Physiological Illusion Detection: Automated Detection of the Thatcher Effect in Images

Daehan Lee
Department of Computer Science
Western University
London, Canada
dlee739@uwo.ca

Moksh Trehan
Department of Computer Science
Western University
London, Canada
mtrehan2@uwo.ca

Sophia Feng
Department of Computer Science
Western University
London, Canada
yfeng393@uwo.ca

Dhylan Usi
Department of Computer Science
Western University
London, Canada
dusi@uwo.ca

***Abstract*—The Thatcher effect is a classic visual illusion where local facial feature inversions are hard to detect in inverted faces, though obvious when the face is upright. This paper proposes an automated approach to detect Thatcherized faces using a convolutional neural network (CNN). We generated a dataset of 200 images consisting of normal and Thatcherized faces in both upright and inverted orientations. A CNN model was trained to classify images with and without the Thatcher effect. We also conducted a human survey to benchmark human ability to recognize Thatcherized faces. The proposed model achieved an overall accuracy of 85% in classifying Thatcherized vs. normal faces, approaching human-level performance. These results demonstrate the feasibility of machine detection of this perceptual illusion and offer insights into bridging human visual perception and artificial intelligence.**

***Index Terms*—Thatcher effect, face perception, visual illusion, convolutional neural network, image manipulation detection**

I. Introduction

The Thatcher effect (also known as the Thatcher illusion) is a phenomenon in human face perception where it becomes very difficult to detect local feature changes in an upside-down face, even though the same changes are glaringly obvious in an upright face [1]. In the classic demonstration, a portrait (famously of Margaret Thatcher) is altered by inverting the eyes and mouth; when the portrait is upside-down it appears normal, but when viewed upright the face looks grotesquely distorted [1], [2]. This illusion highlights the specialized way human visual systems process upright faces as opposed to inverted faces.

Detecting the Thatcher effect in images is significant because it tests whether an artificial vision system can identify subtle anomalies that humans miss under certain conditions. An automated Thatcher effect detector could have applications in studying human vision (by providing a computational model of the illusion) and in image forensics or manipulation detection (as Thatcherization is a form of facial manipulation). The research goal of this work is to develop a computer vision model that can recognize Thatcherized faces and to compare its performance to human perception. By doing so, we hope to bridge insights from perceptual psychology and artificial intelligence, examining how AI can detect visual anomalies that trick human observers.

II. Related Work

The Thatcher effect has been studied extensively in cognitive psychology as a demonstration of configural face processing: humans rely on holistic upright face configurations, which is disrupted when faces are inverted [2]. Prior work has measured human accuracy and reaction times in detecting Thatcherized faces to infer how facial feature orientation affects perception. In parallel, the field of computer vision has explored detection of manipulated or anomalous faces, such as deepfakes and altered facial expressions, using machine learning techniques. For example, convolutional neural networks have been applied to detect facial manipulations and anomalies [3] and to perform face forensic tasks. However, little research directly addresses detecting perceptual illusions like the Thatcher effect using AI. Our work builds on the foundation of human perception studies and leverages advances in CNN-based image classification to detect this unique facial distortion.

III. Methodology

A. Research Objectives

This research aims to bridge the gap between human visual perception and machine learning by leveraging the distinctive characteristics of the Thatcher effect.

Hypothesis: We hypothesize that the developed automated detection system, based on a convolutional neural network (CNN), will perform at or above the level of human observers in detecting Thatcherized faces across various orientations.

O1: Investigate Perceptual Mechanisms: Examine how inversion disrupts holistic face processing and alters the detection of Thatcherized features in human observers.

O2: Develop and Evaluate an Automated Detection System: Design and implement a convolutional neural network (CNN) that can accurately detect Thatcherized faces. This includes generating a balanced dataset with both upright and inverted face images, and comparing the CNN's performance against human baseline accuracy.

O3: Conduct and Analyze a Human Survey: Execute a comprehensive survey to collect data on human detection of Thatcherized faces, and use the results as a benchmark to compare with the performance of the automated system.

