification]). In addition, the table reports regions observed while performing the conjunction analysis on the two tasks as well.



(a) Controlled-Sketch Verification > Controlled-Controlled Verification



(b) Controlled-Low Resolution Verification > Controlled-Controlled Verification

Fig. 2: Peak activations observed while comparing controlled-sketch verification and controlled-low resolution verification, with controlled-controlled face verification. Cross-hair points at the peak activation observed for both the contrasts.

C. Difference between Low Resolution and Sketch Face Verification Tasks

In order to further understand and evaluate the differences between the two types of cross-domain face verification tasks, contrasts between them are also studied. Table II(a-b) presents the key regions of activation obtained for the contrast of [controlled-sketch > controlled-low resolution face verification], and its opposite. Under p-value < 0.005, uncorrected, regions such as the Right Middle Frontal Gyrus, Left Cuneus, and Right Hippocampus show increased activation for sketch verification, as compared to low resolution face verification (Fig. 3). On the other hand, regions such as the Left Angular Gyrus, Left Inferior Parietal gyrus, and Left Postcentral Gyrus present increased activation for low resolution face verification, as compared to sketch face verification.

D. Effect of Ethnicity

Table II(c-d) presents the areas of activation observed when participants perform same-ethnicity and different-ethnicity face verification of controlled-low resolution images. The contrast of [same-ethnicity > different-ethnicity] results in activation of regions such as Left Inferior Occipital Gyrus, Left Middle Temporal Gyrus, and Right Middle Occipital Gyrus, among others. These are obtained for a p-value $<$0.005, uncorrected, under cluster labeling. Similarly, [same-ethnicity < different-ethnicity] illicit higher activation in regions such as the Left Anterior Cingulate Gyrus, Left and Right Precentral Gyrus. On the other hand, in case of controlled-sketch face verification, no regions of activation was observed for the above mentioned contrasts for an uncorrected p-value < 0.005. The behavioral results also show a difference of less than 1% for same and different ethnicity stimulus-participant pairs, for controlled-sketch face verification.

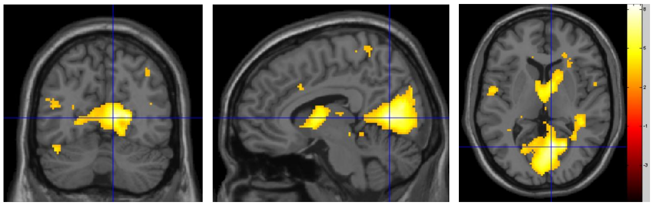
TABLE II: Areas of activation observed for (a-b) controlled-low resolution face verification, as compared to controlled-sketch verification, and (c-d) controlled-low resolution face verification while the participants performed same-ethnicity or different-ethnicity face verification. Peak cluster co-ordinates have been reported.

Cortical Areas	MNI Coordinates x	y	z	T Value	Cluster Size
(a) Controlled-sketch > Controlled-low resolution					
R Middle Frontal Gyrus	8	-64	-6	7.88	11238
L Cuneus					
R Hippocampus					
L ParaHippocampal					
L Thalamus					
(b) Controlled-sketch < Controlled-low resolution					
L Angular Gyrus	-34	-42	42	6.24	649
L Inferior Parietal Gyrus					
L Postcentral Gyrus					
L Superior Parietal Gyrus					
R Inferior Parietal Gyrus	36	-40	40	5.88	1236
R Middle Occipital Gyrus					
R Superior Parietal Gyrus					
(c) Same-ethnicity > Different-ethnicity					
L Inferior Occipital Gyrus	-32	-86	-2	6.42	422
L Middle Temporal Gyrus					
R Middle Occipital Gyrus					
R Fusiform Gyrus					
R Calcarine Sulcus					
(d) Same-ethnicity < Different-ethnicity					
L Anterior Cingulate Cortex	8	-80	0	10.62	26843
R Precuneus					
L Precuneus					
L Fusiform Gyrus					
R Middle Frontal Gyrus					
R Superior Frontal Gyrus					

