ACTIONS ~ TRANSFORMATIONS

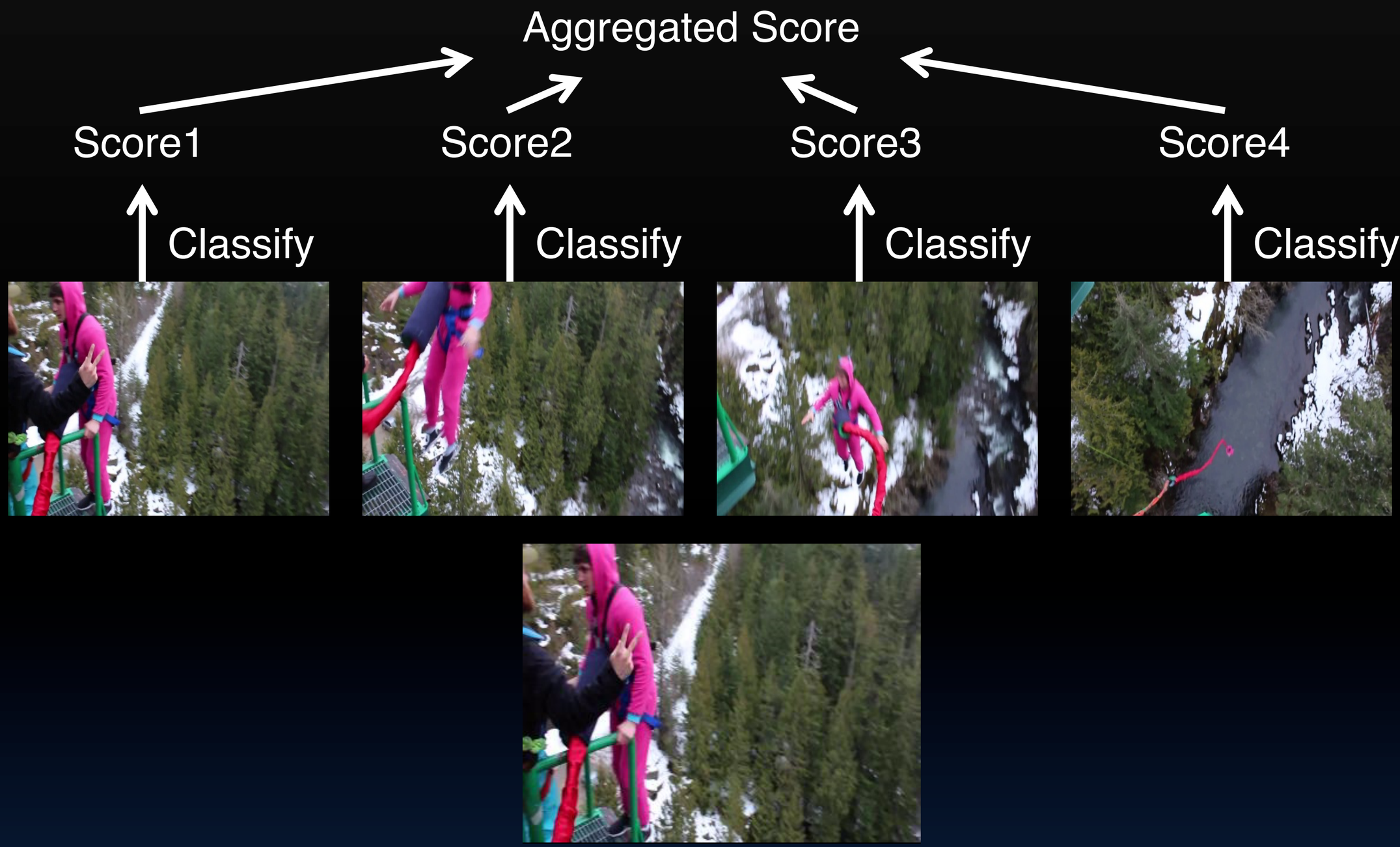
Xiaolong Wang, Ali Farhadi, Abhinav Gupta
Carnegie Mellon University, University of Washington
The Allen Institute for AI