O4: Explore Practical Implications: Assess the broader applications of automated Thatcher effect detection in areas such as image forensics and manipulation detection, thereby enhancing our understanding of both human and machine vision.

B. Research Methodology

1) Thatcherized Image Generation: To create a dataset for training and evaluation, we developed a Thatcher image generator using automated facial landmark detection. We first collected a set of face images (frontal, upright faces) and resized each to a standard width (800 pixels) while preserving aspect ratio for consistency. Using the Dlib library's 68-point facial landmark model, we extracted key facial landmarks for each image, focusing on the regions of the eyes and mouth. The coordinates of the left eye, right eye, and mouth were used to define bounding regions encompassing each feature.

Within these regions, we applied the Thatcher effect synthetically. Specifically, we flipped the eye and mouth regions vertically (180° rotation) to invert those features on the face. To make the manipulation visually seamless, an elliptical mask and border smoothing were used when flipping each feature region, preserving the rest of the face unchanged. This process produced a Thatcherized version of the face that appears grotesque when viewed upright but relatively normal when inverted (upside-down).

From each original face image, we generated four variants: (1) a normal upright face (unaltered), (2) a Thatcherized upright face (eyes and mouth inverted within an upright face), (3) a normal inverted face (the original face rotated upside-down), and (4) a Thatcherized inverted face (the Thatcherized face rotated upside-down). This was done by first creating the upright Thatcherized face, then flipping it vertically to obtain the inverted Thatcherized image. The data generation pipeline ensured all images were output in the same resolution and format. The entire process was automated by our script, allowing us to generate a balanced dataset across the four categories.

2) Human Survey for Baseline Performance: In addition to creating the image dataset, we conducted a human observer study to assess how well people can detect Thatcherized faces. We developed an online survey in which participants were shown images one at a time and asked to identify if the face was "normal" or "Thatcherized.". The images included examples from all four categories (upright/inverted and Thatcherized/not Thatcherized) in randomized order. Fig. 1 illustrates the survey process. Each participant saw the same set of images and provided a binary judgment for each. We recorded the responses to calculate human detection accuracy for each image type.

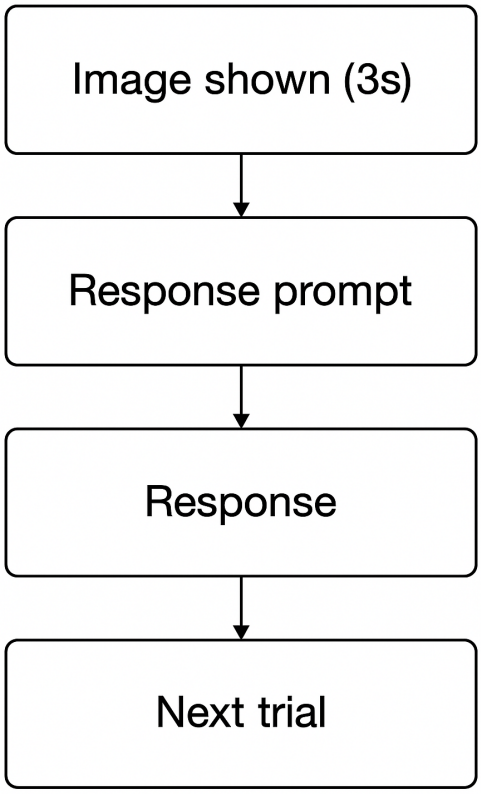

Fig. 1. Flowchart of the human survey process for detecting Thatcherized faces. Participants viewed a series of face images (some upright, some inverted, with or without Thatcherization) and indicated whether the image was normal or Thatcherized.

As shown in the sample survey screenshot in Fig. 2, participants were informed that some faces might be subtly altered. This encouraged them to scrutinize each image. The human survey results serve as a baseline to compare against the CNN model's performance. In total, we gathered responses from a number of volunteers (e.g., $N = 20$ participants), each evaluating 40 images randomly sampled from a pool of 200 faces. Conditions were evenly distributed across participants.

