

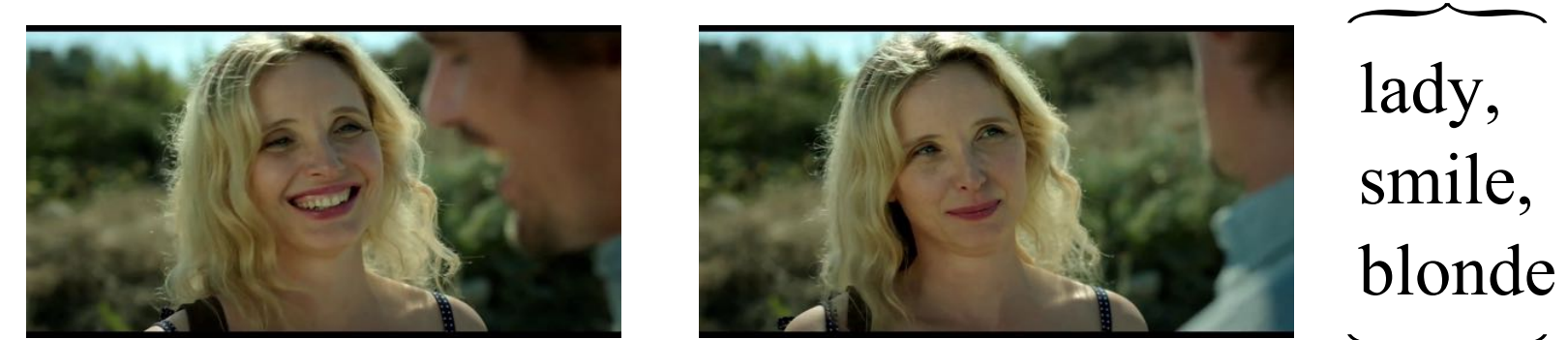
## Sparsified VideoQA

- Accomplish VideoQA with sparse video inputs (frames or words)



Descriptions: {lady, black dress, dress, woman, blue, blue shirt, man, smile, eyes, people walking, .....}

