Detection and Correction of Photosensitive Epilepsy Triggers in Videos

A system for identifying and mitigating epileptic seizure triggers in video content

Berra Nur Öztürk (2220356190) – Kaan Eren (2210356121)

Aydın Kaya (Advisor) – Görkem Akyıldız (Vice Advisor)

Photosensitive Epilepsy
A Critical Accessibility Challenge

Affects approximately 5% of people with epilepsy

Triggered by specific visual patterns:

- Rapid flashing lights
- High contrast patterns
- Certain color combinations

Current digital media landscape lacks safety measures

Need for automated detection and correction systems

Research Objectives & Key Contributions

- Novel Detection Framework: First implementation of transformer-based temporal modeling for photosensitive trigger detection
- Adaptive Correction Algorithm: Context-aware modification system preserving content quality while ensuring safety
- Comprehensive Evaluation: Extensive testing using PEAT validation and real-world video datasets
- Accessibility Impact: Practical solution for making digital media safer for photosensitive individuals

Methodology Overview
Processing Pipeline has 6 Main Stages

- Video Input & Preprocessing
- Feature Extraction
- Temporal Modeling
- Trigger Classification
- Adaptive Correction
- Quality Validation

Stage 1 - Video Input & Preprocessing

Frame Extraction:

30-frame sliding windows with 15-frame stride

50% overlap for comprehensive temporal coverage.

Prevents missing critical transitions

Preprocessing Steps:

BGR to RGB conversion

Resize to 224×224 pixels

ImageNet normalization for optimal feature extraction

Stage 2 - Feature Extraction (CNN-Based Spatial Feature Extraction)

- CNN Backbone
  - EfficientNet-B0 extracts 1280-dimensional spatial features
  - Captures both local patterns and global scene characteristics
- Feature Processing
  - Linear projection layer reduces dimensionality to 1024
  - Efficient transformer processing while preserving critical visual information

Stage 3 - Temporal Modeling (Transformer Architecture for Sequence Analysis)

Transformer Configuration

3-layer encoder with 4-head multi-attention mechanism

Models complex temporal dependencies across 30-frame sequences

Positional Encoding:

Learnable embeddings provide crucial temporal order information

Essential for accurate pattern recognition

Stage 4 - Trigger Classification (Binary Decision Making)

Classification System

Sigmoid classifier with optimized threshold (0.5)	Balances sensitivity and specificity for reliable detection

Confidence Scoring:

Probability outputs enable fine-tuned decision making	Uncertainty quantification for robust performance

Stage 5 - Adaptive Correction (Context-Aware Modification Techniques)

Parameters

- **Flash Detection:** Threshold of 180 on scale 0-255
- **Intense Color Detection:** Threshold of 200 on scale 0-255 for RGB channels.
- **Flicker Detection:** Average brightness across frame, with threshold of 30.
- **Flicker Frequency:** If there are 3 or more detections in a second

Pixel Variance Analysis for Flicker

- Stacking greyscale frames and calculating the variance of each pixel overtime.
- A steady pixel will have low variance like someone’s face.
- A rapid changing area like a disco ball will have high variance.
- If there is a low variance area in front a high variance one the low variance will be left as is, only high variance area will be masked.

Intense Color Detection

- If it has 200 on scale of 0-255 and if its intensity is at least x1.5 larger than other channels which makes sure we only flag dominant colors.
- Pixels of frames which either include intense red or blue are flagged true.
- Morphological operations are done to the masks to connect nearby pixels that should be the part of the same mask area.
- Saturation reduction is done using masks.

Stage 6 - Quality Validation

PEAT Verification:

- Automated validation using Photosensitive Epilepsy Analysis Tool
- Ensures corrected content meets safety standards

Model Architecture Details (Hybrid CNN-Transformer Design)

Video Input Processing

Batch size 4, input tensor (B, 30, 3, 224, 224)

Reshape to (B×30, 3, 224, 224) for CNN processing

CNN Feature Extraction

EfficientNet-B0 (ImageNet pretrained)

Linear projection: 1280 → 1024 dimensions

0.3 dropout rate for regularization

Transformer Encoder

3 layers, 4-head attention, d_model=1024, d_ff=4096

Layer normalization with residual connections

Training Configuration (Optimization and Data Handling)

Training Setup

- AdamW optimizer: learning rate 3e-5, weight decay 1e-4
- Reduce learning rate on plateau

Data Augmentation

- Random sequence reverse by 50% of change

Loss Function

- Binary cross-entropy

Performance Results

Performance Metrics

Accuracy: 88.3% on test set
Precision: 90.7% (low false positives)
Recall: 94.2% (high sensitivity)
F1-Score: 92.4% balanced performance

Test Dataset

Total Sequences: 273 test samples
Correct Predictions: 241 sequences
True Positives: 195 trigger detections
True Negatives: 46 safe classifications

System Implementation (Technical Implementation Details)

Processing Efficiency

- 30-frame sliding windows with 15-frame stride
- Gradient checkpointing and mixed precision training

Safety Assurance

- PEAT validation for all corrected content
- Automated safety standard compliance checking
- Context-preserving corrections without content loss

Demonstration Results (Before & After Comparisons)

Correction Effectiveness

- Adaptive overlay opacity based on content brightness
- Advanced flickering detection and temporal smoothing
- Intelligent color intensity reduction for red/blue channels
- Context-preserving corrections without content loss
- Deep learning-based trigger detection with transformer architecture
- PEAT validation for safety assurance

Visual Quality Preservation

- Maintains content integrity while ensuring safety
- Minimal impact on viewer experience
- Professional-grade video processing

Future Work and Limitations (Current Limitations and Future Directions)

Current Limitations

Application works offline instead of being real time

Limited to 30-frame temporal windows

Requires further validation on diverse content types

Future Work

Real-time processing optimization

Extended temporal window analysis

Deep learning models for correction algorithm

Conclusion
(Summary and Impact)

Research Achievements

Successfully developed transformer-based detection system

Achieved 88.3% accuracy with 94.2% recall

Created adaptive correction algorithm preserving content quality

Validated effectiveness using PEAT standards

Societal Impact

Enhances digital accessibility for photosensitive individuals

Provides practical solution for content creators

Contributes to safer digital media landscape

Technical Contribution

Novel application of transformer architecture

Context-aware correction methodology

Comprehensive evaluation framework

Appendix - Technical Specification (Detailed Technical Specifications)

Model Architecture

Input: 30-frame sequences (224×224×3)

CNN: EfficientNet-B0 backbone

Transformer: 3 layers, 4 heads, 1024 dimensions

Output: Binary classification (safe/dangerous)

Training Details

Batch size: 4 sequences (120 frames)

Learning rate: 3e-5 with cosine annealing

Regularization: 0.3 dropout, weight decay 1e-4

Hardware: GPU-accelerated training

Performance Benchmarks

Processing speed: ~0.4 seconds per 30-frame sequence

Memory usage: ~8GB VRAM for batch processing

Model size: ~175 MB compressed

THANK YOU FOR LISTENING