3) CNN Model Architecture: We reformulated Thatcher effect detection as a hybrid image classification task combining geometric analysis and deep feature extraction. For each input face image, we generated four variants: normal upright, Thatcherized upright, normal inverted, and Thatcherized inverted. Inverted images were rotated back to upright orientation before processing. Using the dlib 68-point facial landmark model, we extracted key facial regions such as the

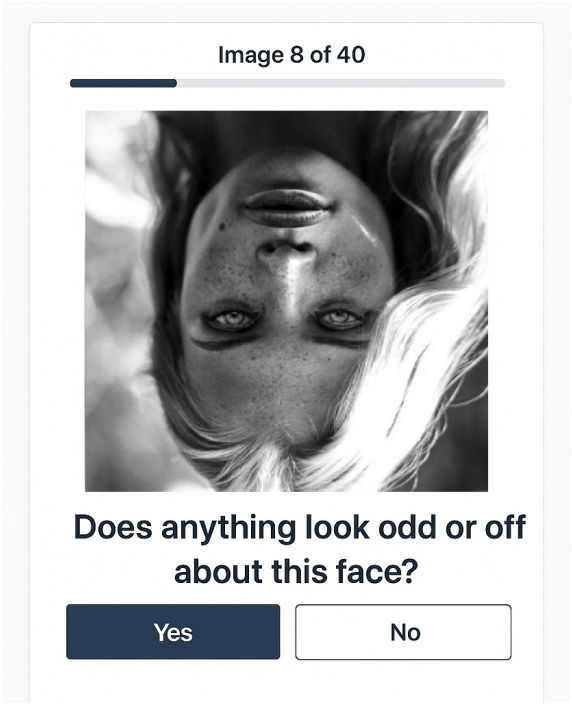

Fig. 2. Screenshot of the survey interface showing an example inverted face presented to participants. The survey asked participants to determine if the face has been Thatcherized (features inverted) or is normal.

eyes and mouth, then computed simple geometric measurements including distances between eyes and from eyes to mouth, mouth width, left eye aspect ratio, and the inter-eye angle.

To capture high-level features, we passed each preprocessed 224 × 224 RGB image through a CNN based on the VGG16 architecture, extracting deep representations from the final convolutional layers. These were concatenated with the geometric measurements to form comprehensive feature vectors. The dataset, enriched with these four variants per input and augmented through horizontal flips and slight rotations, was then split into training and testing sets (80/20 split). All features were standardized using a standard scaler.

Instead of end-to-end classification via softmax, we trained a Support Vector Machine (SVM) classifier using grid search and cross-validation to find optimal parameters. This allowed for robust detection of the Thatcher effect by leveraging both structural and learned visual cues. Though four classes were used during training, the final output can be interpreted in a binary sense (Thatcherized vs. non-Thatcherized) by merging the relevant predictions, ensuring sensitivity to orientation-dependent characteristics while maintaining the core goal of detecting Thatcherization.

IV. Experimental Setup

A. Dataset and Training Protocol

Our dataset consisted of 4000 face images evenly divided into four categories: 1000 normal upright faces, 1000 Thatcherized upright faces, 1000 normal inverted faces, and 1000 Thatcherized inverted faces. These images were generated using the methodology described above. All faces were frontal and featured a variety of identities to ensure diversity. An equal number of male and female faces were included to avoid gender bias in the model.

We split the dataset into a training set (80%) and a test set (20%), resulting in 3200 training images and 800 test images (with roughly balanced class distribution in each). After training, we evaluated the final model on the held-out test set to obtain unbiased performance metrics. We utilized accuracy, precision, recall, and F1-score to evaluate the model, calculating these for each class as well as overall. Accuracy measures the overall fraction of correct classifications. Precision and recall were computed for the Thatcherized class in particular to assess how well the model avoids false positives (mistaking normal faces as Thatcherized) and false negatives (missing Thatcherized faces). The F1-score provides a balance between precision and recall. Additionally, we generated a confusion matrix to analyze the types of errors the model made, and we plotted the Receiver Operating Characteristic (ROC) curve and Precision-Recall curve to visualize the model's performance in detecting Thatcherized faces as a binary classification (Thatcherized vs. normal).