Actions ~ Transformations

- Motivation
- Dataset
- Modeling Actions as Transformations
- Experiments

Motivation

- Action Recognition

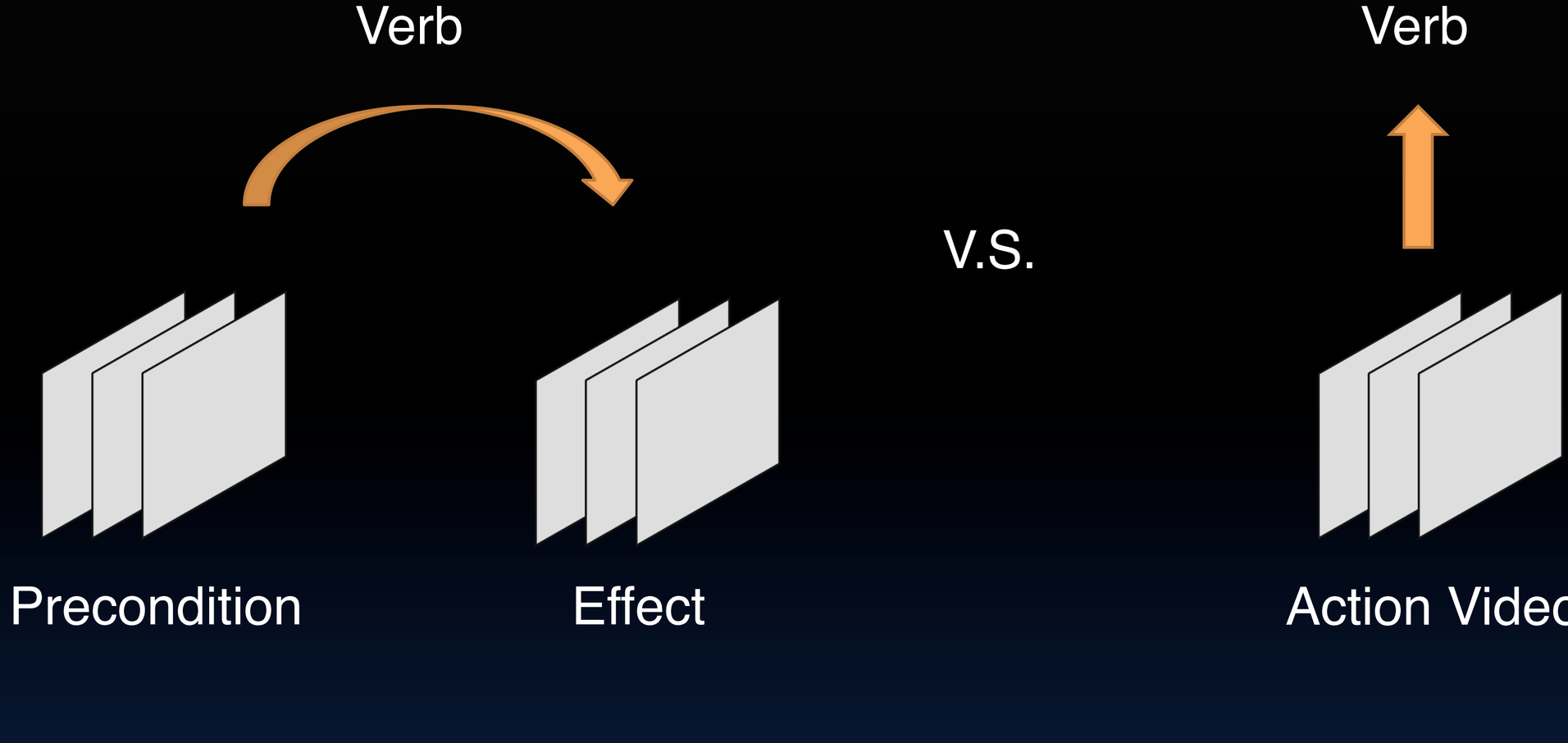


Motivation

Precondition × **Action** = **Effect**

