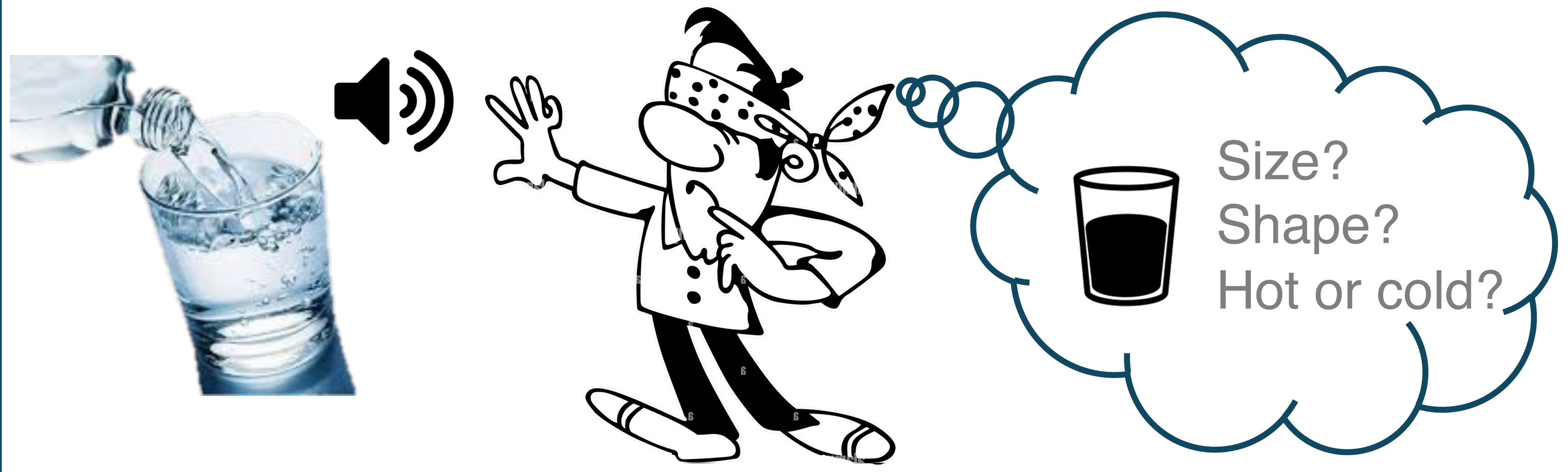
The Sound of Water

Inferring Physical Properties from Pouring Liquids

Piyush Bagad, Makarand Tapaswi, Cees G.M. Snoek, Andrew Zisserman

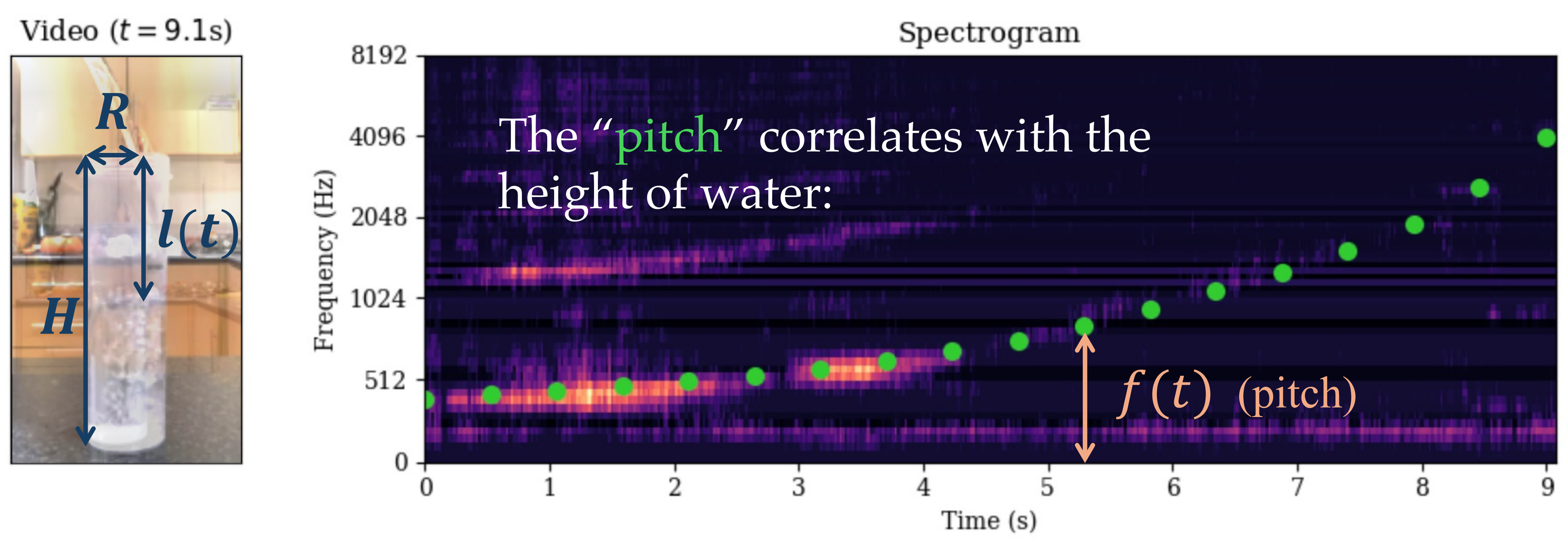
2025 ICASSP HYDERABAD, INDIA
Celebrating Signal Processing
UNIVERSITY OF OXFORD
INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY HYDERABAD

A Remarkable Human Ability



Humans are surprisingly good at estimating physical properties merely from the sound of pouring (Cabe et al., 2000)! Can we train machines to replicate that?