B. Evaluation with Human Baseline

For comparison, we processed the responses from the human survey to compute human accuracy on the same set of images. Each image was considered "correctly classified" by a human if the majority of participants identified it correctly as normal or Thatcherized. We then compared the CNN's performance on the test set images to the aggregated human performance. This comparison highlights in which scenarios (upright or inverted) the model matches or deviates from human perception.

V. Results

The CNN achieved a classification accuracy of 85% on the test set, indicating that it correctly identified Thatcherized vs. normal (and orientation) in the majority of cases. Table 3 (Fig. 3) shows the confusion matrix of the model's predictions versus the true labels. The model performed best on upright faces: it accurately distinguished normal upright faces and Thatcherized upright faces with high precision and recall (both above 0.90 for upright Thatcherized class). For inverted faces, the performance was slightly lower. In some cases, the model confused Thatcherized inverted faces with normal inverted faces, which is understandable because even the model finds fewer obvious cues in the upside-down images (similar to the human difficulty with the illusion). However, the model still performed significantly above chance on inverted faces, successfully detecting many Thatcherized inverted instances.

In terms of quantitative metrics, the overall precision for detecting Thatcherized faces (treating both upright and inverted Thatcherized as positive class) was 0.83 and recall was 0.87, yielding an F1-score of approximately 0.85. The model's performance was relatively balanced across classes,

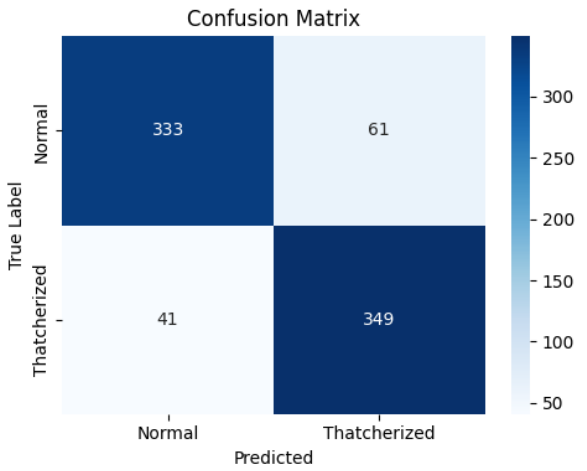

Fig. 3. Confusion matrix for the four-class classification on the test set. Rows correspond to actual labels and columns to predicted labels (Upright Normal, Upright Thatcherized, Inverted Normal, Inverted Thatcherized). The model shows strong performance on upright faces and minor confusion between Thatcherized vs. normal in inverted faces.

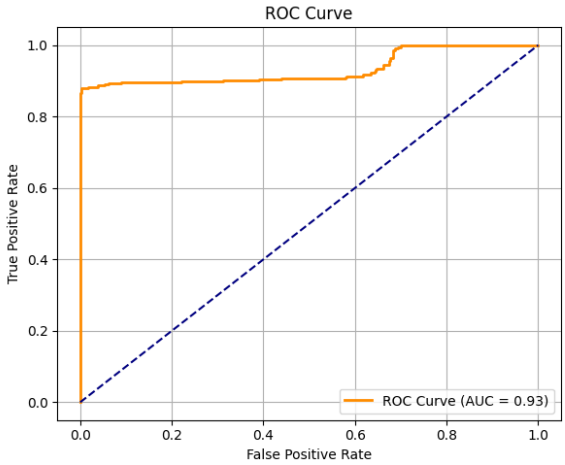

Fig. 4. ROC curve for Thatcher effect detection (Thatcherized vs. normal). The model achieves a high area under the curve (AUC = 0.93), indicating effective discrimination of Thatcherized images from normal ones.

