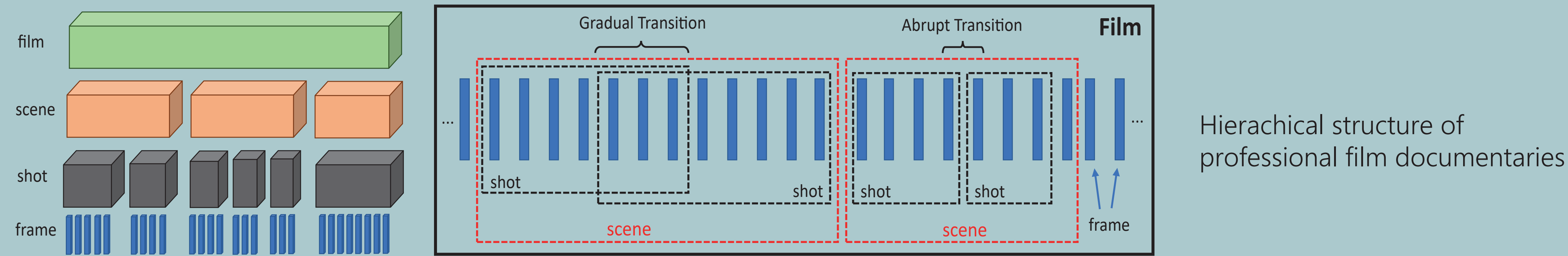


## INTRODUCTION

HISTORIAN, a new dataset, captures novel challenges and characteristics of historical footage.

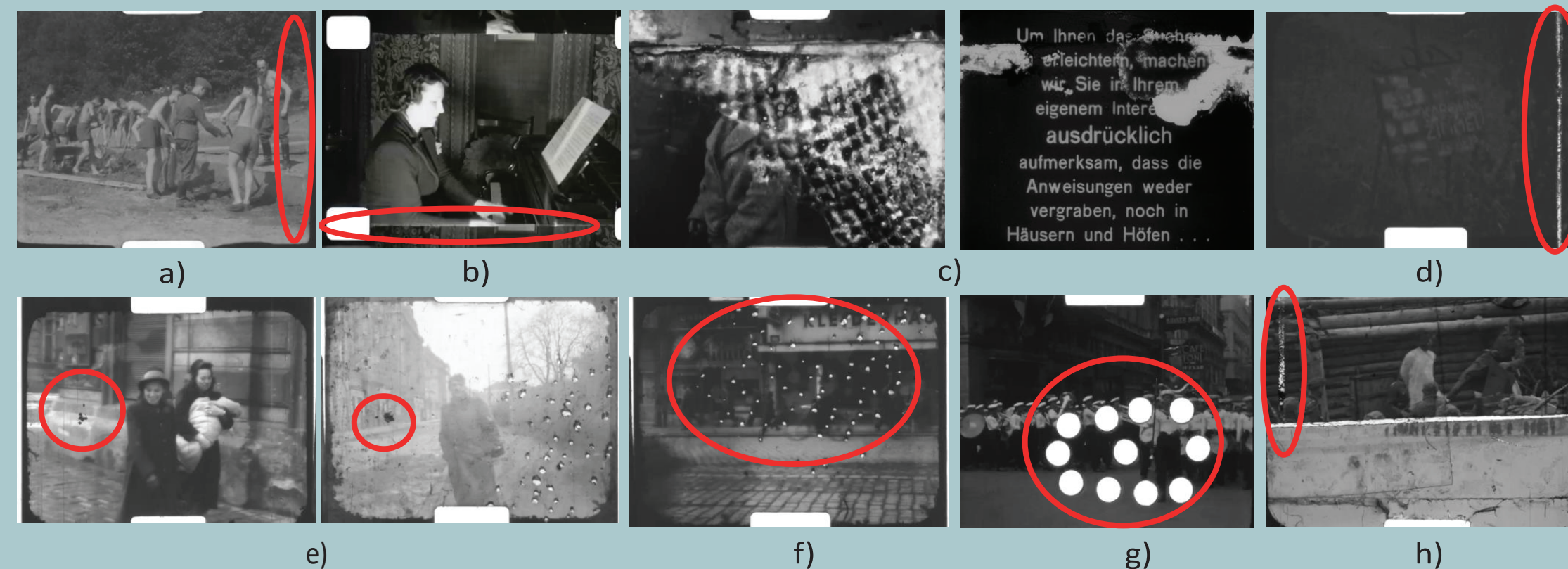
The entire dataset is published on Zenodo to promote research on automatic film analysis.

