Cambridge Large Two (2020), 1–4

ARTICLE TYPE

Using Deep CNN to test Human Face recognition Optimization and Holistic Effect

Guneesh Vats, Abhay Patil, and Pradhuman Tiwari

Abstract

This research paper aims to investigate the use of Deep Convolutional Neural Networks (CNN) in testing Human Face recognition Optimization and Holistic Effect. Holistic effect is the behavioral signature in which humans are better at recognizing faces as a whole rather than by individual parts or facial features. The methodology of this study involved conducting a Target Matching Task (face recognition task) for humans with faces that were dissected and not dissected, and we observed a drop in accuracy due to the Holistic Effect. The CNN was then trained on both complete faces to establish a baseline and dissected faces to investigate whether similar drops in accuracy occurred in face recognition tasks. The study also conducted various iterations of the face recognition task, one of which involved providing images of very similar-looking people as prompts to match with the target image. The results of the study showed a significant drop in accuracy in face recognition tasks for both humans and CNNs trained on both complete and dissected images. The conclusion drawn from this research suggests that CNNs show characteristic features of human facial recognition, including the Holistic Effect, indicating that holistic effect signature in human face recognition is a consequence of optimization.

Keywords: Holostic Effect, Face recognition, Deep CNN

1. Introduction

We introduce the concept of the "Holistic Effect" as a novel signature in trained Deep CNN models, which posits that humans are more proficient at identifying faces in their entirety rather than by individual facial features. In essence, this effect asserts that the perception of facial features is highly influenced by the context in which they are presented, and that processing the entire face provides important cues and information that are not readily available when examining individual parts in isolation.

We suspect that this feature in humans is the result of optimization of the face recognition task so we are going to train Deep CNN models to test whether they exhibit the same properties or not because if a specific human behavior is consequence of optimization of a task then we should see similar phenomena in Deep CNNs that are optimized for that same task.

Holistic Effect : Humans are better at recognizing faces as a whole rather than by individual parts(facial features).

2. Literature Review

Face recognition is a fundamental task that humans excel at, but the underlying mechanisms and processes are still not fully understood. In recent years, deep learning models, particularly deep convolutional neural networks (CNNs), have shown great potential in face recognition tasks. The present review is based on two papers that explore the underlying mechanisms of human face recognition and how they relate to deep learning models.

The first paper, "*Inductive biases for deep learning of higher level cognition*", proposes a framework for incorporating inductive biases into deep learning models to improve their ability to learn higher level cognitive tasks. The authors argue that current deep learning approaches may not be sufficient for learning tasks that require higher level cognition, and that inductive biases are important for guiding learning in the absence of large amounts of data. The authors provide an overview of inductive biases and how they can help guide learning by constraining the space of hypotheses that need to be considered. They also propose a taxonomy of inductive biases relevant to higher-level cognition, including biases related to causality, agency, social cognition, and conceptual knowledge.

The second paper, "*The whole advantage: Holistic face recognition over part-based face recognition*", investigates whether humans process faces holistically, meaning whether they recognize faces as a whole rather than by individual parts. The authors conducted several experiments to test this hypothesis, including one where they asked participants to recognize faces that were presented either as whole faces or as scrambled faces (where individual parts were mixed up). The results showed that participants were better at recognizing faces when they were presented as whole faces than when they were presented as scrambled faces, supporting the hypothesis of holistic face processing. The authors suggest that this holistic processing may be due to specialized neural mechanisms in the brain that are specifically tuned to process faces as a whole.

From the foundation paper for our idea "Using Deep CNN to test why to test why human face recognition works the way

it does" tells us about the Inversion effect and The other race effect which gives us the first glimpse at the consequence of the task of face recognition in human brain they prove their hypothesis by answering these three questions in 4 different types of trained CNNs - Trained on only faces, trained on both faces and objects, trained on objects taking face as one category and with an untrained CNN:
a. Does human-like face recognition performance reflect optimization for face recognition in CNN
b. Do CNNs represent faces in a similar fashion to humans?
c. Do CNNs show classic signatures of human face processing?
We are using a similar approach to The aim of our project is to build on these two papers by using deep CNNs to test the optimization and holistic effect of human face recognition

3. Methodology

3.1 Dataset

We are using an open source dataset of the 16 personalities for the priliminary Target Matvhing task. 8 Men and 8 Women all are famous personalities of hollywood and all of them are caucasion so we rule out the any effect of "*Other Race Effect*".

Task Details

There is one target image which is shown to the participant and they have to tell which of the 2 other options match with the target image. All images were of same gender and race which confirms that these 2 features did not help the participant to identify it easy in both target matching tasks when images were whole as well as dissected into separate facial features.

How CNNs Performed the task?

