Deep Learning Detection of Subtle Dynamic Ocular Torsion from Video Ocular Counter Roll (vOCR)

Tianyi Ye[1], Krishna Mukunda[1,2], Yi Luo[1], Asimina Zoitou[1], Kyungmin (Esther) Kwon[1], Richa Singh[1], JiWon Woo[3], Nikita Sivakumar[3], Joseph L. Greenstein PhD[1,3], Casey Overby Taylor PhD[1,3,4], Amir Kheradmand[2], Kemar Earl Green[2]

[1]Department of Biomedical Engineering, The Johns Hopkins University, Baltimore, Maryland, USA; [2]Department of Neurology, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; [3]Institute for Computational Medicine, The Johns Hopkins University, Baltimore, Maryland, USA; [4]Department of Medicine, The Johns Hopkins University, Baltimore, Maryland, USA;

Introduction

- Ocular Torsion: rotation of the eye around optical axis
  - Dynamic: rotational eye movement during the tilt
  - Static: steady gaze maintained at the end of head tilt
- Degrees of dynamic torsion can be used to detect loss of otolith-ocular function and other disorder diagnosis (central/peripheral)
- Traditional methods involving iris tracking is unreliable:
  - Hard to obtain good waveforms due to frequent eye closures and blinks
  - Videos taken with mobile devices have low quality