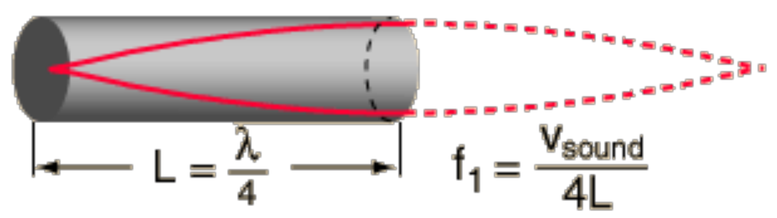
The Physics of Pouring Sounds



Fundamental equation for the sound of pouring

$$\frac{c}{4}\frac{1}{f(t)} = l(t) + \beta R; \quad l(t) = \begin{cases} H, & t = 0 \\ 0, & t = T \end{cases}$$

- c is the speed of sound in air; and β experimental constant
- Underlying principle is the same as that in a resonant pipe

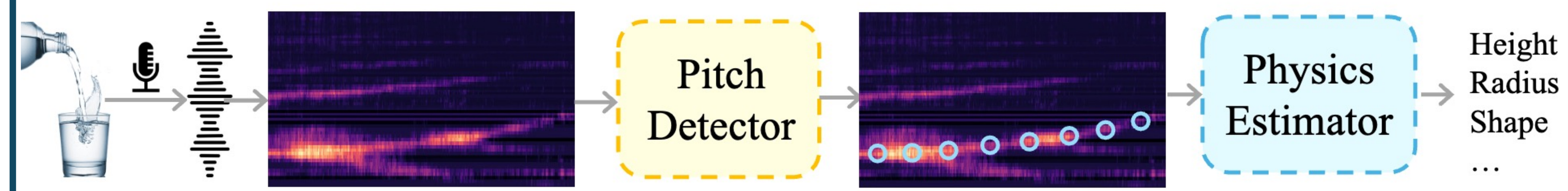


Recovering physical properties from pitch

$$l(t) = \frac{1}{4}\left(\frac{c}{f(t)} - \frac{c}{f(T)}\right); \; H = l(0); \; R = \frac{c}{4\beta}\frac{1}{f(t)}$$

