Learning Time in Static Classifiers

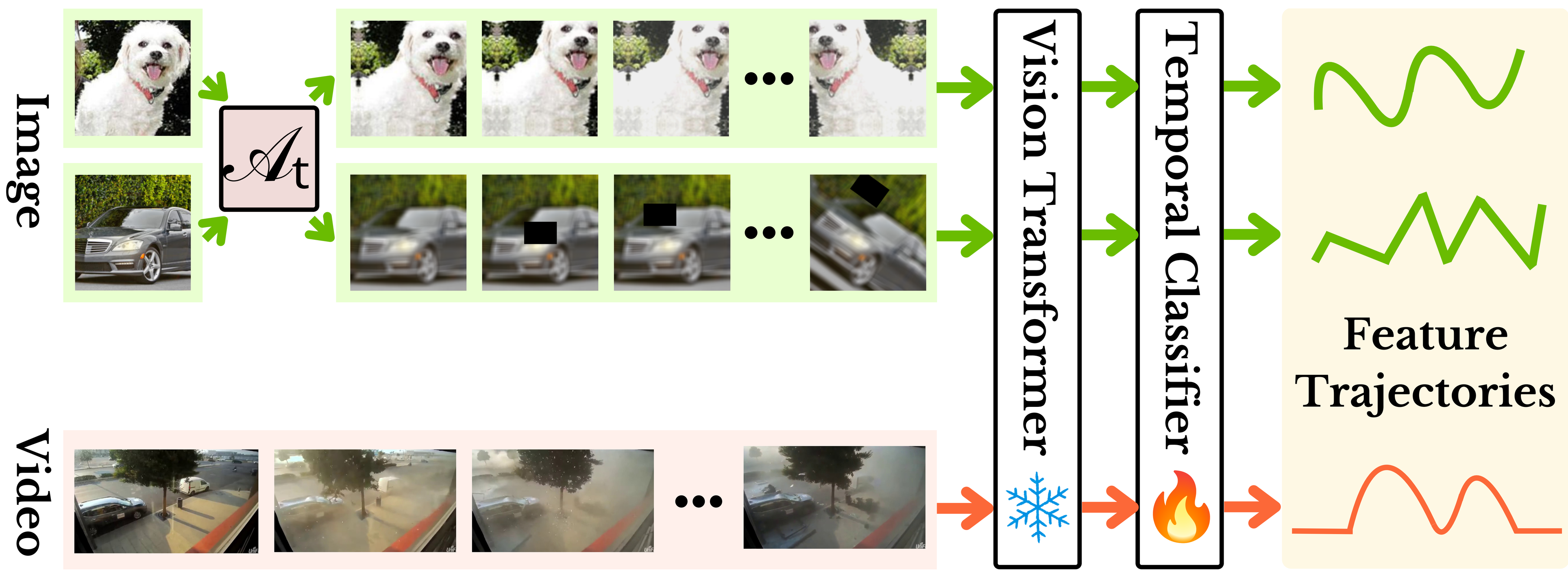
Xi Ding* Lei Wang*
Piotr Koniusz Yongsheng Gao
*Co-first authors (equal contribution).
Griffith University, Queensland, Australia · DATA 61 · CSIRO · UNSW Sydney · Australian National University · AAAI 2026

