Conformal Prediction for Zero-Shot Models

Julio Silva-Rodríguez · Jose Dolz · Ismail Ben Ayed - ÉTS Montréal

New to conformal prediction? Then, start here!

(brief) Introduction to Split Conformal Prediction

- **Goal: constructing predictive sets with valid coverage from any pre-trained black-box predictor scores.**

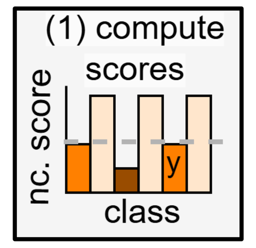
Coverage guarantee

$$\mathcal{P}(Y \in C(\mathbf{x})) \geq 1 - \alpha$$

predicted sets — *error rate (e.g., 10%)*

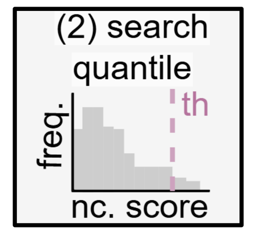
Marginal over $\mathcal{XY}$

Under cal/test exchangeability



$$\mathcal{S}(\mathbf{x}, y) = 1 - \hat{p}_{k=y}$$

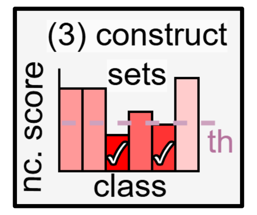
Non-conformity score



Calibration (labeled)

$$\hat{s} = \inf\left[s : \frac{|i \in \{1, \dots, N\} : s_i \leq s|}{N} \geq \frac{\lceil (N+1)(1-\alpha) \rceil}{N} \right]$$

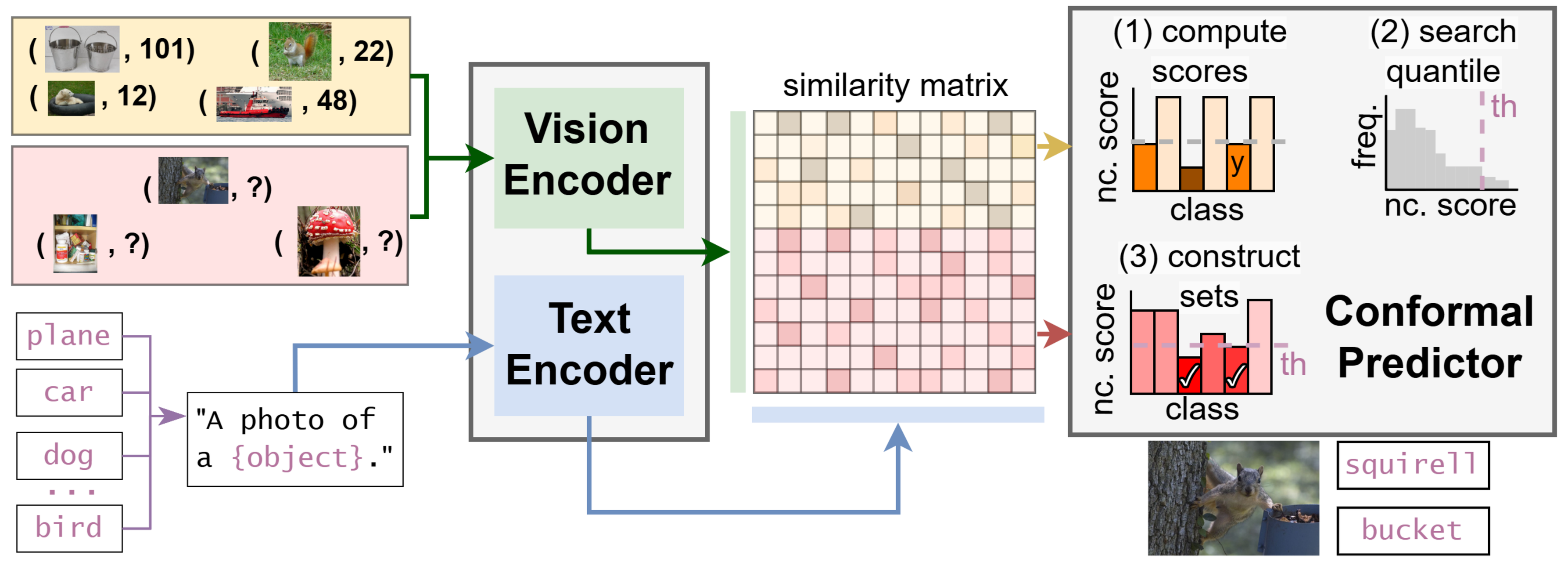
quantile



Test (unlabeled)

$$\mathcal{C}(\mathbf{x}) = \{ y \in \mathcal{Y} : \mathcal{S}(\mathbf{x}, y) \leq \hat{s} \}$$