Quantitative assessment is proposed by evaluating an automatic film analysis pipeline.



Global and local artifacts:

- (a) Horizontal shrinking,
- (b) frame lines, (c) melt,
- (d) vertical scratch (dust),
- (e,f) brights and darks,
- (g) press cuts/cue dots,
- (h) tears



Characteristics and artifacts of digitized analog historical footage

**Table 1:** Comparison of state-of-the-art film-related datasets. (SB) Shot Boundaries, (ST) Shot Types, (CM) Camera Movements, (OA) Overscan Areas

Dataset	color b/w	Cinematographic Settings			
		SB	ST	CM	OA
DeepSBD [1]	color	•			
Clipshot [2]	color	•			
Cinescale [3]	color		•		
MovieNet (Trailers) [4]	color	•	•	•	
HistShotDS [5]	color,b/w		•		
Overscan Areas [6]	color,b/w				•
<b>HISTORIAN (ours)</b>	<b>color,b/w</b>	•	•	•	•

Comparison of state-of-the-art datasets and HISTORIAN

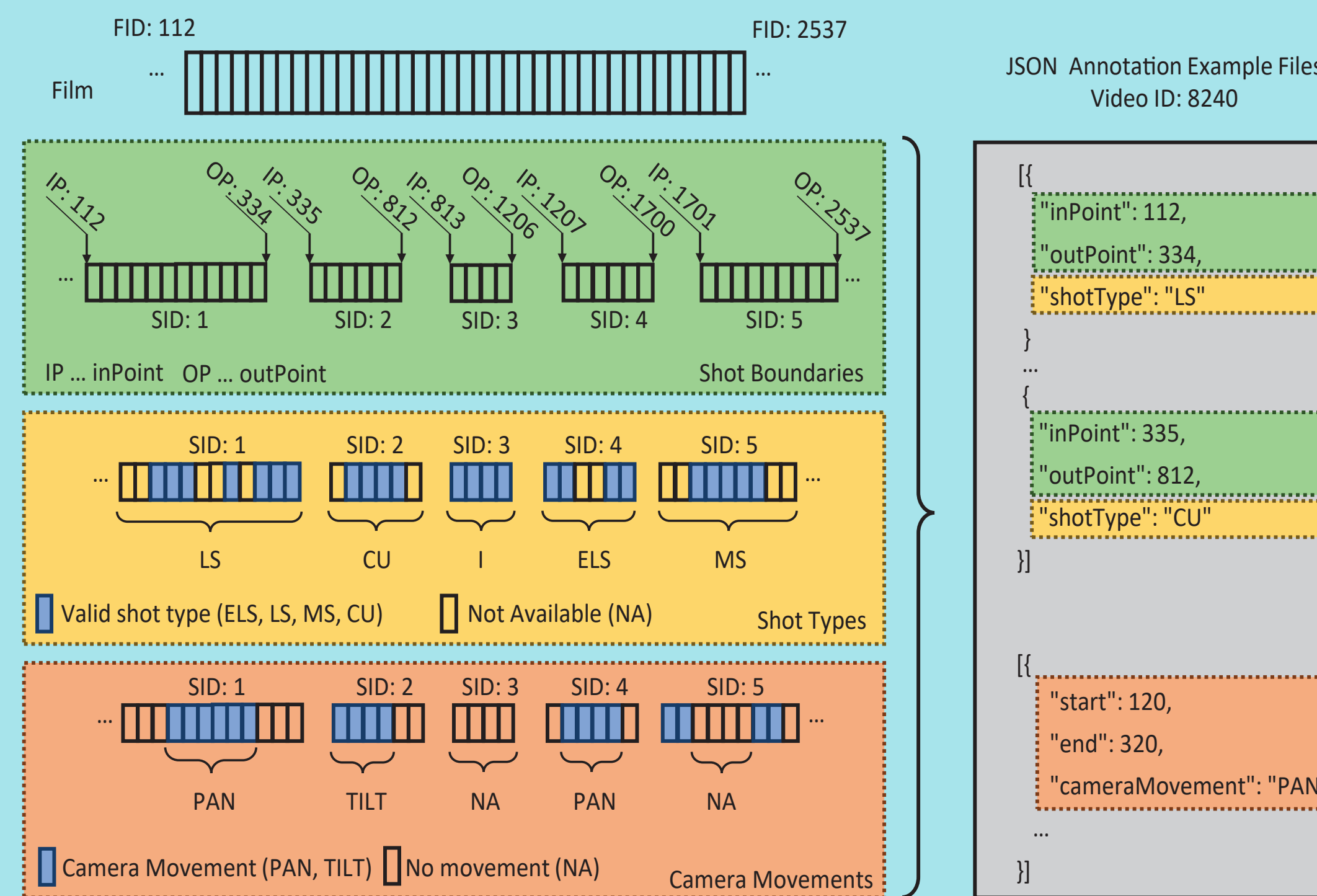
## DATASET

HISTORIAN consists of 98 unprocessed original digitized analog film documentaries recorded during the Second World War (1943-1945).

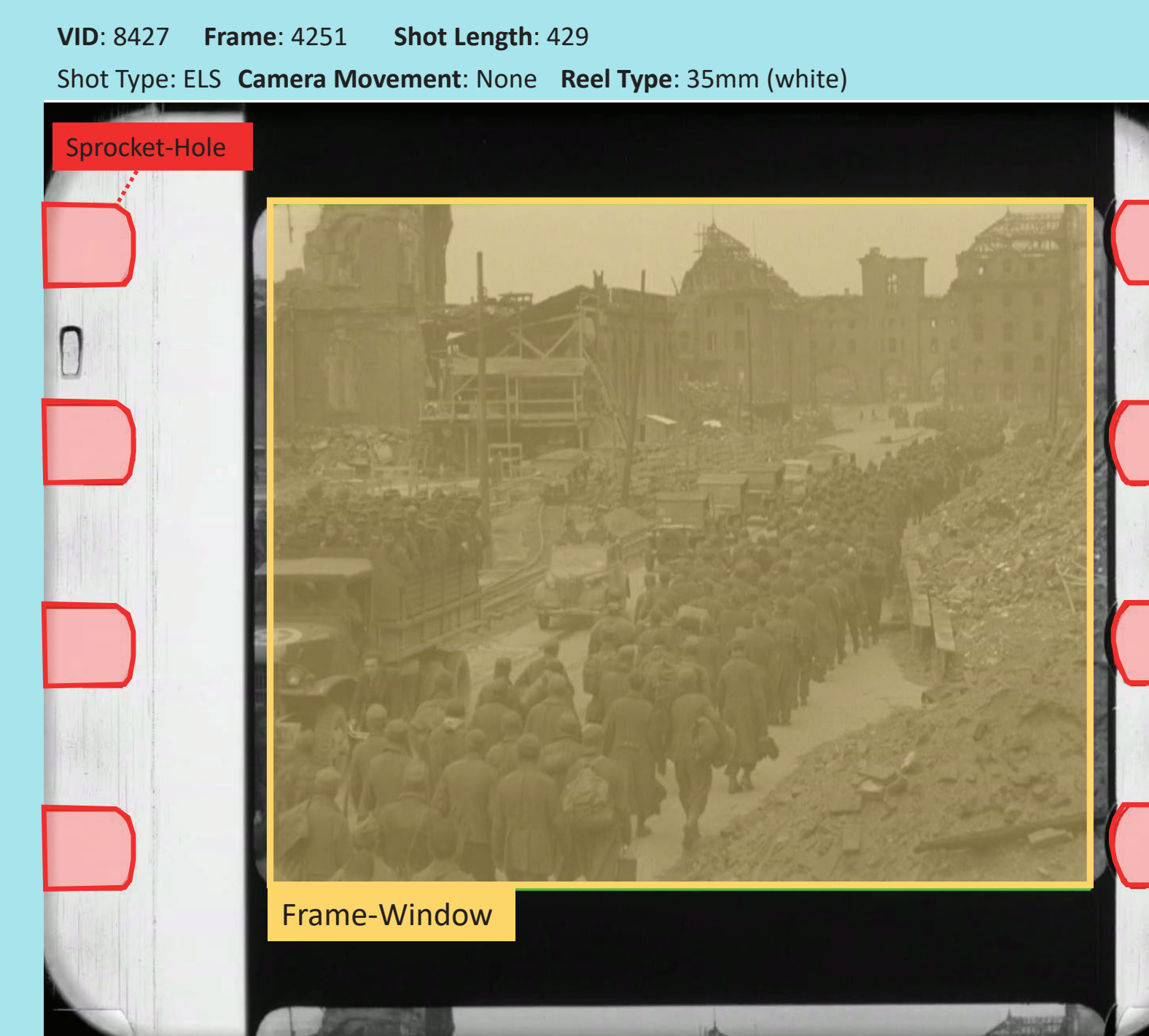
Collaborated annotation process of computer vision and film experts.