Motivation and key ideas

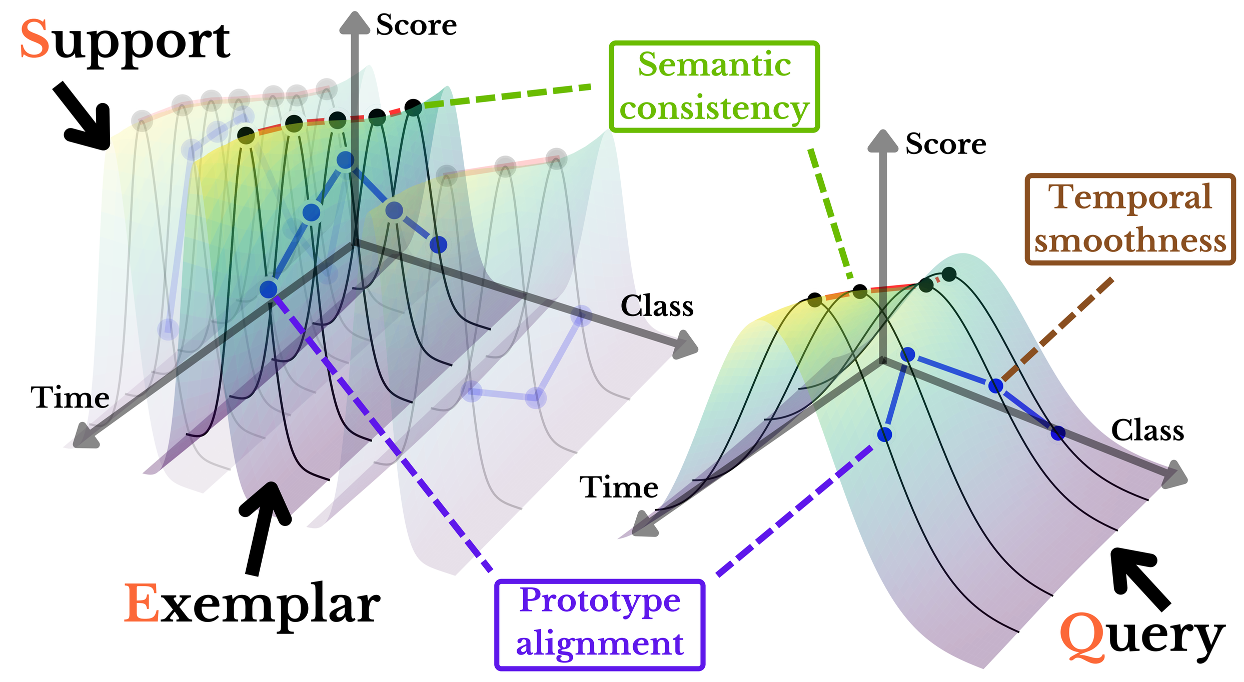
- Real-world visual data rarely presents as isolated, static instances. Instead, it often evolves gradually over time through variations in pose, lighting, object state, or scene context. However, conventional classifiers are typically trained under the assumption of temporal independence, limiting their ability to capture such *dynamics*.
- We ask: *Can standard feedforward classifiers reason over time without modifying their architecture, simply through rethinking how we supervise them?*.

The pipeline: further details

Overview of our framework. (a) Temporally smooth sequences are generated via time-indexed transformations $\mathcal{A}_t$ (or sourced from natural videos) and processed by a frozen, image-pretrained vision transformer to extract frame-wise features. A lightweight temporal classifier is then trained to produce feature trajectories. (b) These trajectories are optimized using a multi-term objective with the Support-Exemplar-Query (SEQ) learning framework to (i) align with class-specific prototype trajectories that capture typical temporal patterns (violet block), (ii) achieve accurate classification through semantic supervision (vivid green block), and (iii) ensure smooth and consistent temporal evolution (gray brown block).