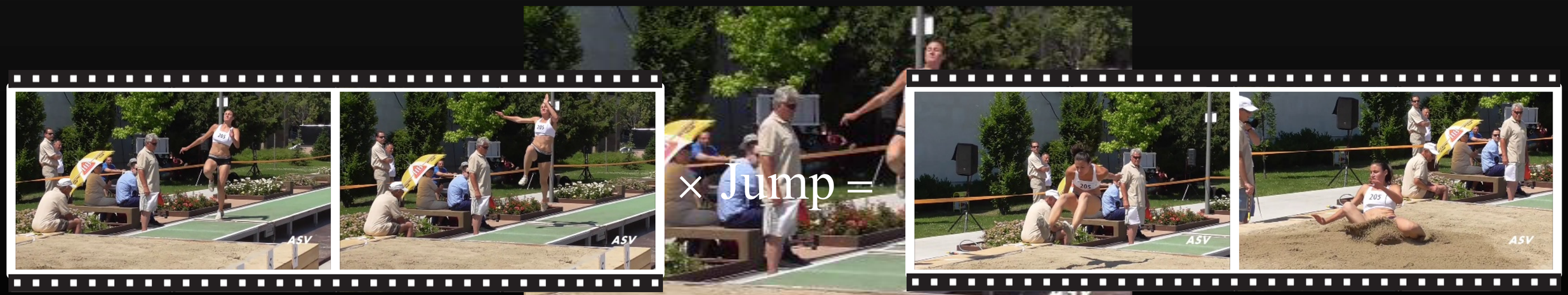
Motivation

- Action: the transformation between precondition and effect.
  - Baseline: Direct classification on action

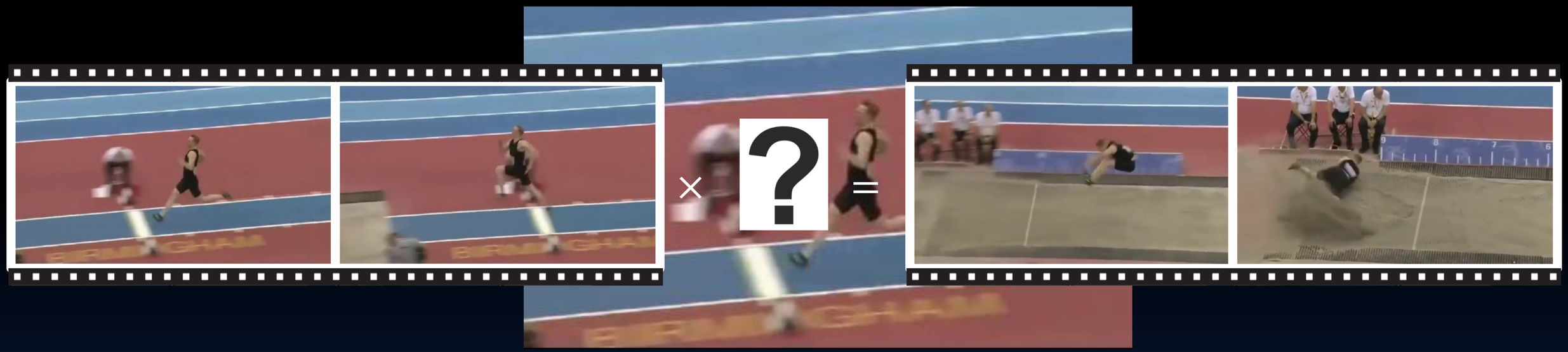


Action Recognition

Training:

