SynBlink and BlinkFormer: A Synthetic Dataset and Transformer-Based Method for Video Blink Detection

Bo Liu[1], Yang Xu[1], Feng Lu[1,2] ✉

[1] State Key Lab. of VR Technology and Systems, Beihang University, Beijing, China
[2] Peng Cheng Laboratory, Shenzhen, China

Motivation

- The dataset volume for blink detection remains relatively small due to the cost of data collection and annotation.
- There is still room for improvement in the accuracy of current algorithms.

Contribution

- **The Synthetic Data Workflow:** A controllable and flexible workflow for synthesizing human video clips with blinks, allowing for the unlimited generation of data with precise labeling.
- **SynBlink:** The largest video blink dataset available as of now.
- **BlinkFormer:** The first Transformer-based method that performs both blink detection for video clips and estimates the blink strength in frame level.

Blink Detection Datasets Overview

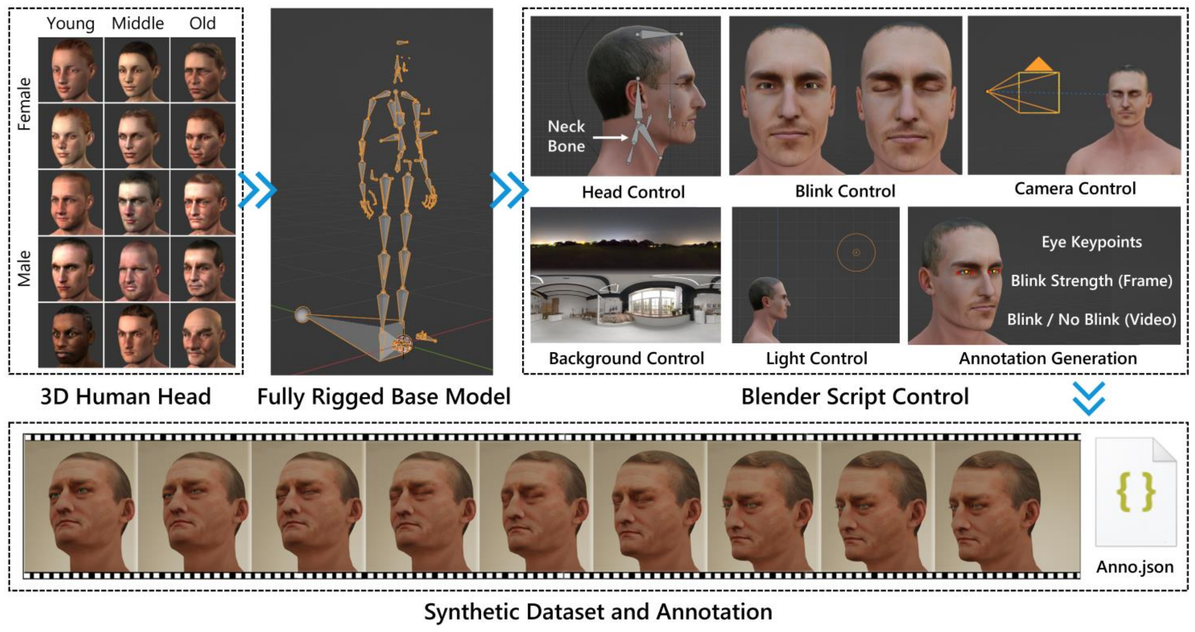
SynBlink: A Synthetic Dataset

- 50,000 video clips / 13 frames / 350×350 pixels / 126GB / 25085 blinking / 24915 non-blinking.
- Advantages:

  Large Data Volume: Provide feasibility and convenience for deep learning methods that require a massive amount of data.

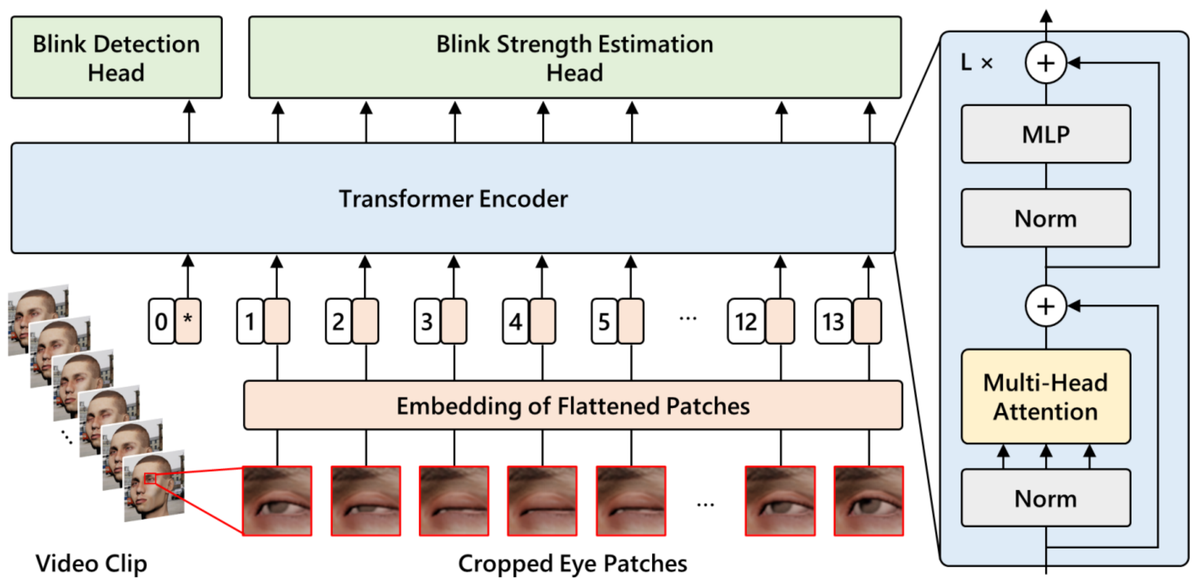
  Complex Scenarios: Different head postures, head movements, eye sizes, age, ethnicity, gender, light intensity, camera angles, background types.

  Accurate Annotations: Include whether the entire clip blinks or not, keypoints of the eyes in each frame, pupil center positions, and blink strengths.



Generation Workflow of SynBlink Dataset

BlinkFormer: A Transfomer-Based Method



Architecture of BlinkFormer

A blink detection head is used to classify the whole sequence into blink detection result, while another head output the blink strength of each frame.

Result

Comparation on HUST-LEBW dataset

Method	F1	Precison	Recall
Variance Map (ver.) [2]	51.58%	49.03%	54.41%
Variance Map (hor.) [2]	55.53%	52.25%	59.35%
Variance Map (flow.) [2]	47.10%	48.30%	46.02%
Eye Template [5]	33.28%	**98.28%**	20.12%
Motion Vector [13]	51.58%	49.03%	54.41%
EAR [4]	42.95%	61.15%	33.12%
LRCN [26]	**78.52%**	69.69%	**89.92%**
mEBAL [9]	75.42%	67.14%	**87.77%**
HUST [20]	78.18%	75.82%	80.69%
3D PBNN* [3]	**82.03%**	**78.36%**	86.07%
BlinkFormer (Ours)	**84.31%**	**82.06%**	**86.69%**

BMVC 2023

Data & Code

PHI-AI Lab