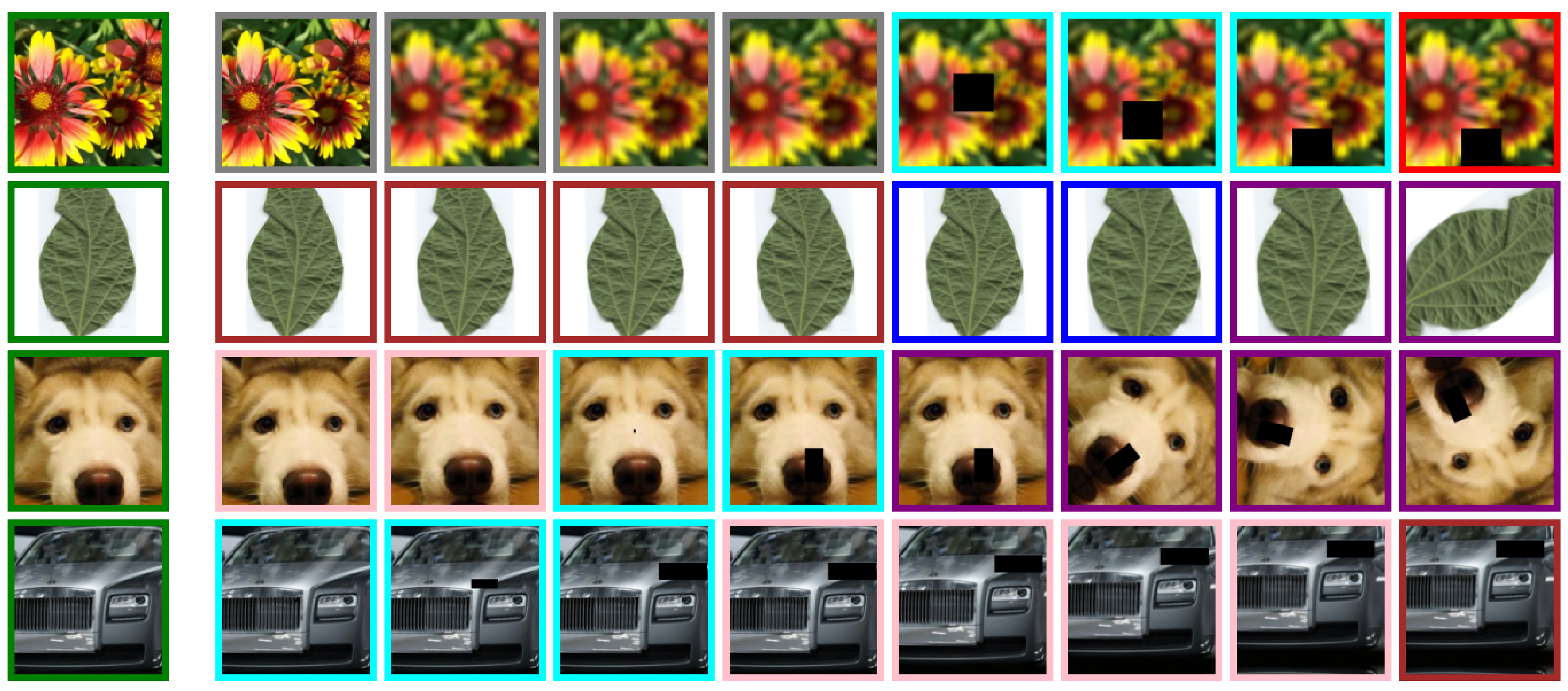


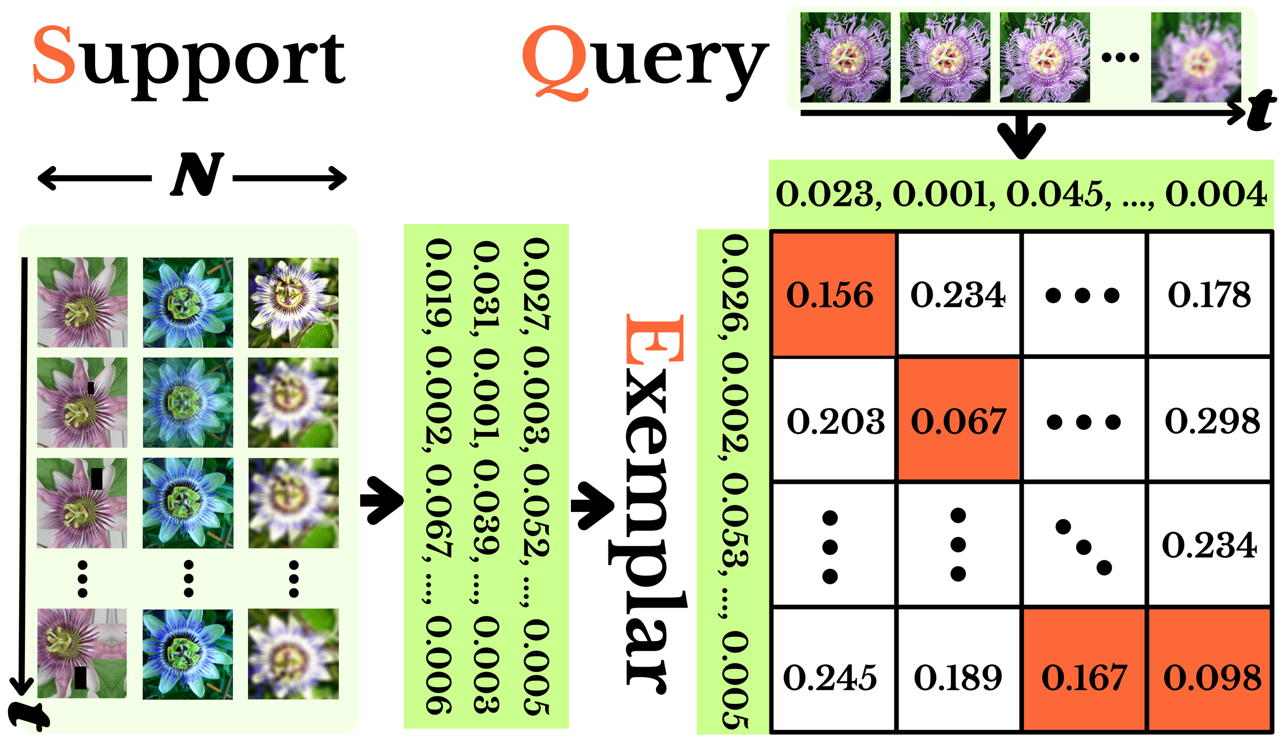
(a) Capturing evolving feature trajectories.



(b) The unified multi-term objective.

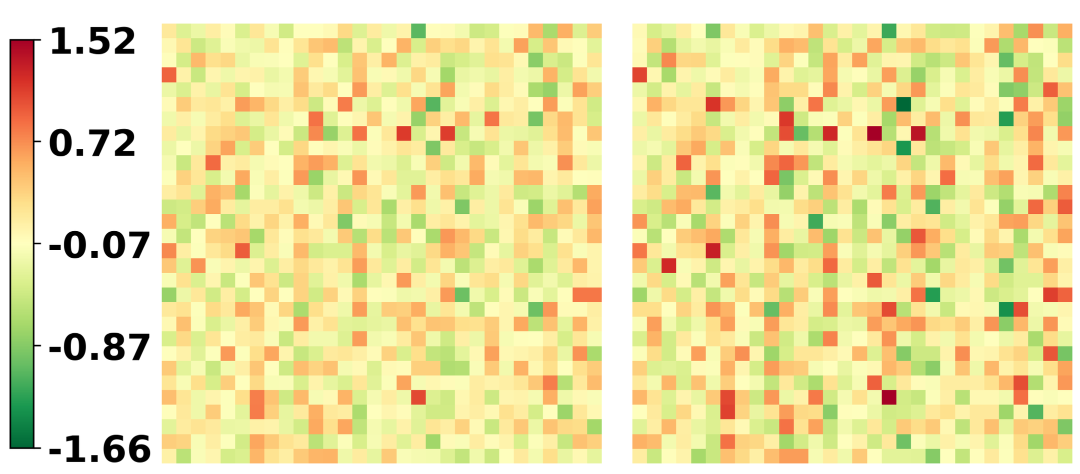
(**Left**) Examples from Flowers-102, SoyAging, Stanford Dogs, and Cars show how augmentations create temporal variations from one image. The first column shows originals (green); others apply augmentations by color: flip (red), zoom (blue), rotation (purple), color jitter (orange), shear (brown), translation (pink), blur (gray), and cutout (cyan), enriching the feature space with varied appearances.
(**Right**) SEQ models class-consistent temporal dynamics by constructing a *support set* of sequences to form a class-specific *exemplar* that captures typical prediction trajectories over time. A *query sequence* is then aligned against this exemplar to enforce temporal consistency and reveal deviations from expected class behavior.