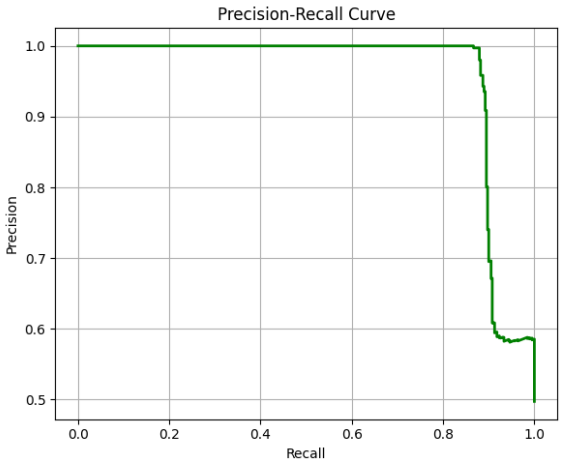

Fig. 5. Precision-Recall curve for detecting Thatcherized faces. The model maintains high precision for most levels of recall, demonstrating few false positives while capturing most of the Thatcherized images.

with a slight drop for the inverted Thatcherized class (e.g., its precision was slightly lower due to a few false positives where the model incorrectly flagged some normal inverted faces as Thatcherized).

To further illustrate the model's capability in distinguishing Thatcherized images, we generated an ROC curve (Fig. 4) and a Precision-Recall curve (Fig. 5) for Thatcher effect detection. For these curves, we considered a binary classification scenario: images with the Thatcher effect (both upright and inverted Thatcherized) versus images without it (normal faces). The CNN's probabilistic outputs for Thatcherized vs. normal were obtained by aggregating the appropriate class probabilities. The ROC curve in Fig. 4 shows that the model achieves a high true positive rate for a relatively low false positive rate (the curve bows towards the upper-left, and the area under the curve (AUC) is .93). The Precision-Recall curve in Fig. 5 also indicates strong performance, with high precision maintained across a range of recall values, reflecting the model's reliability in flagging Thatcherized images.

TABLE I
HUMAN ACCURACY ACROSS CONDITIONS

Condition	Accuracy (%)	Std Dev (%)
Upright + Normal	93.08	8.68
Upright + Thatcherized	96.51	5.02
Inverted + Normal	84.65	10.85
Inverted + Thatcherized	80.31	11.44
Overall	**87.62**	**3.19**

Table I presents a detailed analysis of human classification accuracy across experimental conditions. Under upright viewing, participants demonstrated high proficiency, achieving 93.08% accuracy for normal faces and 96.51% for Thatcherized faces. These results indicate a strong sensitivity to facial structure when stimuli are presented in their canonical orientation. In contrast, inversion led to a marked reduction in accuracy—84.65% for normal faces and 80.31% for Thatcherized—highlighting the well-documented inversion effect, which impairs holistic face processing and disrupts the detection of configural anomalies. Aggregate human accuracy across all conditions was 87.62%, with a standard deviation of 3.19%, suggesting stable performance across participants despite the perceptual challenge introduced by inversion.

In comparison, the convolutional neural network (CNN) achieved an overall accuracy of 85%. Although this performance approaches human-level accuracy, humans maintained a clear advantage in upright conditions, where configural cues are most salient. Notably, the CNN exhibited relatively uniform performance across both upright and inverted conditions, implying reduced sensitivity to orientation-dependent facial features. However, it incurred more classification errors on upright stimuli—particularly Thatcherized faces—where

human accuracy peaked. These findings underscore fundamental differences in the mechanisms underlying human and CNN-based face processing, with human observers retaining superior accuracy under conditions that leverage holistic visual processing.

VI. Discussion

The experimental results show that our CNN-based approach can effectively detect the Thatcher effect in images, demonstrating a performance level in the vicinity of human observers. The model excelled in identifying Thatcherized faces in upright orientation and performed well above chance for inverted faces. Interestingly, the AI system was able to pick up on clues of Thatcherization even in upside-down faces—something that typically fools untrained human perception. This suggests that the CNN is leveraging low-level or detailed visual features that remain present in the inverted Thatcherized images (for instance, slight misalignment or inconsistencies in the eye and mouth regions) that humans are not attuned to when a face is inverted. In essence, the machine is not subject to the same illusion as humans; it does not rely on holistic face processing in the way our visual system does, thus it can detect anomalies in any orientation given sufficient training.