Height H depends on accurate pitch at the start of pouring and radius R on the end of pouring

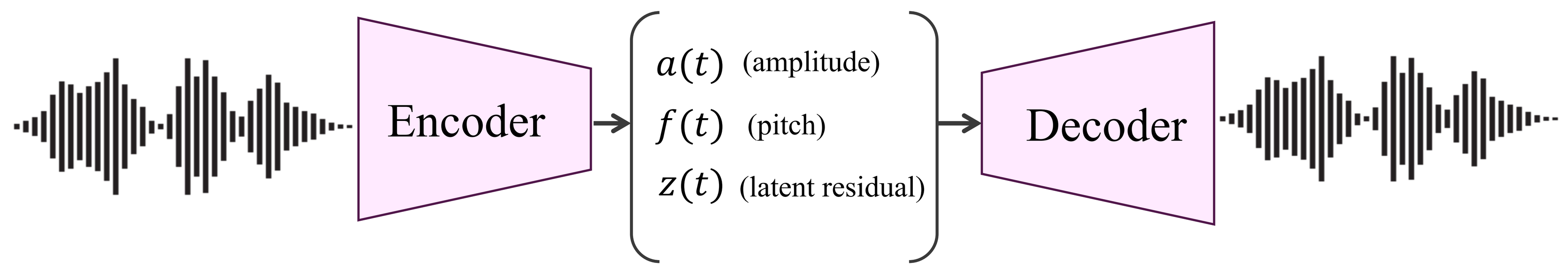
Training Pitch Detector by Visual Co-supervision



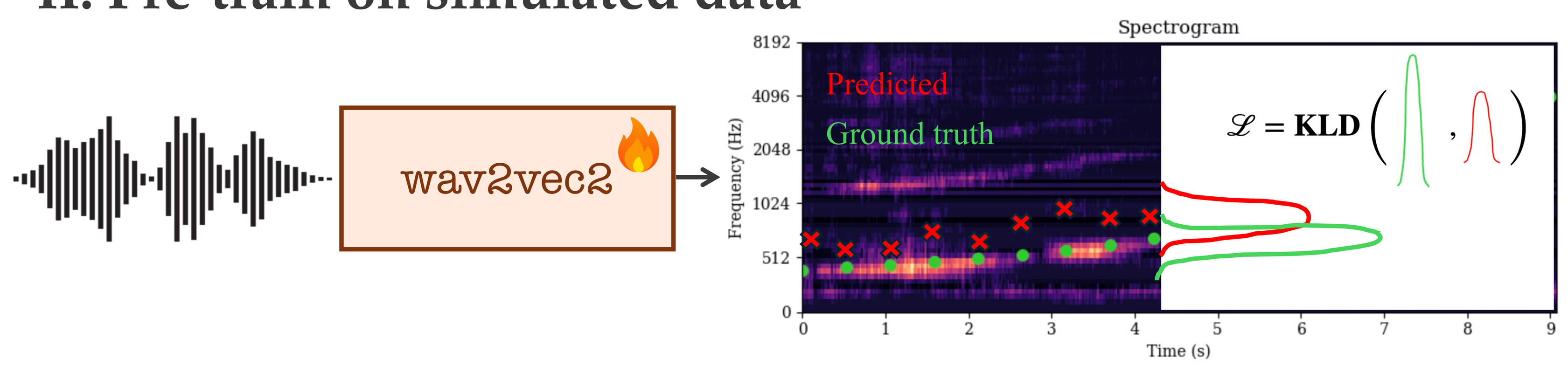
How to detect pitch in pouring sounds?

1. Simulate sounds of pouring with desired pitch profile
2. Pre-train a pitch detector network (wav2vec2) on simulated data
3. Fine-tune on real data with co-supervision from the video stream

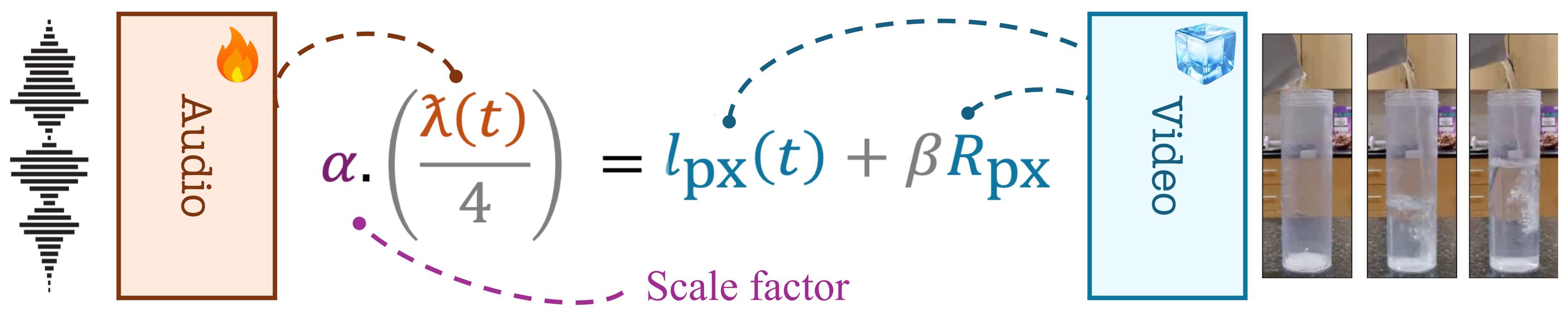
I. Simulate sounds of pouring



II. Pre-train on simulated data

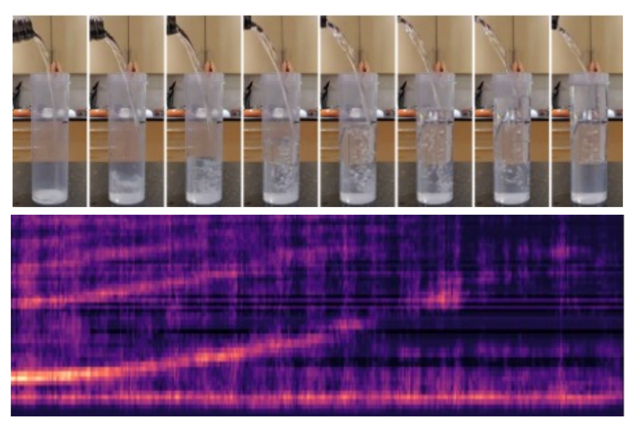


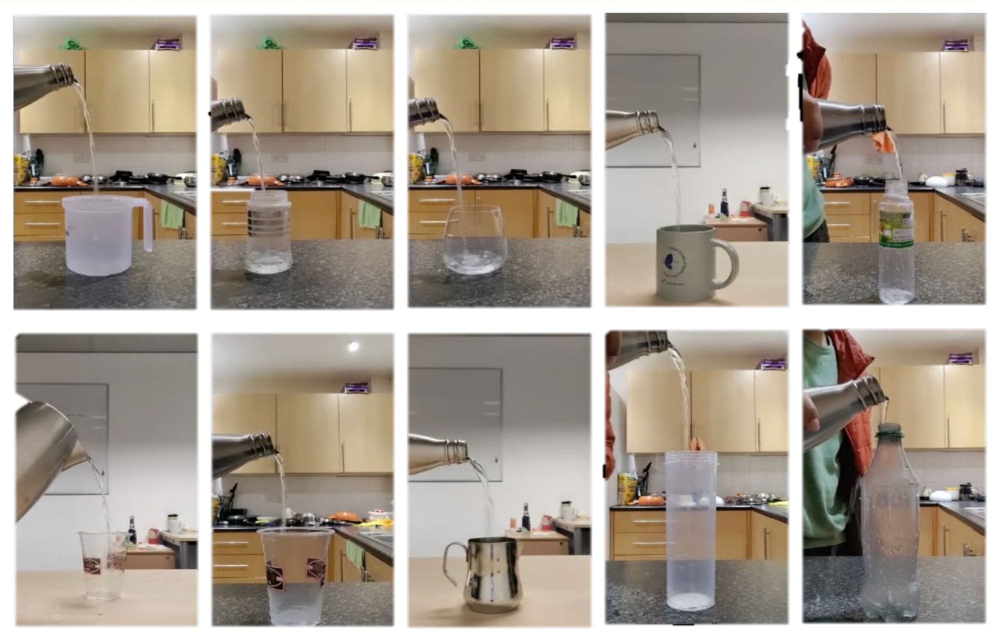
III. Fine-tune on real data with video teacher



Train and Evaluation Dataset: **Sound of Water 50**

For train/evaluation, 805 videos spanning 50+ containers (4 shapes, 5 materials, 2 liquids)





Experimental Results

Achieves an error rate of < 1 cm; and co-supervision helps!

Method	Test set I seen containers ↓
Baselines	
Yin [26]	30.80
PESTO [80]	11.70
CREPE [50]	7.61
argmax on spectrogram	4.60
Ours	
Audio-only	0.78
Co-supervised	**0.60**

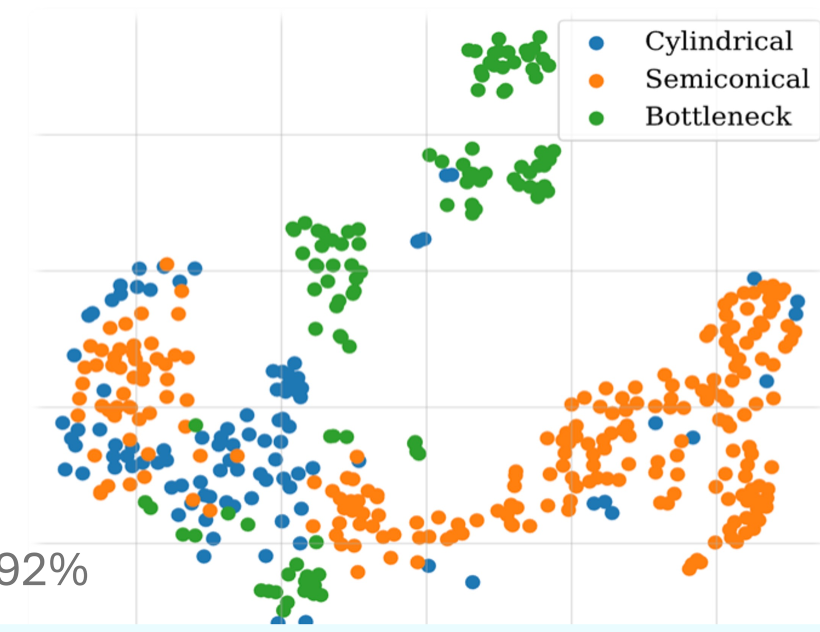
Property	Units	Notation	Test set I *Synthetic* ↓	Test set I *Co-supervised* ↓
Static properties				
Height	cm	H	2.23	2.27
Radius	cm	R	1.62	1.39
Dynamic properties				
Flow rate	ml/s	$Q(t)$	25.20	22.50
	s	$\tau_{\frac{1}{4}}(t)$	3.96	4.16
Time to fill	s	$\tau_{\frac{1}{2}}(t)$	1.62	1.49
	s	$\tau_{\frac{3}{4}}(t)$	1.53	1.07

The learned features encode liquid mass and container shape!

Samples from Wilson et al (2019)

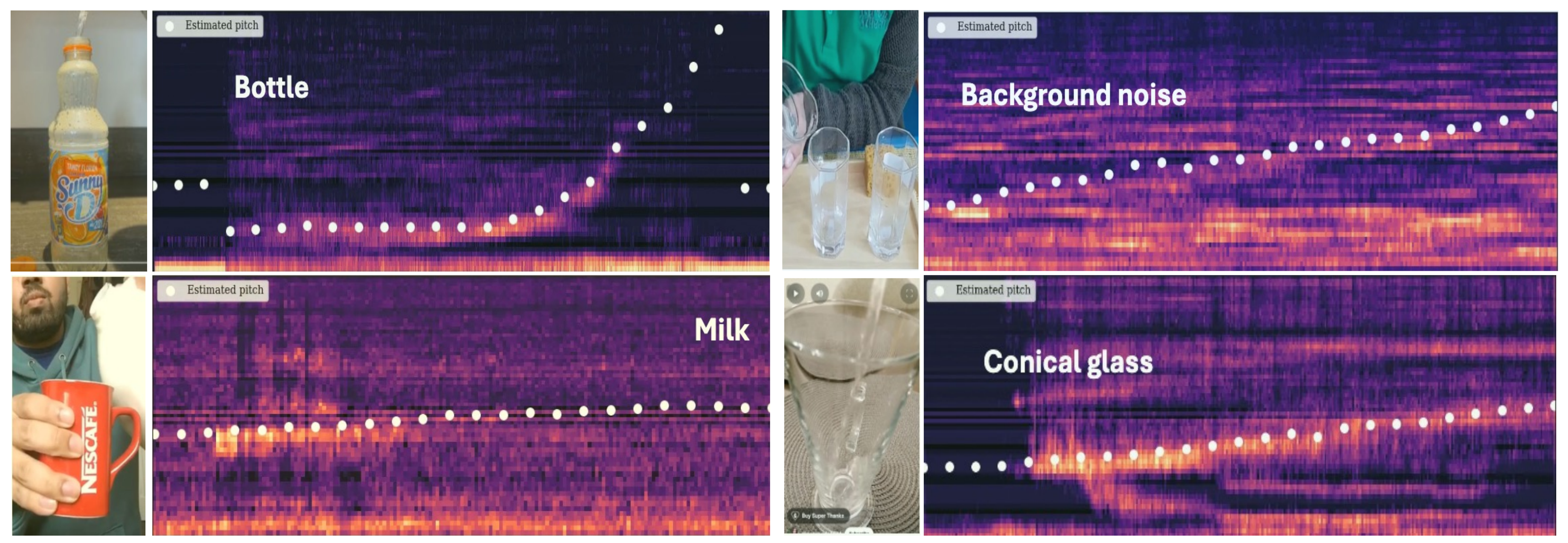
Liquid mass estimation on Wilson et al: MAE: 34ml

TSNE of learned representations

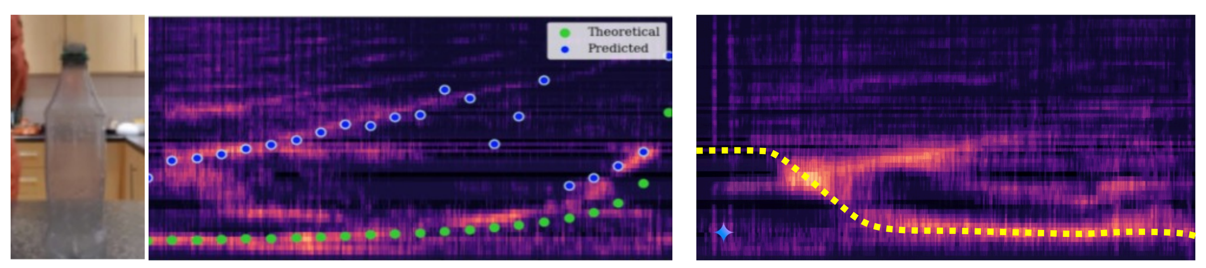


Shape classifier accuracy: 92%

Generalization to novel container shapes, materials, liquids and even in-the-wild YouTube samples.



Failure cases and future work



Higher harmonics

Radial resonance

Code & Models

piyush@robots.ox.ac.uk