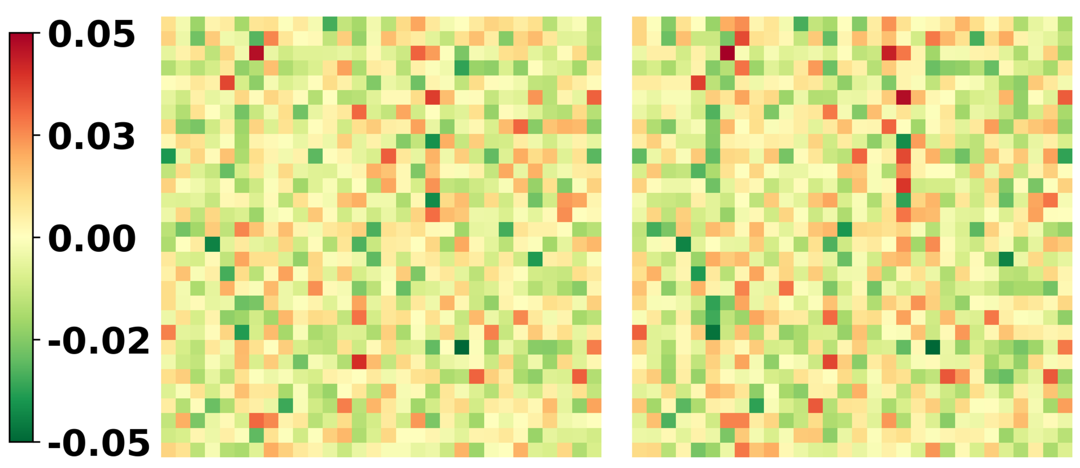


Results

Visualization of selected FC weight regions shows a clear comparison between the baseline (left) and our temporal modeling (right). Temporal modeling yields stronger, more distinct patterns, enhancing feature discrimination.