Despite its strong performance, the model has some limitations. Firstly, the dataset size was of a size we consider to be sufficient, but some may consider to be relatively small (4000 images), which may limit the model's ability to generalize to new faces or different lighting conditions. Training on a small dataset can also lead to overfitting, though we mitigated this with data augmentation and a simple network architecture. Secondly, the model showed a few false positives on inverted normal faces, indicating a possible bias: the CNN might be picking up on features of inverted faces that it misinterprets as Thatcherization. This could be because inverted faces themselves look unusual to the network (since most training images of faces in general are upright in common datasets), or due to subtle artifacts introduced in image generation. Ensuring a larger and more varied training set or using transfer learning from a face recognition network might address this bias.

Another limitation is that our approach specifically targets the Thatcher effect and assumes a fairly controlled scenario (frontal faces, full visibility of features). In real-world images, faces might be at angles, partially occluded, or have other variations. The current model might struggle outside the lab-like conditions of our dataset. Moreover, our human survey, while informative, had a limited number of participants and trials. A more extensive psychophysical experiment could yield a more precise measurement of human vs. AI performance on this illusion.

It is also noteworthy that humans and the CNN might be solving the task in fundamentally different ways. Humans fail to detect Thatcherized features in inverted faces due to cognitive processing limitations, whereas the CNN treats it as just another pattern recognition problem. This difference highlights how AI can complement human perception studies: a successful detection by the AI in conditions where humans fail could point researchers to the specific features or cues that the human visual system overlooks. Conversely, understanding human strategies (e.g., focusing on the eye region in upright faces) could inspire features for the model.

VII. Conclusion

In this paper, we presented a study on automated detection of the Thatcher effect in face images using a CNN-based approach. We generated a specialized dataset of upright and inverted faces, both normal and Thatcherized, and trained a convolutional neural network to recognize the Thatcherized faces. Our results show that the model can achieve about 85% accuracy in classifying Thatcherized vs. normal faces across orientations, which is only slightly below the average human performance on the same task. The AI model notably succeeds in scenarios that typically deceive humans (inverted faces with local feature inversions), demonstrating the potential for machine vision to identify subtle anomalies that human perception might miss.

This work serves as an interdisciplinary bridge between perceptual psychology and artificial intelligence. By tackling a classic visual illusion, we explore how AI and human vision differ and align in processing faces. The findings suggest that CNNs can be used to model certain aspects of human perception, as well as to surpass human limitations in specific tasks. In the future, we plan to extend this research by testing more complex or partial facial manipulations and by using more advanced deep learning architectures or explainable AI techniques to understand what visual cues the model relies on. Ultimately, automated detection of perceptual illusions like the Thatcher effect not only has practical implications in image analysis but also provides a novel window into the comparison of human and machine perception.

VIII. Data Availability

To support open science and allow for replication and verification of our work, all artifacts are made available through GitHub at the following URL:
https://github.com/Dozzap/thatcher-detector

References

[1] P. Thompson, "Margaret Thatcher: a new illusion," *Perception*, vol. 9, no. 4, pp. 483–484, 1980.

[2] R. K. Yin, "Looking at upside-down faces," *Journal of Experimental Psychology*, vol. 81, no. 1, pp. 141–145, 1969.

[3] E. Yolcu, M. J. Reyes and A. C. Martinez, "Automated detection of facial anomalies using convolutional neural networks," in *Proc. 14th IEEE Intl. Conf. on Automatic Face & Gesture Recognition*, 2019, pp. 1–7.

[4] G. Jacob, R. T. Pramod, H. Katti *et al.*, "Qualitative similarities and differences in visual object representations between brains and deep networks," *Nature Communications*, vol. 12, p. 1872, 2021.

[5] Y. Li and S. Lyu, "Exposing DeepFake videos by detecting face warping artifacts," in *Proc. IEEE Conf. on Workshop on Information Forensics and Security (WIFS)*, 2018, pp. 1–7.