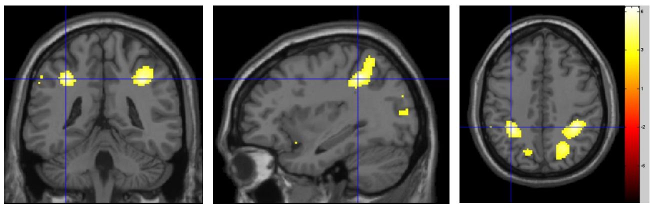
V. Discussion

A. Verification of Sketch Face Images

Sketch faces images, unlike photographs of faces, capture only the outlines (edges) of facial features, often comprising primarily of sharp edges or high spatial frequency. Upon analyzing the regions obtained for the contrast [controlled-sketch face verification > controlled-controlled face verification] (Table I(a)), it was observed that most of these regions have previously been shown to be responsible for object processing over face processing [18], [43], as well as complex matching tasks involving higher visual load [43], [44]. These findings suggest that humans perceive (high spatial frequency) sketch face images more as objects, as compared to faces, and treat sketch face verification as a more complex task as compared to controlled face verification. Most of the regions obtained for the reverse contrast ([controlled-controlled > controlled-sketch] face verification) indicate higher low frequency processing [45], along with higher processing of faces [18], [46]. This is consistent with the results obtained above, since sketches contain only high spatial frequency information, therefore, low frequency processing is expected to be missing while performing sketch verification. Moreover, as observed, since sketch face images may be perceived as objects, it was expected that controlled-sketch face verification would yield lower activations in face processing regions than controlled-



(a) Controlled-Sketch > Controlled-Low Resolution Verification



(b) Controlled-Low Resolution > Controlled-Sketch Verification

Fig. 3: Peak activations observed for the contrasts of controlled-sketch face verification versus controlled-low resolution face verification. In both cases, cross-hair points at the location of peak activation.

controlled face verification. Greater activations were also observed in regions associated with viewing familiar faces [16], [46]. This is in agreement with the behavioral results obtained in this study, where we observed 36.31% stimuli familiarity for controlled face verification, as opposed to 27.75% for sketch face verification. The regions of activation observed for the conjunction of these two tasks are primarily responsible for high frequency processing and face processing [47]–[49]. Combined, the above observations enable us to understand how humans perform face sketch verification, as compared to controlled face verification. Face sketch images may be perceived as objects, activate regions responsible for complex matching tasks, and present lack of activation in low frequency processing areas.

B. Verification of Low Resolution Face Images

Table I(d-f) presents the areas of activations obtained for different contrasts of controlled-controlled and controlled-low resolution face verification. For the contrast of [controlled-low resolution face verification > controlled-controlled face verification], regions of increased activation may primarily be attributed to object (over face) perception, performance of complex matching tasks, and visual load [43], [48]. Since low resolution images contain both low and high frequency information, regions pertaining to only a single kind of frequency processing cannot be expected here. The reverse contrast of [controlled-controlled face verification > controlled-low resolution face verification] results in increased activation in regions of Left Hippocampus, Right Occipital Gyrus, Right Calcarine Sulcus. The increased activation in Left Hippocampus may be attributed to the percentage of familiarity calculated above (35.37% for controlled and 4.54% for low resolution face verification) [50], [51]. As observed in literature, regions of Right Occipital Gyrus (Inferior, Middle), and Right Calcarine Sulcus may be attributed to upright face recognition and greater face processing [18], [49]. The increased activation in face processing related areas for controlled-controlled face verification as compared to controlled-low resolution face verification further strengthens the inference that humans perceive low resolution face images more as objects, or as individual face components, as opposed to perceiving the face in a holistic manner. Further, the conjunction analysis between these two tasks revealed activation of areas responsible for both low and high frequency face processing, along with regions responsible for decision making. This can be attributed to the fact that while low resolution images appear blurred, there exist traces of both low and high frequency facial information.

C. Difference between Low Resolution and Sketch Face Verification

The analysis between controlled-sketch face verification and controlled-low resolution face verification (Table II(a-b)) further provides insights into the difference in processing of these representations. Areas which showed higher activation for controlled-sketch verification as compared to controlled-low resolution face verification are primarily responsible for familiar face processing [50], [51]. On the other hand, regions showing activation for the reverse contrast ([controlled-low resolution > controlled-sketch verification]) are primarily responsible for low frequency processing [47], [52].