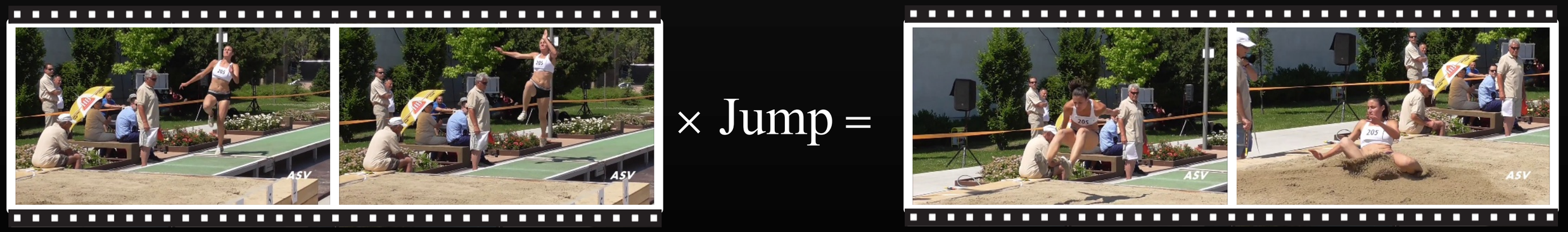


Testing:

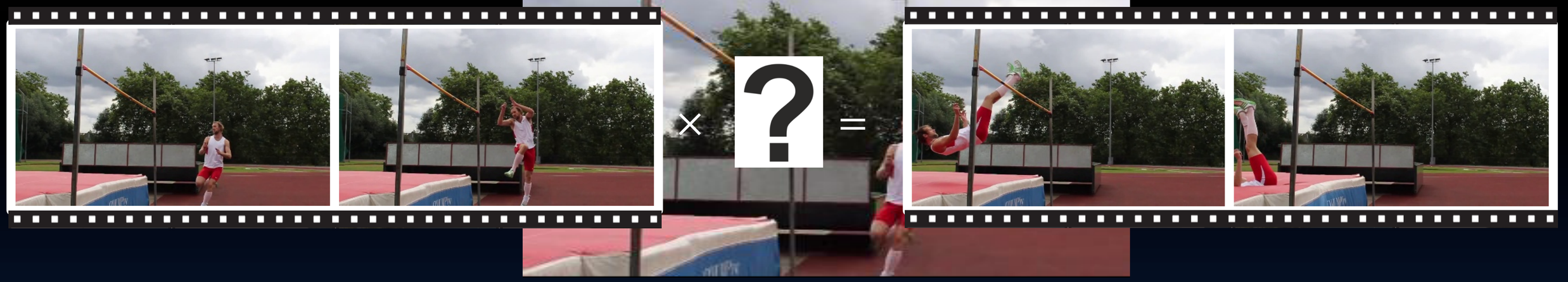


CROSS CATEGORY GENERALIZATION

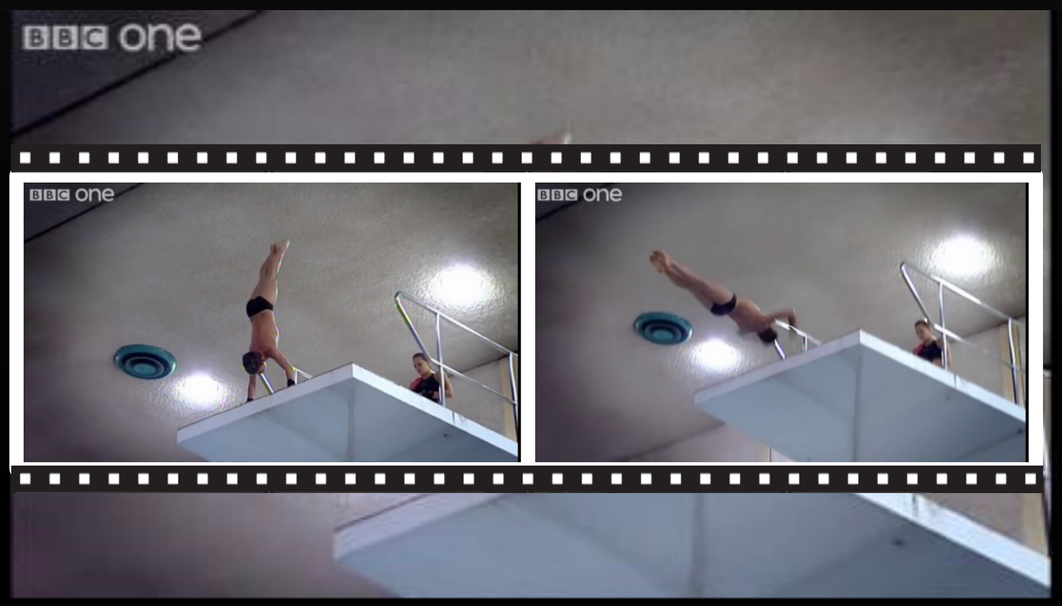
Training:



