Eye Gaze Tracking:

Eye gaze is an important insight into human cognitive behavior which can be used for various applications.

Eye tracking finds applications in diverse domains, from human–computer interaction to medical diagnoses, psychological studies, and computer vision. It serves as an externally-observable indicator of human visual attention, making it valuable in various fields.

In the context of gaze estimation, convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are commonly employed to analyze spatial and temporal dependencies in eye images, enhancing the accuracy of predicting gaze direction.

Appearance-based gaze estimation methods utilize visual information captured from the human eye to infer gaze direction. Appearance-based approaches are emphasized over other methodologies, signaling their significance in leveraging deep learning for accurate gaze estimation.

GazeCapture Dataset :

There is success in deep learning in various computer vision domains but it has limited impact on eye tracking due to the scarcity of large-scale data.

There are public datasets on gaze estimation, which have information about gaze direction and head pose distributions as well. These provide valuable insights into the diversity and challenges within these datasets, for example Gaze360 does not have head information.

GazeCapture, a dataset collected through crowdsourcing, comprises 2,445,504 images from 1,474 participants using mobile phones or tablet[review]. Participants are tasked with gazing at a circle on the devices, allowing for unconstrained head movements. The dataset is known for covering various lighting conditions and head motions, making it suitable for evaluating unconstrained 2D gaze point estimation methods.

The dataset is designed to capture a diverse range of backgrounds, lighting conditions, and unconstrained head motion, addressing limitations in existing datasets.

The GazeCapture dataset is very useful in evaluating unconstrained 2D gaze point estimation methods.

iTracker, a CNN for gaze prediction, is trained using GazeCapture, achieving real-time performance on modern mobile devices.

Krafka et al provide technical details on the performance of iTracker, reporting more prediction errors without calibration and with calibration, these errors reduced proving that calibration is essential for fine-tuning the model's predictions.

The performance of iTracker is extensively evaluated using the GazeCapture dataset. The model outperforms existing approaches, achieving an average error of approximately 2 cm without calibration.

A comparison with other gaze datasets highlights GazeCapture's distinctions, such as encouraging head movement during recording, generating a random distribution of gaze points, and leveraging crowdsourcing for unprecedented scale. GazeCapture is positioned as a valuable resource for future work in the domain.

The work done with iTracker is positioned as a benchmark for the next generation of eye tracking solutions.

Lightweight architectures contribute to high performance on the GazeCapture dataset, achieving a prediction error of 1.591 cm on the test set with two eyes as input and 1.951 cm with just one eye[].

The proposed gaze estimation model employs ResNet and Inception architectures, making predictions using a single eye image.

Existing methods, while efficient, lack the utilization of advanced CV techniques like ResNet and Inception architectures and ensemble models. The proposed approach modifies the CNN architecture to incorporate ResNet, Inception, and InceptionResNet versions, enhancing gaze prediction accuracy. Data preprocessing involves using the GazeCapture dataset, extracting eye regions and landmarks. Training involves splitting the dataset into training, validation, and testing sets, employing the Adam optimizer[gazecapture dataset paper-1 (2)].

While GazeCapture has significantly advanced the field, limitations exist. The use of controlled tasks in data collection raises questions about generalizability to real-world scenarios. Additionally, accuracy may be hampered by factors like head pose variations and occlusions.

Ongoing research explores further improvements in accuracy, robustness, and applicability to diverse use cases. Integration with other sensors and personalized calibration procedures are promising avenues for future development.

The study in paper by Rishi et al, 2022 suggests the efficacy of ResNet and Inception architectures and ensemble calibration for gaze tracking, emphasizing potential applications in e-commerce, digital accessibility, and medical diagnosis.

[Optional]

Recent datasets proposed to enhance gaze estimation studies:

RT-Gene Dataset: Proposed by Fischer et al. in 2018[], this dataset provides accurate 3D gaze data collected with a dedicated eye tracking device, offering a valuable resource for precise gaze estimation research.

Gaze360 Dataset: Introduced by Kellnhofe et al. in 2019[], the dataset includes 238 subjects in indoor and outdoor environments, featuring 3D gaze information across a diverse range of head poses and distances. It contributes to understanding gaze behavior in varying scenarios.

ETH-XGaze Dataset: Proposed by Zhang et al. in 2020[], this dataset offers high-resolution images covering extreme head poses and includes 16 illumination conditions. It facilitates the exploration of illumination effects on gaze estimation and extends the scope of research in challenging conditions.