- Cinematographic annotations:
- Shot Boundaries
  - Shot Types
  - Camera Movements
  - Overscan regions

Schematic illustration of the annotation process

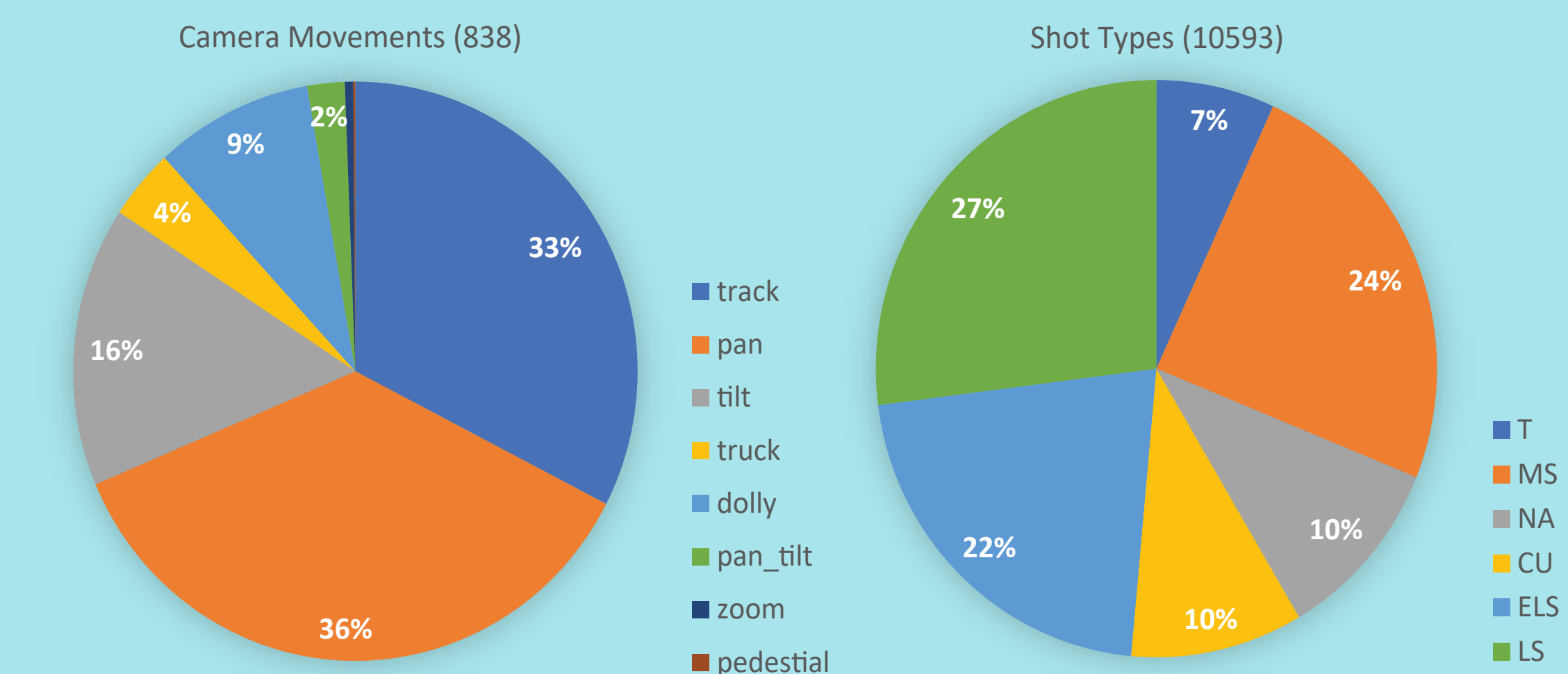


Example of an annotated samples



Dataset Key Facts & Distribution:

- Number of films: 98
- Number of frames ~1.5 million
- Average length: ~11 minutes
- Number of annotated shots: 10593
- Number of annotated sequences: 838
- Number of annotated frames: 192



## EVALUATION

Evaluation of an automatic film analysis pipeline as baseline for further research.