Innovation & Significance

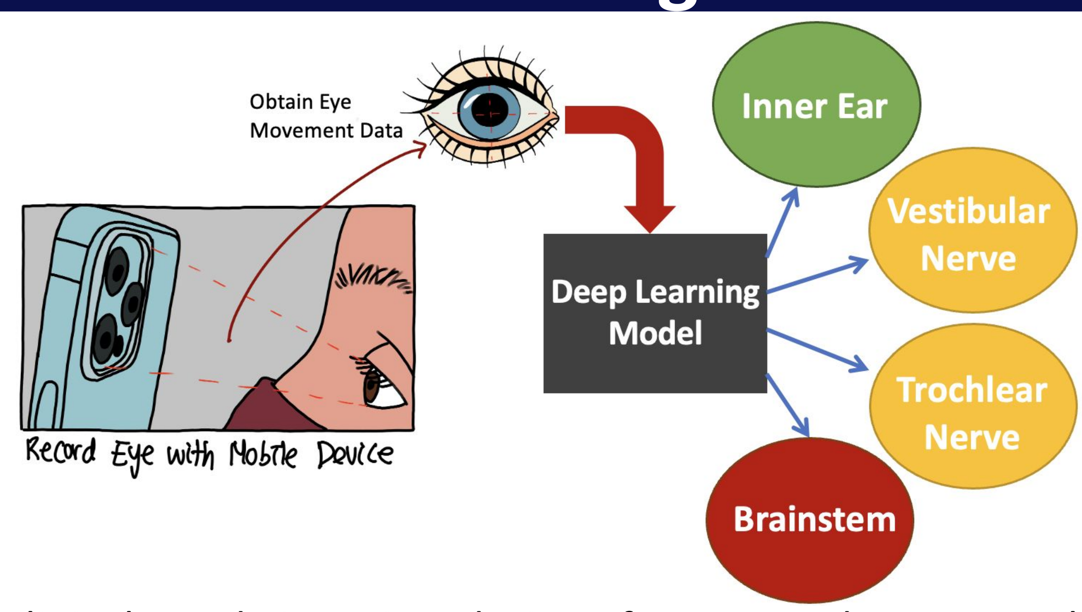


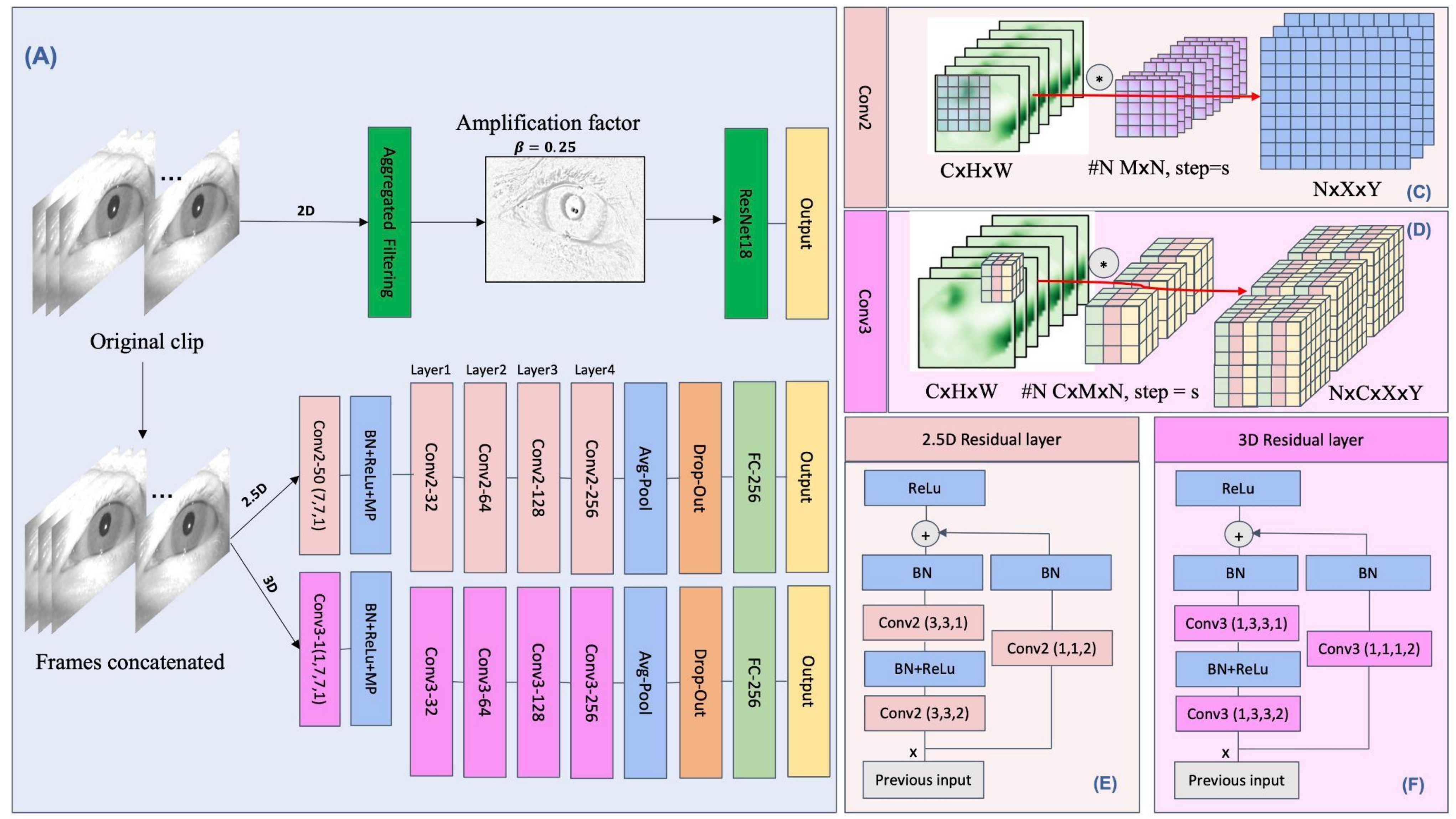
Fig 1. Clinical Need: Overview design of automated, remote solution for triage and diagnosis of dizzy or double vision patients

- **7.4%** lifetime adult prevalence of vestibular vertigo (inner ear disease)
- **74-81%** of peripheral vestibular disorders are often misdiagnosed or incorrectly managed
- Shortage in specialists that can accurately assess/interpret eye movements
- Automated diagnosis in remote settings allows patients to promptly receive necessary treatments

Methods

- **Video Ocular Counter Roll (vOCR) dataset:** 60 videos from 15 health controls with 12 head/trunk movements per video (frame rate: 100 Hz; resolution: 260 × 400 pixels)
- **Data augmentation and balancing:** over-down-sampling and precise labeling algorithm
- Subject-wise train/val/test split representatively (11:2:2)
- Models designed and constructed task-specifically → optimized using grid search
- **Model Interpretability:** 2D/3D GradCAM method
- **Robustness and generalization:** 5-fold cross-validation