D. Effect of Ethnicity

The key regions of activation observed for different contrasts of same-ethnicity and different-ethnicity face verification for low resolution face images are given in Table II(c-d). Face network regions such as the Right Middle Occipital Gyrus and Right Fusiform Gyrus showed higher activations for same-ethnicity stimuli as compared to different-ethnicity stimuli. Similar results have also been observed in literature for experiments of own-race over other-race face categorization [23]. These regions have also shown to be responsible for identity coding and face processing. Similarly, regions of Left and Right Precuneus have shown significant activation for the reverse contrast of [different-ethnicity > same-ethnicity]. Regions of Left and Right Precuneus as well as Anterior Cingulate Cortex have shown higher activations for other-race faces as compared to own faces [23], [53]. These regions have shown to be responsible for self versus other processing and other episodic memory processes [54], [55]. In case of controlled-sketch face verification, no voxel survived the uncorrected p-value < 0.005 for cluster labeling. Coupled with less than 1% difference in verification accuracy of humans for same-ethnicity and different-ethnicity stimuli for sketch verification, it can be inferred that processing sketch images is not highly influenced by ethnicity, likely due to the reduced ethnic information in the outline-style drawings.

VI. Conclusion and Impact on Automated Face Recognition

This research presents a fMRI study for understanding the neural responses of the human behavior while performing cross-domain face verification. The study has been conducted across two ethnic groups and two cross-domain representations, namely, sketch to digital image matching and cross-resolution face matching. Comprehensive behavioral and neuro-imaging based results have been provided to draw analysis on the human behavior. It can be observed that as compared to controlled face verification, humans perform the task of face verification differently when comparing face images belonging to varying domains. Sketch face verification results in more activation of areas responsible for high frequency processing and complex matching tasks. Similarly, difference in processing was observed for controlled-low resolution face images, where higher activations were observed in areas of high visual load and complex matching task. Analysis has also been performed to understand the effect of same or different ethnicity of the stimulus while performing verification under varying face representations. Owing to the reduced ethnicity details in sketch images, no difference in brain processing was observed for high frequency sketch images related to ethnicity. However, significant difference in activation of different regions was observed for low resolution face verification, for varying ethnicity.

While recent research has focused on (i) learning representations for fMRI data [56]–[58], and (ii) on developing automated same-domain [59], [60] or cross-domain facial recognition systems [61]–[63], limited research has focused on understanding the human behavior while performing cross-domain face verification. To this effect, we believe that the current study extends the existing pool of literature by providing novel insights into the working of the human brain, thus facilitating researchers to develop brain-inspired algorithms. The variations observed while matching different face representations (sketch or low-resolution) provides better understanding of the salient features used by the human brain during matching. Such insights can thus enable the development of a stronger, more accurate, and more robust face recognition system, capable of matching across domains as well. While in this research we have tried to control the variations observed in different stimuli, there may still be some confounding factors that are not modeled, such as the difficulty level. As part of the future work, we intend to extend this research to understand the effect of different difficulty levels and more unconstrained stimuli for cross-domain face verification.

VII. Acknowledgement

This research is partially supported from a grant via the Ministry of Electronics and Information Technology, India. S. Nagpal is supported via the TCS PhD fellowship, and M.Vatsa is partially supported through the Swarnajayanti Fellowship by the Government of India.

References

[1] J. V. Haxby, E. A. Hoffman, and M. I. Gobbini, "The distributed human neural system for face perception," *Trends in cognitive sciences*, vol. 4, no. 6, pp. 223–233, 2000.

[2] O. Pascalis, X. de Martin de Viviés, G. Anzures, P. C. Quinn, A. M. Slater, J. W. Tanaka, and K. Lee, "Development of face processing," *Wiley Interdisciplinary Reviews: Cognitive Science*, vol. 2, no. 6, pp. 666–675, 2011.

[3] W. A. Freiwald, "The neural mechanisms of face processing: cells, areas, networks, and models," *Current Opinion in Neurobiology*, vol. 60, pp. 184–191, 2020.

[4] G. McCarthy, A. Puce, J. C. Gore, and T. Allison, "Face-specific processing in the human fusiform gyrus," *Journal of Cognitive Neuroscience*, vol. 9, no. 5, pp. 605–610, 1997.

[5] I. Gauthier, M. J. Tarr, J. Moylan, P. Skudlarski, J. C. Gore, and A. W. Anderson, "The fusiform "face area" is part of a network that processes faces at the individual level," *Journal of Cognitive Neuroscience*, vol. 12, no. 3, pp. 495–504, 2000.

[6] O. L. Lopatina, Y. K. Komleva, Y. V. Gorina, H. Higashida, and A. B. Salmina, "Neurobiological aspects of face recognition: The role of oxytocin," *Frontiers in Behavioral Neuroscience*, vol. 12, p. 195, 2018.