Shot Boundary Detection:

$$F_{1,Score} = 87\%$$

Shot Type Classification:

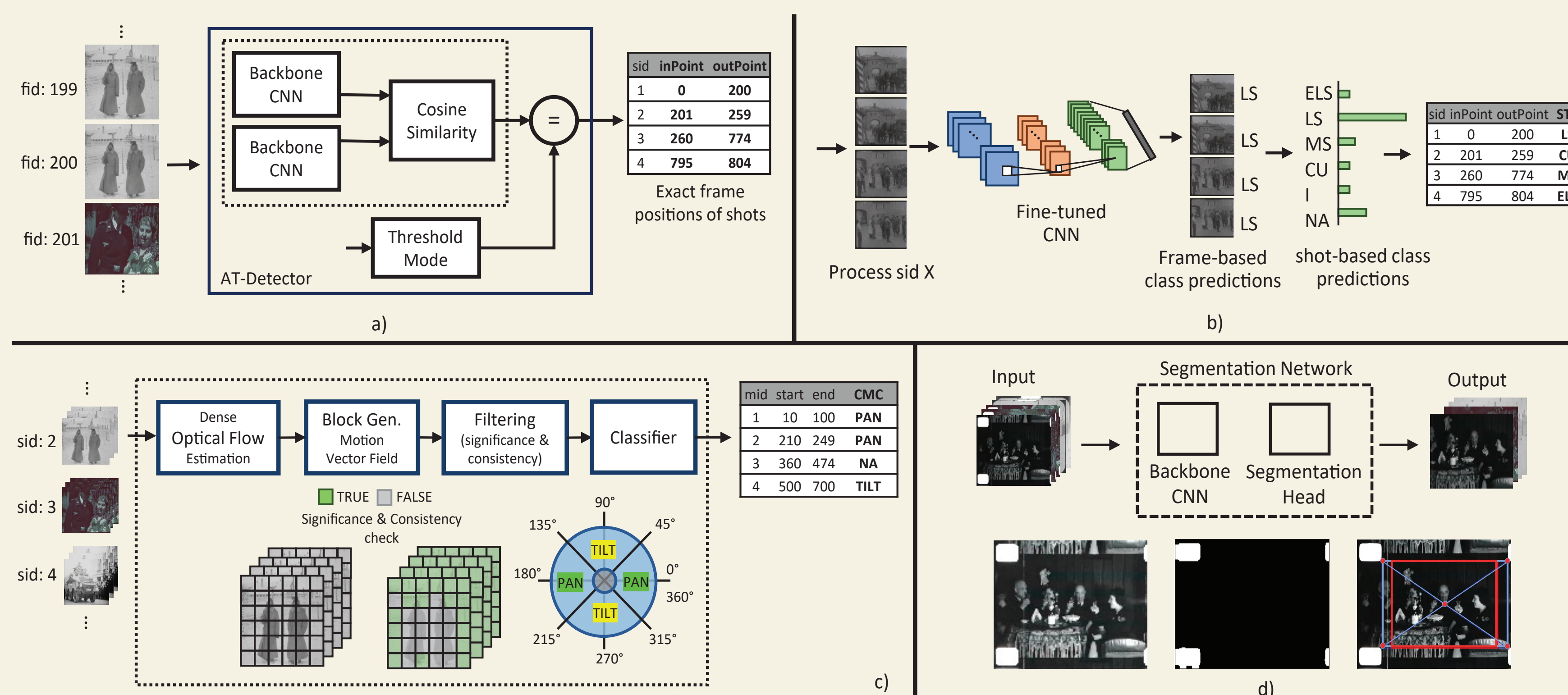
$$Accuracy = 75\%$$

Camera Movements Classification:

$$Accuracy = 82\%$$

Sprocket-Hole (Overscan) segmentation:

$$Dice-Coefficient = 88.1\%$$



Baseline Film Analysis Pipeline:

- a) Shot Boundary Detection
- b) Shot Type Classification
- c) Camera Movements Classification
- d) Overscan Detection

## CONCLUSION

HISTORIAN, a novel dataset, includes cinematographic annotations based on original digitized historical film documentaries.

Captures challenges and characteristics of digitized historical film documentaries



Publicly available on Zenodo