Testing:



PREDICTION



× Dive =

Actions ≈ Transformations

- Motivation
- Dataset
- Modeling Actions as Transformations
- Experiments

Dataset

- We propose a new action dataset ACT, for testing cross-category generalization.

- 43 action classes, and 16 super classes.

- Youtube videos labeled by turkers: Around 11000 videos with 1-10 seconds.

DATASET

- 2-layer hierarchy

Dataset

- Catching: football, soccer;
- Climbing: rock, rope, tree;
- Crawling: baby, army;
- Cutting: apple, orange, pineapple, watermelon
- Diving: bunjeejumping, cliffdiving, divepool
- Jumping: highjump, longjump
- Kicking-fight: bag, people
- Kicking-ball: soccergoal, soccerpass, soccerpenalty
- Lifting: benchpress, cleanandjerk
- Pouring: beer, juice, milk
- Punching: bag, people, speedbag
- Pushing: car, cart
- Sliding: skateboard, ski, snowboard
- Swinging: baseball, golf, tennis
- Weaponfighting: fencing, swordfighting
- Throwing: discus, shotput, baseball, hammer

Dataset

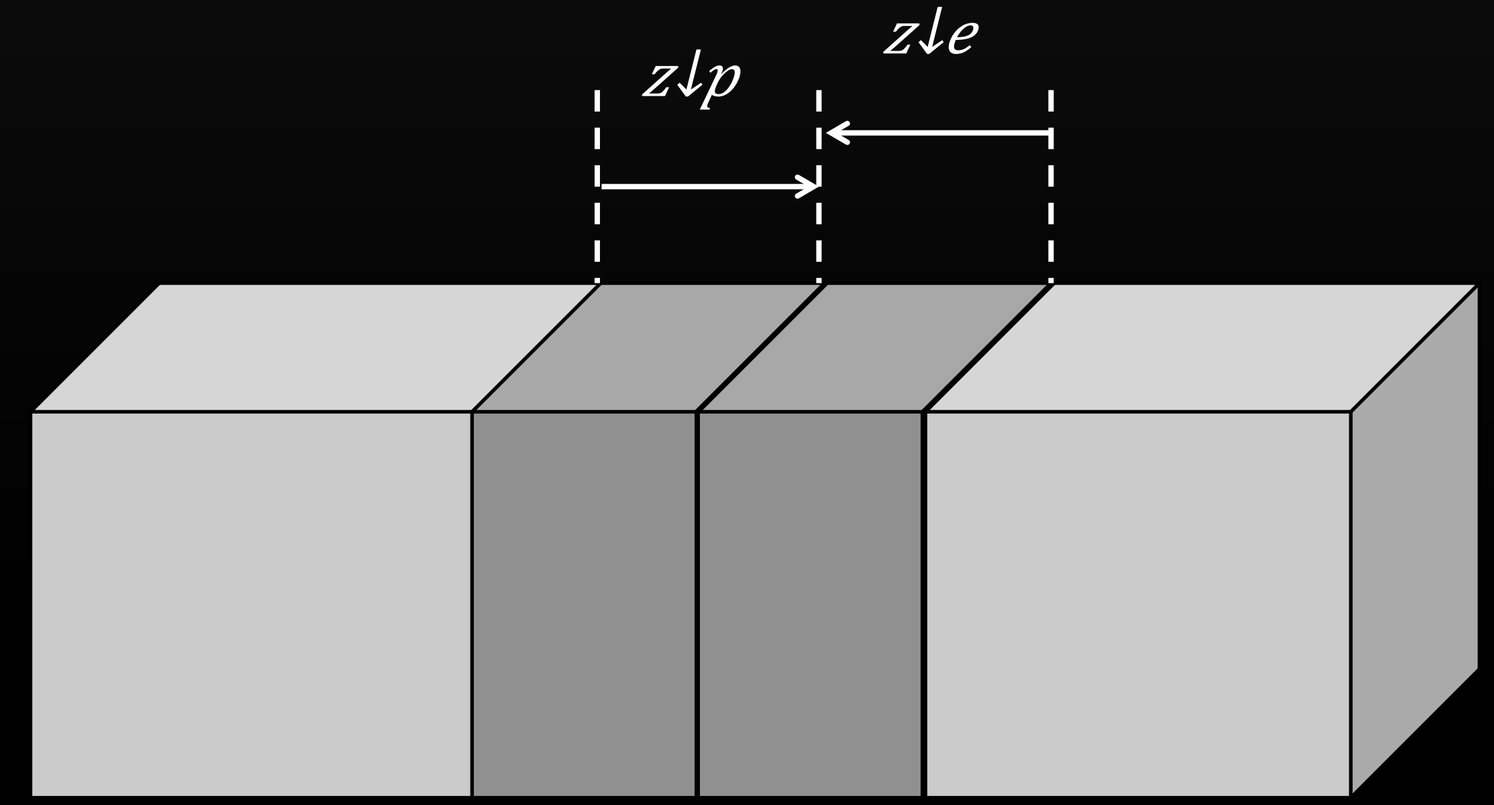
- Task Design
  - Standard Action Recognition (1 split).
  - Cross-category: For each super class, leave one sub-category for testing and the others as training data (3 splits).