[7] K. M. O'craven, P. E. Downing, and N. Kanwisher, "fMRI evidence for objects as the units of attentional selection," *Nature*, vol. 401, no. 6753, pp. 584–587, 1999.

[8] B. A. S. Hasan, M. Valdes-Sosa, J. Gross, and P. Belin, ""hearing faces and seeing voices": Amodal coding of person identity in the human brain," *Scientific reports*, vol. 6, p. 37494, 2016.

[9] I. Muukkonen, K. Ölander, J. Numminen, and V. R. Salmela, "Spatio-temporal dynamics of face perception," *NeuroImage*, vol. 209, 2020.

[10] Z. Zhen, H. Fang, and J. Liu, "The hierarchical brain network for face recognition," *PLOS ONE*, vol. 8, no. 3, pp. 1–9, 03 2013.

[11] K. Grill-Spector, N. Knouf, and N. Kanwisher, "The fusiform face area subserves face perception, not generic within-category identification," *Nature neuroscience*, vol. 7, no. 5, 2004.

[12] B. Wang, T. Yan, S. Ohno, S. Kanazawa, and J. Wu, "Retinotopy and attention to the face and house images in the human visual cortex," *Experimental brain research*, vol. 234, no. 6, pp. 1623–1635, 2016.

[13] B. Wang, T. Li, Y. Niu, J. Xiang, J. Cheng, B. Liu, H. Zhang, T. Yan, S. Kanazawa, and J. Wu, "Differences in neural responses to ipsilateral stimuli in wide-view fields between face-and house-selective areas," *PloS one*, vol. 13, no. 2, 2018.

[14] M. Bernstein, Y. Erez, I. Blank, and G. Yovel, "An integrated neural framework for dynamic and static face processing," *Scientific reports*, vol. 8, no. 1, 2018.

[15] M. Tsantani, N. Kriegeskorte, K. Storrs, A. L. Williams, C. McGettigan, and L. Garrido, "FFA and OFA encode distinct types of face identity information," *Journal of Neuroscience*, 2021.

[16] S. M. Platek and S. M. Kemp, "Is family special to the brain? an event-related fMRI study of familiar, familial, and self-face recognition," *Neuropsychologia*, vol. 47, no. 3, pp. 849 – 858, 2009.

[17] S. M. Landi and W. A. Freiwald, "Two areas for familiar face recognition in the primate brain," *Science*, vol. 357, no. 6351, pp. 591–595, 2017.

[18] R. Watson, E. M. J. Huis in 't Veld, and B. de Gelder, "The neural basis of individual face and object perception," *Frontiers in Human Neuroscience*, vol. 10, p. 66, 2016.

[19] N. Kanwisher, F. Tong, and K. Nakayama, "The effect of face inversion on the human fusiform face area," *Cognition*, vol. 68, no. 1, pp. B1 – B11, 1998.

[20] K. Sergerie, C. Chochol, and J. L. Armony, "The role of the amygdala in emotional processing: A quantitative meta-analysis of functional neuroimaging studies," *Neuroscience & Biobehavioral Reviews*, vol. 32, no. 4, pp. 811 – 830, 2008.

[21] C. A. Cushing, H. Y. Im, R. B. Adams, N. Ward, D. N. Albohn, T. G. Steiner, and K. Kveraga, "Neurodynamics and connectivity during facial fear perception: The role of threat exposure and signal congruity," *Scientific reports*, vol. 8, no. 1, 2018.

[22] L. Q. Uddin, J. T. Kaplan, I. Molnar-Szakacs, E. Zaidel, and M. Iacoboni, "Self-face recognition activates a frontoparietal "mirror" network in the right hemisphere: an event-related fMRI study," *NeuroImage*, vol. 25, no. 3, pp. 926 – 935, 2005.

[23] W. Wei, J. Liu, R. Dai, L. Feng, L. Li, and J. Tian, "Different brain activations between own- and other-race face categorization: an fMRI study using group independent component analysis," *Proceedings SPIE*, vol. 9038, 2014.

[24] H. Farmer, M. Hewstone, O. Spiegler, H. Morse, A. Saifullah, X. Pan, B. Fell, J. Charlesford, and S. Terbeck, "Positive intergroup contact modulates fusiform gyrus activity to black and white faces," *Scientific reports*, vol. 10, no. 1, 2020.

[25] D. White, R. I. Kemp, R. Jenkins, M. Matheson, and A. M. Burton, "Passport officers' errors in face matching," *PLOS ONE*, vol. 9, no. 8, pp. 1–6, 08 2014.

[26] D. White, P. J. Phillips, C. A. Hahn, M. Hill, and A. J. O'Toole, "Perceptual expertise in forensic facial image comparison," *Proceedings of the Royal Society B: Biological Sciences*, vol. 282, no. 1814, p. 20151292, 2015.

[27] D. J. Robertson, E. Noyes, A. J. Dowsett, R. Jenkins, and A. M. Burton, "Face recognition by metropolitan police super-recognisers," *PloS one*, vol. 11, no. 2, 2016.

[28] L. G. Ungerleider, S. M. Courtney, and J. V. Haxby, "A neural system for human visual working memory," *Proceedings of the National Academy of Sciences*, vol. 95, no. 3, pp. 883–890, 1998.

[29] S. J. Teipel, A. L. Bokde, C. Born, T. Meindl, M. Reiser, H.-J. Möller, and H. Hampel, "Morphological substrate of face matching in healthy ageing and mild cognitive impairment: a combined mri-fmri study," *Brain*, vol. 130, no. 7, pp. 1745–1758, 2007.

[30] D. Yadav, N. Kohli, S. Nagpal, M. Singh, P. Pandey, M. Vatsa, R. Singh, and A. Noore, "Region-specific fMRI dictionary for decoding face verification in humans," in *International Joint Conference on Neural Networks*, 2017.

[31] L. D. Harmon, "The recognition of faces," *Scientific American*, vol. 229, no. 5, pp. 70–83, 1973.

[32] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, "Face recognition by humans: Nineteen results all computer vision researchers should know about," *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1948–1962, 2006.

[33] D. Yadav, R. Singh, M. Vatsa, and A. Noore, "Recognizing age-separated face images: Humans and machines," *PLOS ONE*, vol. 9, no. 12, pp. 1–22, 12 2014.

[34] V. Goffaux and B. Rossion, "Faces are" spatial"–holistic face perception is supported by low spatial frequencies." *Journal of Experimental Psychology: Human perception and performance*, vol. 32, no. 4, 2006.

[35] A. M. Burton, D. White, and A. McNeill, "The glasgow face matching test," *Behavior Research Methods*, vol. 42, no. 1, pp. 286–291, 2010.

[36] H. S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa, "Memetically optimized mcwld for matching sketches with digital face images," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 5, pp. 1522–1535, 2012.

[37] P. J. Flynn, K. W. Bowyer, and P. J. Phillips, "Assessment of time dependency in face recognition: An initial study," in *International Conference on Audio-and Video-Based Biometric Person Authentication*, 2003, pp. 44–51.

[38] D. S. Bolme, J. R. Beveridge, M. Teixeira, and B. A. Draper, "The CSU face identification evaluation system: Its purpose, features, and structure," in *International Conference on Computer Vision Systems*, 2003, pp. 304–313.

[39] "Statistical parametric toolbox (SPM12)," http://www.fil.ion.ucl.ac.uk/spm/, 2014.

[40] K. J. Friston, A. P. Holmes, J. Poline, P. Grasby, S. Williams, R. S. Frackowiak, and R. Turner, "Analysis of fmri time-series revisited," *Neuroimage*, vol. 2, no. 1, pp. 45–53, 1995.

[41] Z. Chen and V. Calhoun, "Effect of spatial smoothing on task fmri ica and functional connectivity," *Frontiers in neuroscience*, vol. 12, p. 15, 2018.

[42] E. T. Rolls, M. Joliot, and N. Tzourio-Mazoyer, "Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas," *NeuroImage*, vol. 122, pp. 1 – 5, 2015.

[43] S. Schwartz, P. Vuilleumier, C. Hutton, A. Maravita, R. J. Dolan, and J. Driver, "Attentional load and sensory competition in human vision: Modulation of fMRI responses by load at fixation during task-irrelevant stimulation in the peripheral visual field," *Cerebral Cortex*, vol. 15, no. 6, p. 770, 2005.

[44] T. Klingberg, B. T. O'Sullivan, and P. E. Roland, "Bilateral activation of fronto-parietal networks by incrementing demand in a working memory task." *Cerebral Cortex*, vol. 7, no. 5, p. 465, 1997.

[45] E. Eger, P. G. Schyns, and A. Kleinschmidt, "Scale invariant adaptation in fusiform face-responsive regions," *NeuroImage*, vol. 22, no. 1, pp. 232 – 242, 2004.