They first obtained the Activation patterns and computed the correlation distance b/w activation patterns of each pair of images (1 - pearson's r) and the distance closest to the target image is taken as the right choice by the network.

3.1.1 Dissected Images

In order to generate a dataset of dissected images from normal images, we followed a series of steps. First, we performed facial keypoint detection, which is a computer vision task that involves detecting and localizing specific points on a face, including the corners of the eyes and mouth, the tip of the nose, and the extremities of the eyebrows. These points, also known as facial landmarks, are essential for various face-related applications, such as face recognition, facial expression analysis, and gaze tracking.

Next, we utilized the facial landmarks obtained in the previous step to separate full-face images into smaller individual images containing only the left eye, right eye, nose, and mouth regions. This process is known as separation of facial features and resulted in a new dataset where each person is represented by a collection of images of these distinct facial parts.

3.2 Face recognition tasks [Target Matching Task]

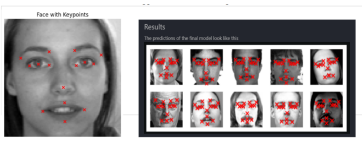

Figure 1. facial keypoint recognition

3.2.1 For humans

We conducted a survey to gather data on the accuracy of humans on the target matching task for both normal and dissected faces, where target images were normal in both cases. We tested 25 participants on 10 face recognition tasks, with 5 tasks requiring them to match one of two normal images with a normal target image (Figure 2), and the other 5 tasks requiring them to match one of two dissected faces with a normal target image (Figure 3). One potential concern with this survey is that the prompt images used were of famous Hollywood celebrities, which could have resulted in participants performing exceptionally well on the target matching task.

We also conducted a similar survey for an another face recognition task, only this time the prompt images were of people who had similar faces

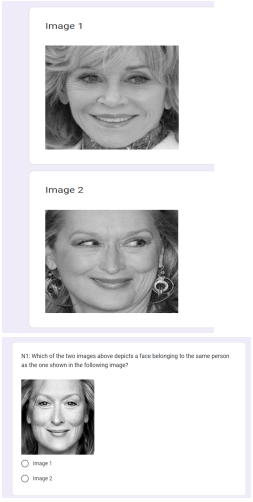

Figure 2. Normal Face recognition task

3.2.2 For CNN

Discuss how CNNs were used for face recognition task In this study, we employed a deep convolutional network (CNN) to perform the face recognition task. Specifically, we used the VGG16 architecture as a pretrained model and froze the existing layers. We construct the next Flattened layer based on the output coming from the previous pretrained VGG layer. We then add an output layer containing 16 neurons, which corresponds to the number of classes (i.e., persons) in

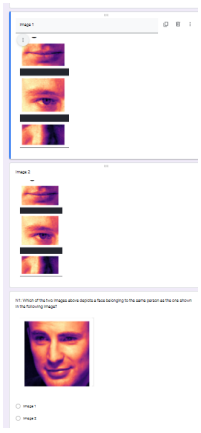

Figure 3. Dissected face recognition task

our dataset. The activation function for this layer was set to softmax, and the optimizer used was Adam. Initially we found that the model, despite having an accuracy of 0.96, performed poorly on the validation data. To improve the generalization capability of our model, we augmented the face images using rotation, noise, brightness, horizontal and vertical shift. This augmentation allowed our model to learn more robust features and fixed the problem of overfitting, with a validation accuracy of 0.96.

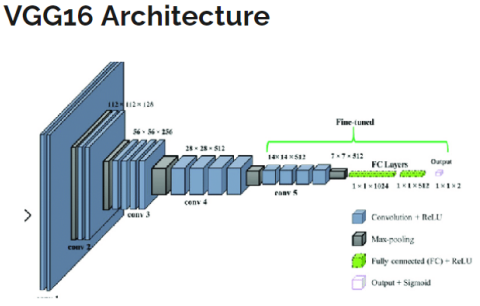

Figure 4. VGG

4. Results

As one can see in the figure 5, the accuracy of both humans and CNN dropped considerably in the face recognition task when the prompt images were dissected.

The iteration in which we trained our model on complete images and found that there was no significant difference in the accuracy drop in face recognition tasks compared to when the model was trained on dissected images. However, it is

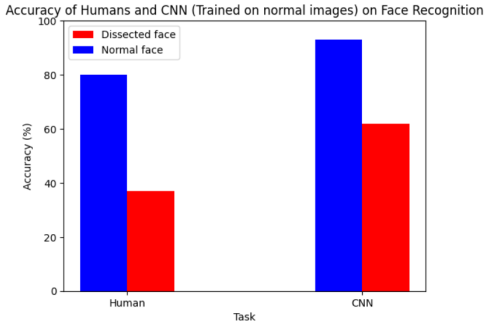

Figure 5. VGG

worth noting that the accuracy drop was slightly less when trained on dissected images, as shown in Figure 6.