Actions ≈ Transformations

- Motivation
- Dataset
- Modeling Actions as Transformations
- Experiments

Model

- Input with latent variables



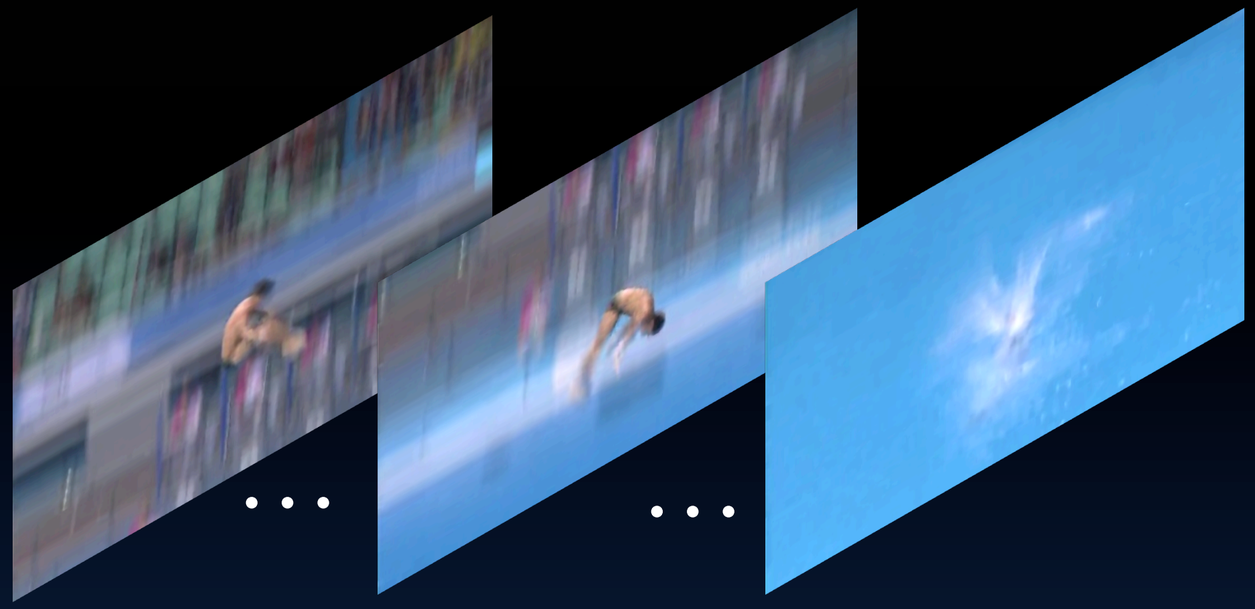
Model

- Input samples

Precond frames

→ Network 1

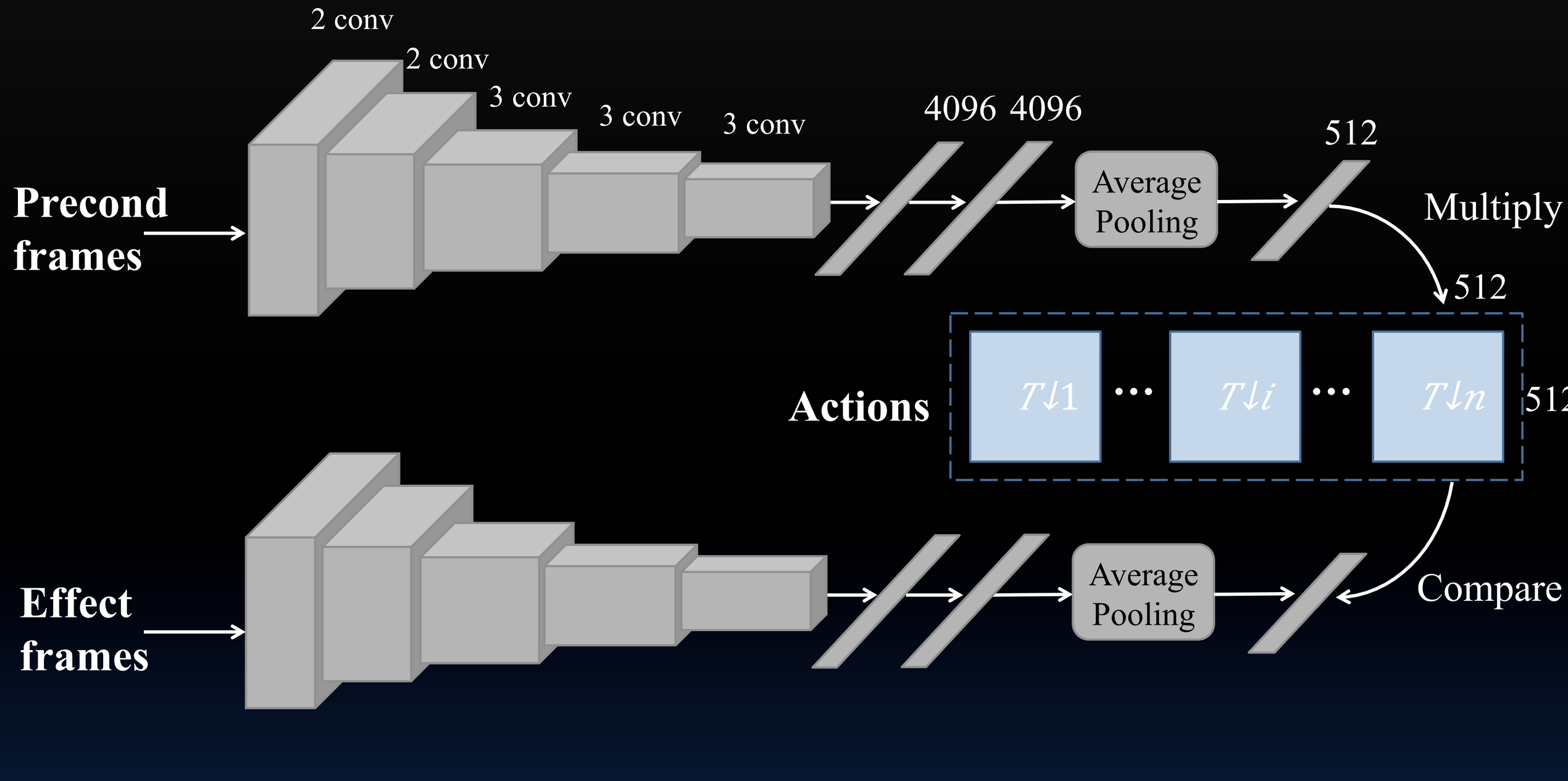
Effect frames



→ Network 2

MODEL

- Siamese Network (VGG 16)



TRAINING

- Learning Objective
  - *minimize* dist(precond_fc×T↓y , effect_fc)
  - *maximize* dist(precond_fc×T↓i , effect_fc) for i ≠ *y*

- Iterate between 2 steps:
  - Learn model parameters
  - Estimate latent variables

TESTING

Precondition

Effect

Testing

Precondition

Effect

Testing

Precondition

Effect

Actions ≈ Transformations

- Motivation
- Dataset
- Modeling Actions as Transformations
- Experiments

Numbers

- UCF101: 13000 videos, 101 classes

Method	RGB	Optical Flow	Fusion
Two Stream [32]	73.0%	83.7%	88.0%
Two Stream (VGG16) [45]	78.4%	87.0%	91.4%
Ours	**80.8%**	**87.8%**	**92.4%**

- HMDB51: 6700 videos, 51 classes

Method	RGB	Optical Flow	Ave Fusion
Two Stream [32]	40.5%	54.6%	58.0%
Two Stream (VGG16)	42.2%	55.0%	58.5%
Ours	**44.1%**	**57.1%**	**62.0%**

Numbers

- Training precondition and effect models separately (UCF)

Method (for split1)	RGB	Optical Flow	Fusion
Two Stream (VGG16) [45]	79.8%	85.7%	90.9%
Two Stream First Half	80.0%	84.7%	90.0%
Two Stream Second Half	79.5%	84.8%	
Ours	**81.9%**	**86.4%**	**92.0%**

- Ensemble with 2-stream (HMDB)
  - 62%→63.4%

Numbers

- ACT dataset, cross category generalization

	RGB				Optical Flow				Fusion			
Model	Split1	Split2	Split3	Average	Split1	Split2	Split3	Average	Split1	Split2	Split3	Average
Two Stream	48.3%	53.2%	49.8%	50.4%	53.5%	56.8%	56.3%	55.5%	59.6%	66.8%	61.8%	62.7%
LSTM+Two Stream	47.7%	53.4%	50.4%	50.5%	55.3%	57.2%	58.2%	56.9%	60.0%	67.3%	61.9%	63.1%
Ours	50.7%	55.2%	52.3%	**52.7%**	57.4%	59.6%	60.3%	**59.1%**	62.5%	70.1%	65.4%	**66.0%**

VISUALIZATION

- Back-prop gradients

Precondition | Direct Classification | Ours | Effect | Direct Classification | Ours

Nearest Neighbor Testing

Query

Ours

Two Stream

Nearest Neighbor Testing

Query

Ours

Two Stream

Nearest Neighbor Testing

Query

Ours

Two Stream

Prediction Results

Precondition Query

Effect Retrieval
(same class)

Effect Retrieval
(different class)

Prediction Results

Precondition Query

Effect Retrieval
(same class)

Effect Retrieval
(different class)

Prediction Results

Precondition Query

Effect Retrieval
(same class)

Effect Retrieval
(different class)