[46] G. Pourtois, S. Schwartz, M. L. Seghier, F. Lazeyras, and P. Vuilleumier, "View-independent coding of face identity in frontal and temporal cortices is modulated by familiarity: an event-related fMRI study," *NeuroImage*, vol. 24, no. 4, pp. 1214 – 1224, 2005.

[47] P. Rotshtein, P. Vuilleumier, J. Winston, J. Driver, and R. Dolan, "Distinct and convergent visual processing of high and low spatial frequency information in faces," *Cerebral Cortex*, vol. 17, no. 11, p. 2713, 2007.

[48] T. Iidaka, K. Yamashita, K. Kashikura, and Y. Yonekura, "Spatial frequency of visual image modulates neural responses in the temporo-occipital lobe. an investigation with event-related fMRI," *Cognitive Brain Research*, vol. 18, no. 2, pp. 196 – 204, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0926641003002490

[49] Z. Zhen, H. Fang, and J. Liu, "The hierarchical brain network for face recognition," *PLOS ONE*, vol. 8, no. 3, pp. 1–9, 03 2013.

[50] J. J. Gold, C. N. Smith, P. J. Bayley, Y. Shrager, J. B. Brewer, C. E. L. Stark, R. O. Hopkins, and L. R. Squire, "Item memory, source memory, and the medial temporal lobe: Concordant findings from fMRI and memory-impaired patients," *Proceedings of the National Academy of Sciences*, vol. 103, no. 24, pp. 9351–9356, 2006.

[51] J. V. Haxby, L. G. Ungerleider, B. Horwitz, J. M. Maisog, S. I. Rapoport, and C. L. Grady, "Face encoding and recognition in the human brain," *Proceedings of the National Academy of Sciences*, vol. 93, no. 2, pp. 922–927, 1996.

[52] C. Peyrin, M. Baciu, C. Segebarth, and C. Marendaz, "Cerebral regions and hemispheric specialization for processing spatial frequencies during natural scene recognition. an event-related fMRI study," *NeuroImage*, vol. 23, no. 2, pp. 698 – 707, 2004.

[53] L. Feng, J. Liu, Z. Wang, J. Li, L. Li, L. Ge, J. Tian, and K. Lee, "The other face of the other-race effect: An fMRI investigation of the other-race face categorization advantage," *Neuropsychologia*, vol. 49, no. 13, pp. 3739 – 3749, 2011.

[54] A. D. Wagner, B. J. Shannon, I. Kahn, and R. L. Buckner, "Parietal lobe contributions to episodic memory retrieval," *Trends in cognitive sciences*, vol. 9, no. 9, pp. 445–453, 2005.

[55] R. Cabeza, F. Dolcos, S. E. Prince, H. J. Rice, D. H. Weissman, and L. Nyberg, "Attention-related activity during episodic memory retrieval: a cross-function fMRI study," *Neuropsychologia*, vol. 41, no. 3, pp. 390–399, 2003.

[56] K. Kusano, T. Tashiro, T. Matsubara, and K. Uehara, "Deep generative state-space modeling of FMRI images for psychiatric disorder diagnosis," in *International Joint Conference on Neural Networks*, 2019.

[57] A. El Gazzar, L. Cerliani, G. van Wingen, and R. M. Thomas, "Simple 1-D convolutional networks for resting-state fMRI based classification in autism," in *International Joint Conference on Neural Networks*, 2019.

[58] S. R. Oota, V. Rowtula, M. Gupta, and R. S. Bapi, "StepEncog: A convolutional LSTM autoencoder for near-perfect fMRI encoding," in *International Joint Conference on Neural Networks*, 2019.

[59] A. Majumdar, R. Singh, and M. Vatsa, "Face verification via class sparsity based supervised encoding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1273–1280, 2017.

[60] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4685–4694.

[61] H. S. Bhatt, R. Singh, M. Vatsa, and N. K. Ratha, "Improving cross-resolution face matching using ensemble-based co-transfer learning," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5654–5669, 2014.

[62] S. Ouyang, T. Hospedales, Y.-Z. Song, X. Li, C. C. Loy, and X. Wang, "A survey on heterogeneous face recognition: Sketch, infra-red, 3D and low-resolution," *Image and Vision Computing*, vol. 56, pp. 28–48, 2016.

[63] S. Nagpal, M. Singh, R. Singh, and M. Vatsa, "Discriminative shared transform learning for sketch to image matching," *Pattern Recognition*, vol. 114, p. 107815, 2021.