Video Sparsification



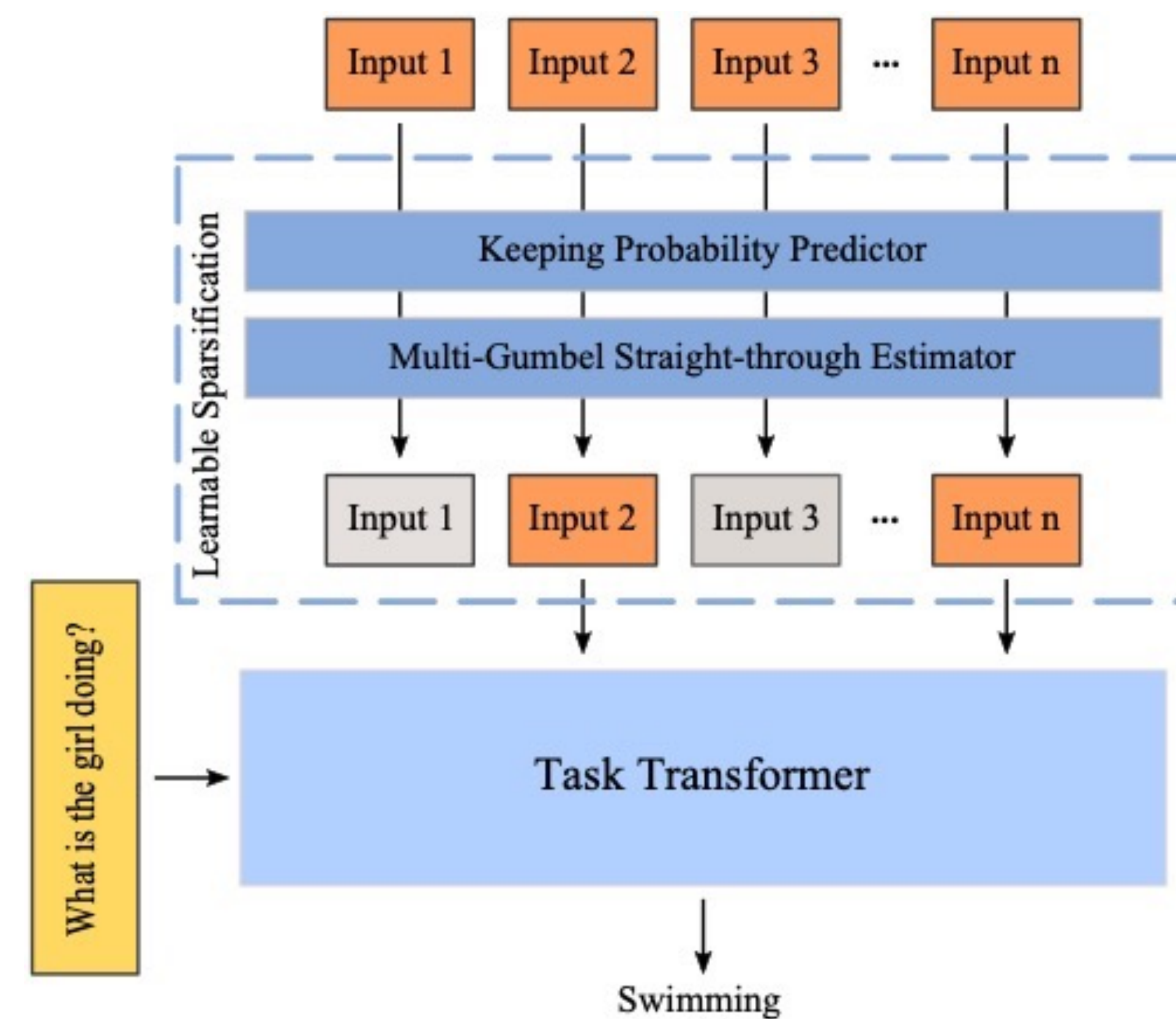
lady, smile, blonde

Q: Is someone laughing?

Yes, the lady is laughing.

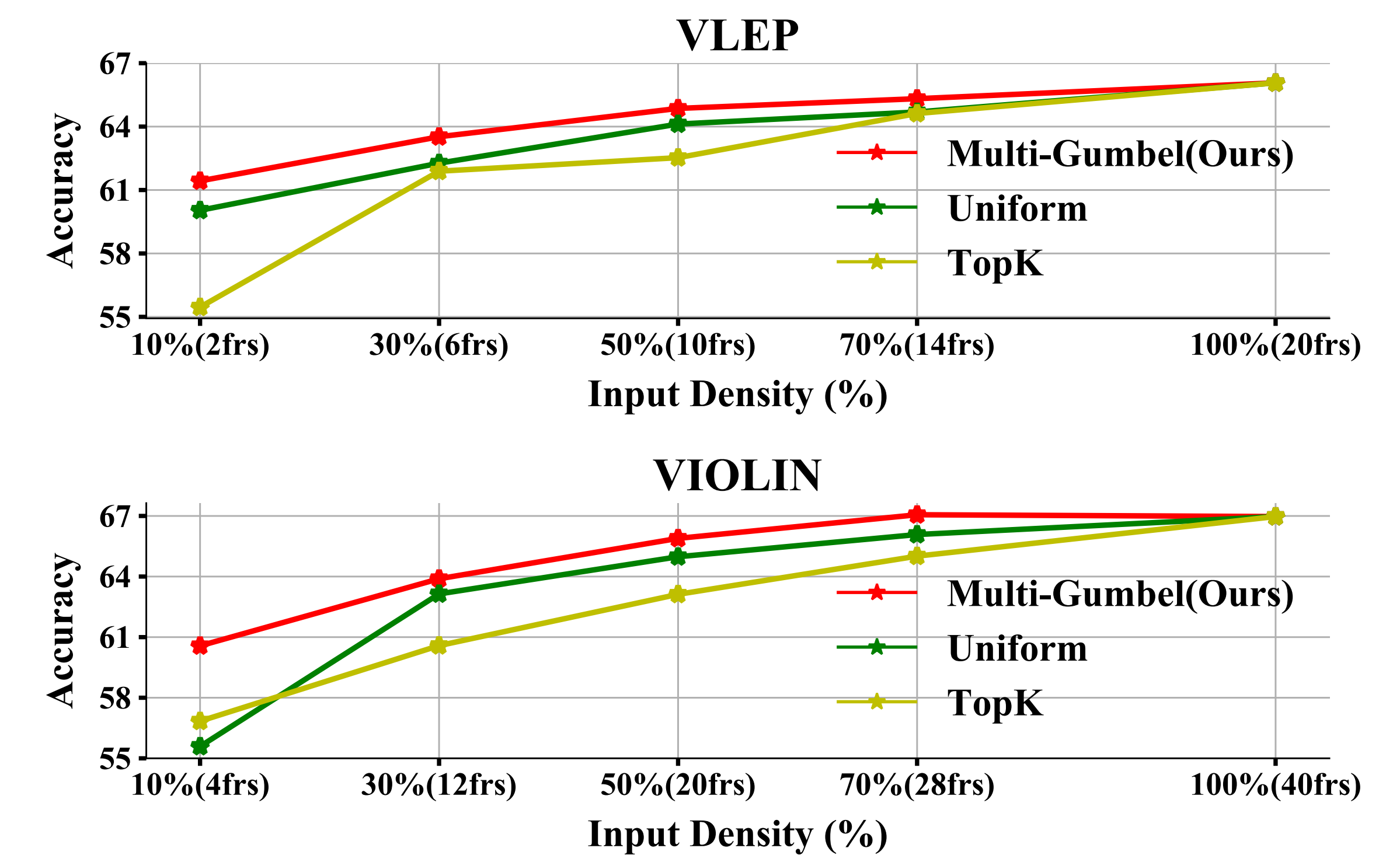
## Learnable Sparsification

- Learn to keep important tokens
- Multi-Gumbel Straight-through Estimator: facilitate sampling process during training