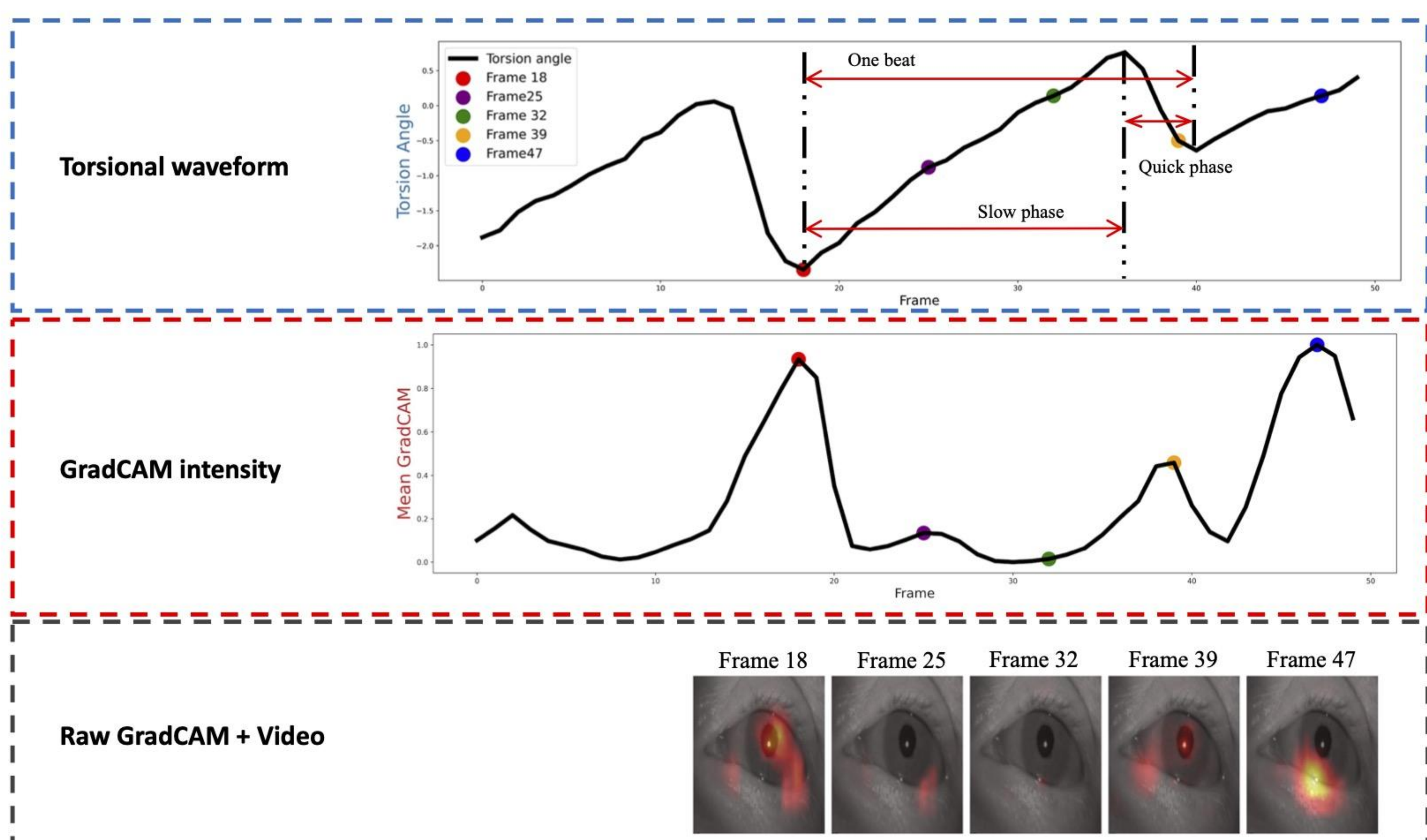
Overview of Model



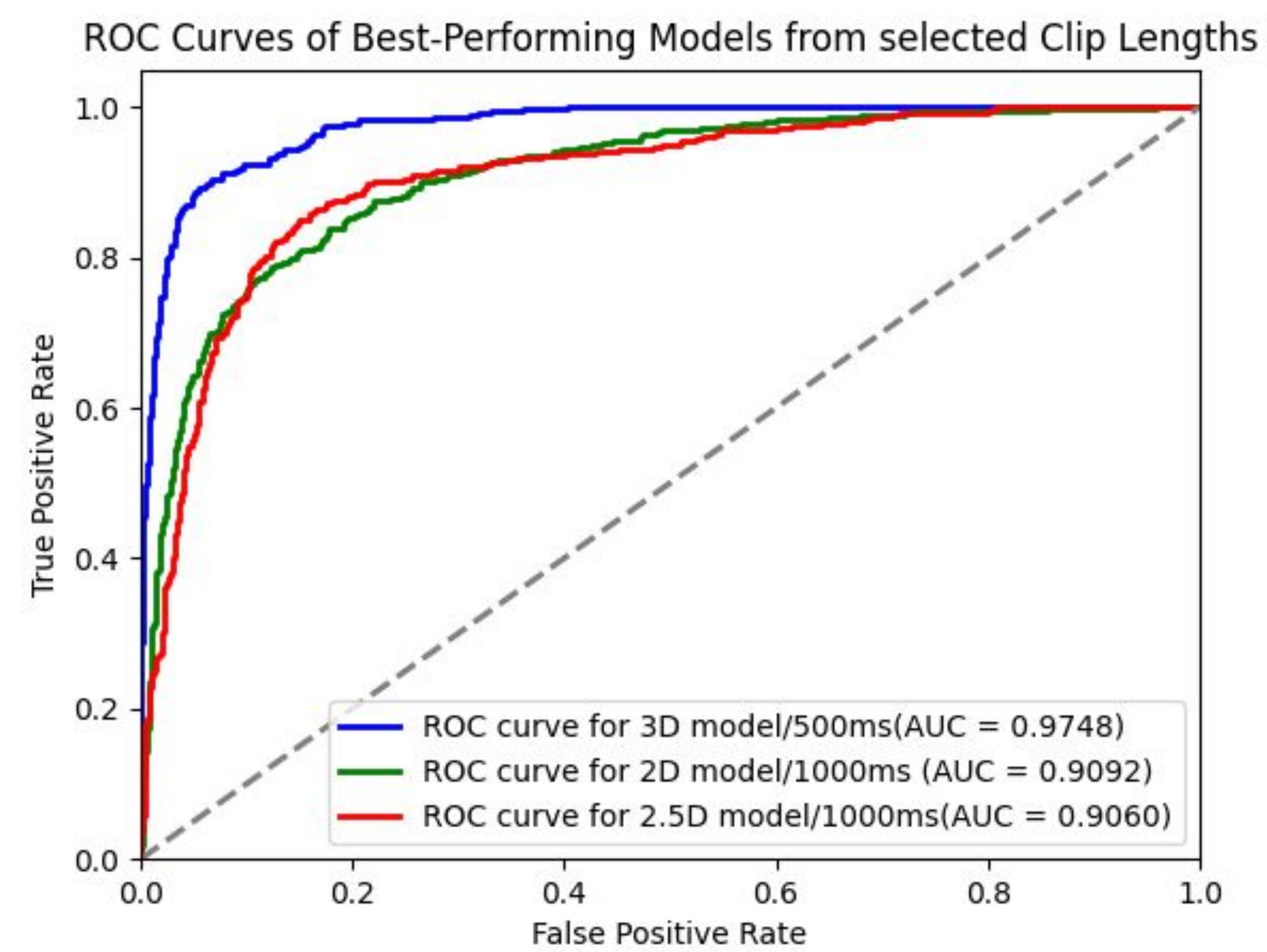
Results

Clip Length / Model	subject-wise train/val/test split			5-fold validation		
	500 ms	750 ms	1000 ms	500 ms	750 ms	1000 ms
2D	0.7987 / 0.9011	0.7987 / 0.9080	0.8275 / 0.9092	0.8807 / 0.9510	0.8508 / 0.9377	0.8688 / 0.9478
2.5D	0.8063 / 0.8738	0.8081 / 0.8685	0.8406 / 0.9060	0.9018 / 0.9754	0.9273 / 0.9782	0.9303 / 0.9797
3D	**0.9119 / 0.9748**	0.8825 / 0.9554	0.8981 / 0.9660	**0.9735** / 0.9964	0.9728 / **0.9971**	0.9694 / 0.9945

Table 1: ACC / AUC for subject-wise train/val/test split and average of 5-fold cross-validation



Results



Conclusions

- Task-specific dataset and precise labeling completed
- Model comparisons and optimization completed
- Decent accuracies were achieved
- Time-series Interpretability model indicates that ocular features and neurophysiology-supported phenomena led to the model's prediction

Future Direction

- Statistical analysis
- External validation & Generalization ability improvement
- Further interpretability and clinical relevance
- Torsion degree regression

References

[1] Y. Agrawal, B. K. Ward, and L. B. Minor. Vestibular dysfunction: prevalence, impact and need for targeted treatment. Journal of Vestibular Research, 23(3):113–117, 2013.
[2] J. Otero-Millan et al. The video ocular counter-roll (vocr): a clinical test to detect loss of otolith-ocular function. Acta Oto-Laryngologica, 137(6):593–597, 2017.
[3] Y. Yang et al. Video ocular counter roll: A bedside test of otolith-ocular function. Annals of Clinical and Translational Neurology, 2023.
[4] K. He et al. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 770–778, 2016

Scan to view GradCAM video: