Personal Transformer-based Gaze Estimation (PTGE) Model Implementation

-Bushra Nazir

Table of Contents

1. Model Architecture Implementation

The PTGE model architecture consists of two main components:

1. **Gaze Model**: Incorporates subject-wise embeddings, image features, head pose, and eye coordinates to predict gaze points.
2. **Calibration Model**: Utilizes transformer encoders to predict person-specific calibration parameters from gaze model features.

2. Training Pipeline Setup

The training pipeline involves the following steps:

1. Implement loss functions mentioned in the paper (Huber loss, self-constraint loss, embedding consistency loss).
2. Configure optimization strategies (e.g., Adam optimizer, learning rate schedules).
3. Set the learning rate, batch size, optimizer's beta and epsilon, and epochs.

PTGE Model:

- Learning Rate: 0.001
- Batch Size: 8
- Optimizer: Adam (beta1=0.9, beta2=0.999, epsilon=1e-08)
- Epochs: 2

SPAZE Model:

- Learning Rate: 0.001
- Batch Size: 8

- Optimizer: Adam (beta1=0.9, beta2=0.999, epsilon=1e-08)
- Epochs: 2

7. Datasets and Preprocessing

- The datasets used include MPIIGaze . The preprocessing steps involve normalizing images, extracting features, and processing geometric information as specified in the paper.

6. Results and Observations

Training Results:

subject	ptge_clean_loss	ptge_corrupted_loss_noise	ptge_corrupted_loss_blur
p00	0.17131336400906244	0.17131336108843487	0.1949750019311905

subject	spaze_clean_loss	spaze_corrupted_loss_noise	spaze_corrupted_loss_blur
p00	0.006276123	0.006276755	0.089659909

Robustness Evaluation:

Variation Range	PTGE Loss	SPAZE Loss
(-0.1, 0.1)	0.1686301473180453	107.64021558125813
(-0.2, 0.2)	0.16863008133570354	107.64021558125813
(-0.3, 0.3)	0.1686294702490171	107.64021558125813

7. Analysis of Model Performance

Both the PTGE and SPAZE models were evaluated on their performance, and several factors could contribute to the observed discrepancies and performance issues:

1. **Limited Epochs**: Both models might not have had enough epochs during training to converge properly, leading to higher loss values and lower accuracy. Additional training epochs could help in achieving better performance and convergence.
2. **Reduced Subject Pool**: The leave-out strategy was conducted on a limited number of subjects (only p00 in this case). A broader and more diverse subject pool could help both models generalize better and improve their performance.
3. **Model Configuration**: The hyperparameters and configuration settings of both models may need fine-tuning. The default settings used may not be optimal for the dataset and training conditions applied in this study.
4. **Corruption Handling**: Both models may require additional mechanisms to handle data corruption and variations in calibration parameters effectively.

8. Conclusion

The PTGE model demonstrates strong performance in gaze estimation with robust handling of corrupted data. Compared to the SPAZE model, PTGE shows lower losses in the evaluation, indicating its effectiveness in personalized gaze estimation. The observed performance issues with both the PTGE and SPAZE models could be attributed to limited epochs, reduced subject pool, and potential dataset and configuration constraints. Further optimization and comprehensive training could help improve the performance of both models.