## VideoQA Results

- About 5% performance drop with 10% video lengths



## Multi-modal VideoQA Results

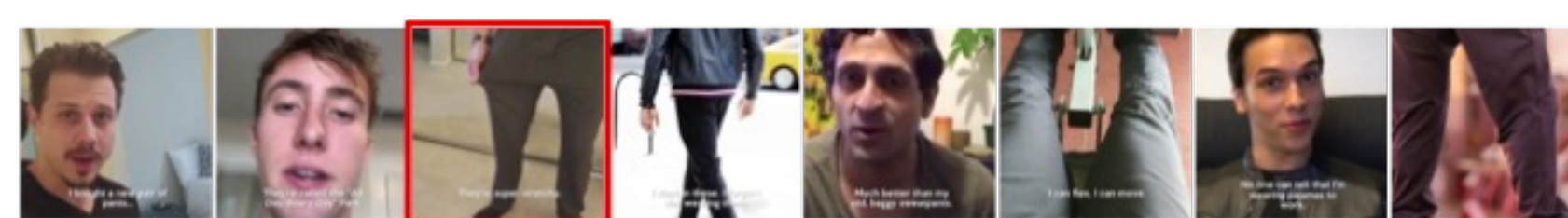
		Visual (Snippets)				
		0	1 snippet	2 snippets	5 snippets	20 snippets
Textual (Words)	0	14.6 (Q-only)	28.65	30.24	31.26	35.43 [22]
	5 words	17.5	28.68	30.31	31.70	35.43
	10 words	18.22	29.87	31.43	31.88	36.01
	25 words	20.14	30.16	31.59	32.03	36.09
	100 words	26.75	31.47	32.11	33.21	36.42

## Ablation on Multi-Gumbel Process

- More explorative selection is beneficial for denser inputs

Input Percentage	VLEP			VIOLIN		
	$\tau = 0.01$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.01$	$\tau = 0.1$	$\tau = 0.5$
10%	60.25	56.01	58.94	56.25	60.57	58.80
30%	60.95	63.52	59.13	61.72	57.64	62.34
50%	63.05	63.64	64.30	65.57	64.48	66.06
70%	63.73	65.14	65.32	65.94	66.52	67.06

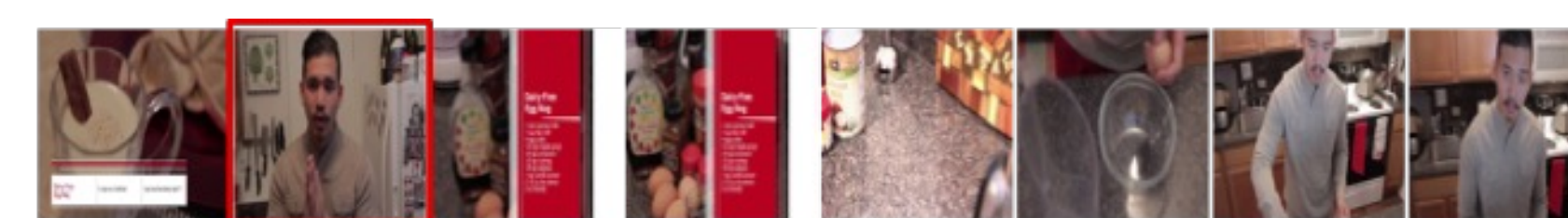
## Sparsified Video Examples



Key Frame

Key Words

pants button, sweatshirt, chef's jacket, jacket, suit jacket, hoodie, hazard vest, blue jeans, athlete, humans



Key Frame

Key Words

espresso, coffee cup, nutmeg, energy drink, milk, ground cinnamon, mustache, ukulele, beard, goat



Key Frame

Key Words

multi meter, charging, unit, air conditioner, gas range, valve, electrical currents, meter, hoses, water faucet, hose



Key Frame

Key Words

buildings, city skyline, mobile homes, homes, houses, construction bids, home remodeling, brick wall, gazebo, furniture

Q: What is the man having in his hand in the first part of the video?

Predicted Answer: Pants

Q: What facial hair does the man have?

Predicted Answer: Mustache

Q: What is the white striped item in the video?

Predicted Answer: Wire (GT: Shirt)

Q: What is the lady holding?

Predicted Answer: Plant (GT: Glass)