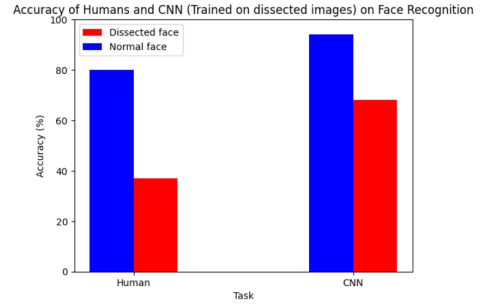

Figure 6. VGG

Lastly, our survey and CNN revealed that when prompts containing similar images to match to a target dissected image and a target normal image were used, the accuracy of both humans and CNN decreased in both iterations, as demonstrated in Figure 7.

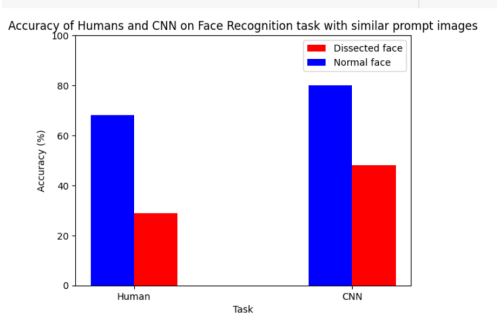

Figure 7. VGG

5. Discussion

Comparison of Target matching task:

From the obtained plots of the performance of both CNN and Humans we can clearly observe the difference in accuracies. CNN is able to achieve near to the human accuracy when whole images are given but when it is tested on dissected images the accuracy drops for both the CNNs and

humans. When we train the CNNs on dissected images we see they perform good in dissected image matching taks but the accuracy is still similar in whole faces as compared to previous task when CNN was trained differently.

Analysis on Similar faces pair:
In this project, we aimed to evaluate the effect of image similarity on face recognition performance in both human participants and a Deep CNN model. We designed a target matching task that included a dataset of images with varying degrees of similarity, and calculated similarity scores between pairs of images using a distance metric. We then recruited a sample of human participants and trained a Deep CNN model on the same dataset. Participants performed the target matching task online through Micrsoft Forms, and the model was evaluated on the same task. Performance was analyzed on different sets of images with varying degrees of similarity. Our analysis showed that image similarity had a significant effect on face recognition performance, with lower similarity leading to higher accuracy for both humans and the Deep CNN model. Our results provide insights into the effect of image similarity on face recognition performance and highlight the importance of considering image similarity when designing target matching tasks. And when we dissected the images we observed the similar patter with even lower accuracies than this considering the features are dissected and it anyways lowers the chances of matching to correct target image.

Attention Perspective of the Modified Target Matching task:
Using eye tracking equipment and software to analyze which features of faces humans look at during a target matching task is based on the idea that eye movements provide valuable information about where a person is directing their attention. Eye tracking technology allows us to measure where a person is looking on a screen, and to record the timing and duration of their fixations and saccades. Participants will be asked to perform a target matching task on a computer screen, where they will be presented with three images - one target image and two distractors. They will be asked to identify the image that matches the target image. Meanwhile, the eye tracking equipment will record their eye movements and generate a series of fixations and saccades. By analyzing the fixation patterns across participants, researchers can identify which facial features attract the most attention during the task. For example, if participants tend to fixate on the eyes more frequently than other features, this suggests that the eyes are an important cue for target matching. To conduct this experiment, the researcher will need to select appropriate eye tracking equipment and software, recruit a sample of participants, and design a task that elicits the target matching behavior. The data collected will then be analyzed using statistical techniques to identify patterns in gaze behavior and the frequency of fixations on specific facial features. This information can then be used to refine the design of target matching tasks and improve their accuracy and efficiency. Brain regions activation (FMRI analysis):

6. Summary

We can conclude that CNN's trained on face recognition perform similar to Humans. From previous work we know that humans represent faces in a similar way as face-recognition optimized CNN models. This is the first indication that the Holistic Effect might be a consequence of Optimization of task of Face recognition.

7. Code

Link to the code : https://drive.google.com/drive/folders/1UkqvQy2A57
szH9aBulksfSwTBLlrqp?usp=sharing

8. References

1. Dobs, K., Yuan, J., Martinez, J., Kanwisher, N. (2022). Using deep convolutional neural networks to test why human face recognition works the way it does. https://doi.org/10.1101/2022.11.2
2. Tanaka, J. W., Farah, M. J. (1993). The whole advantage: Holistic face recognition over part-based face recognition. Journal of Experimental Psychology: General, 122(4), 422-433. https://doi.org/10.1037/0096-3445.122.4.422