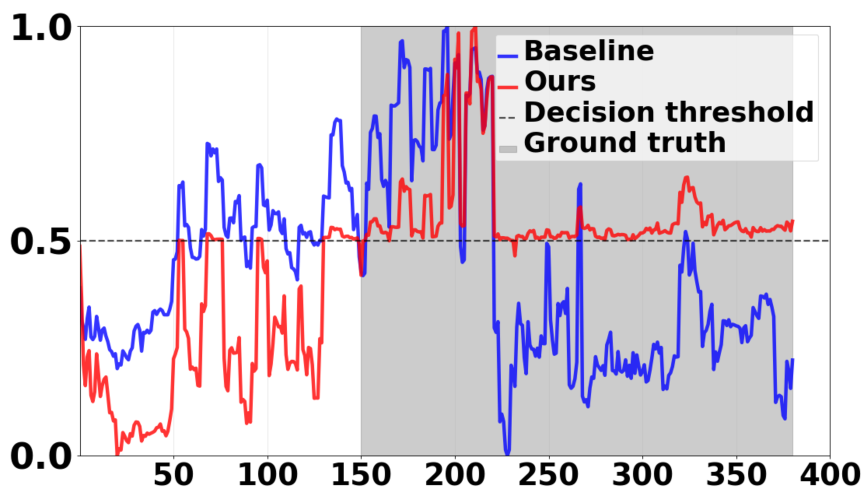


Flowers-102

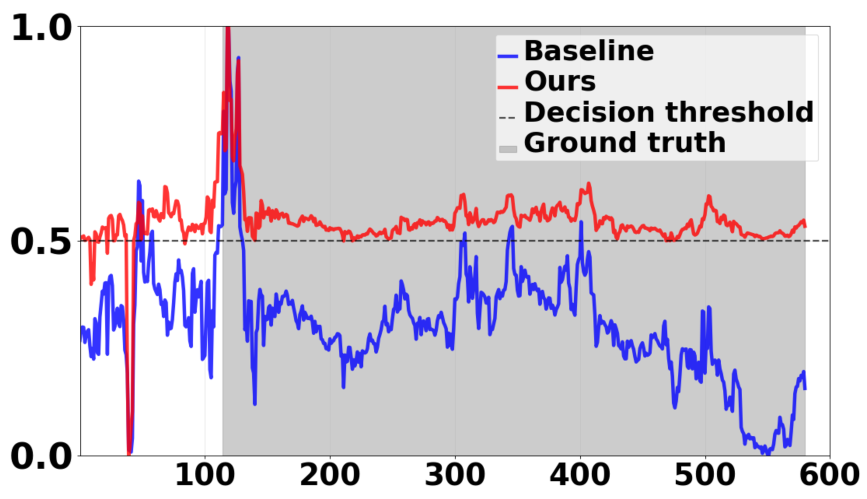


SoyAging

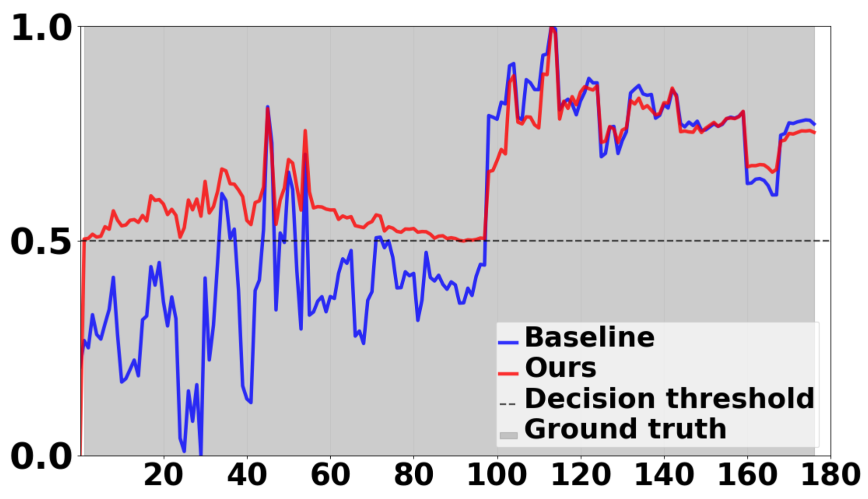
Anomaly prediction comparison. Grey regions indicate ground-truth anomalies. Blue and red curves show the baseline and our method. Our approach detects anomalies more accurately and earlier, with scores crossing the 0.5 threshold in closer alignment with the ground truth.



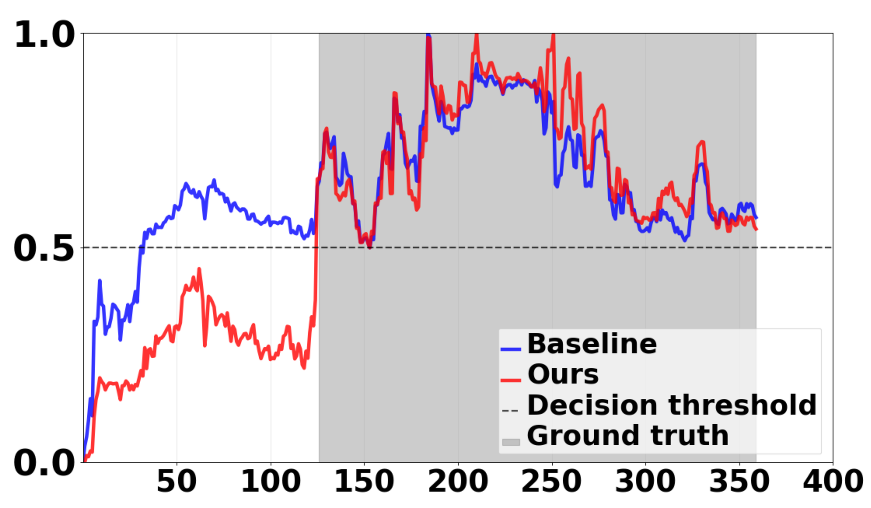
Explosion



People falling



Shooting



Traffic accident