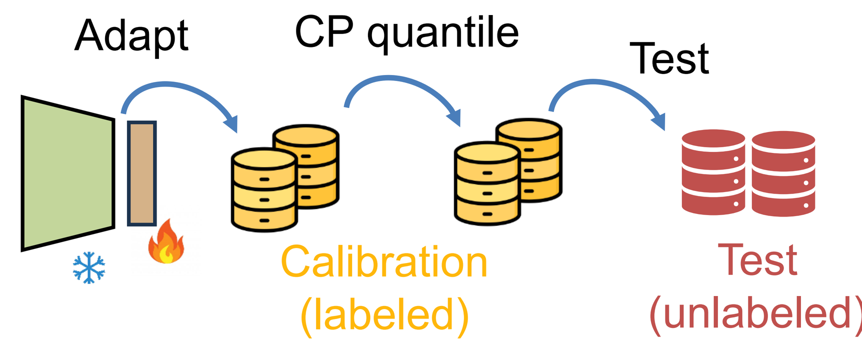
1 Conformal Prediction in Vision-Language Models

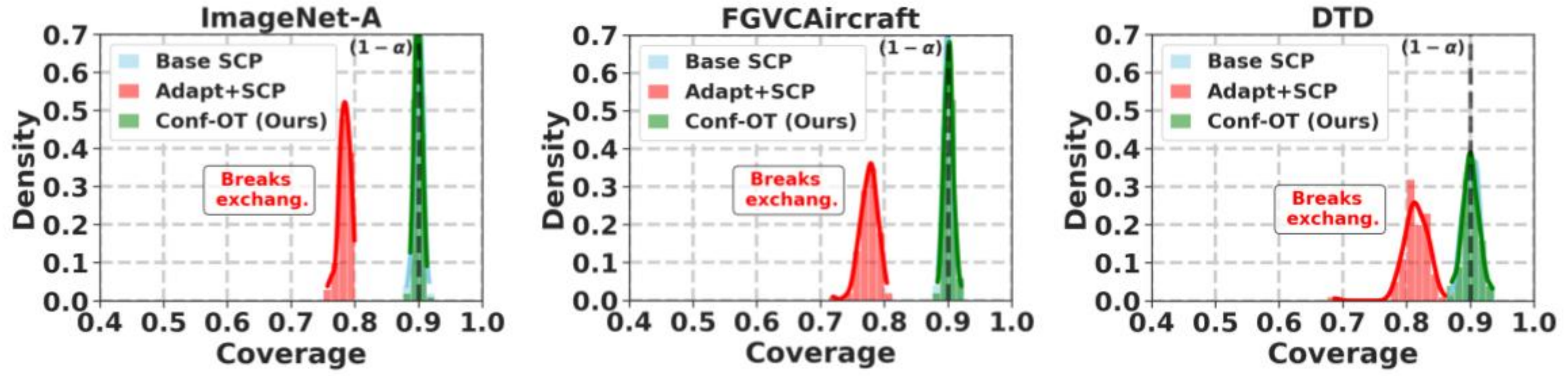


How to improve the conformal sets when transferring VLMs to new domains/tasks?

2

Transfer learning scenario: Can we adapt and conformalize using the same calibration data?

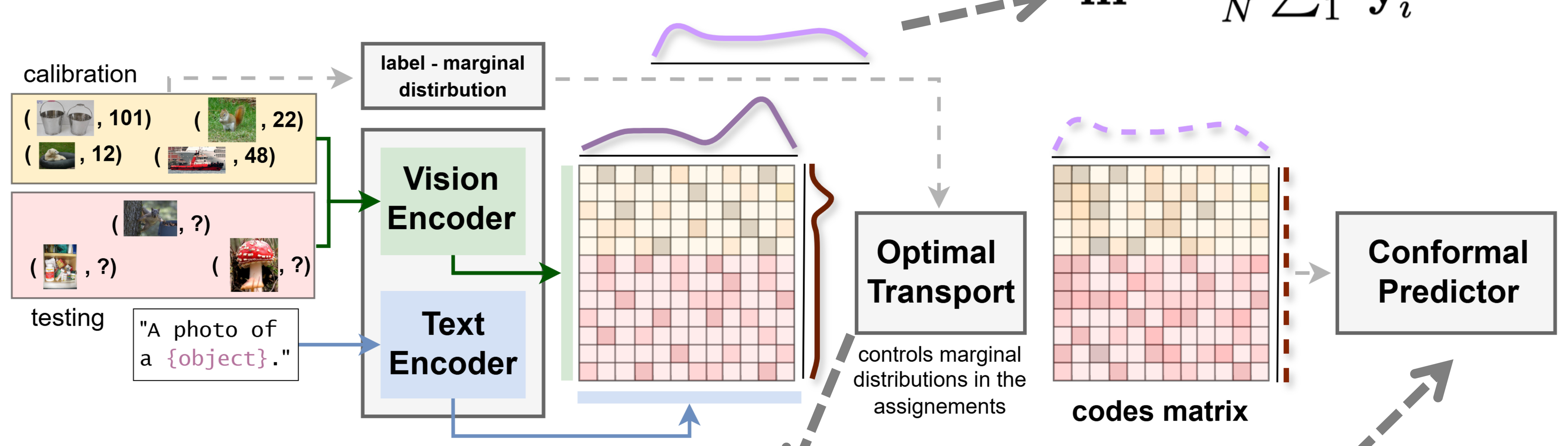




3

Conf-OT: enhancing the produced conformal sets trough **<u>transductive</u>**, **<u>unsupervised</u>** optimal transport.

Maintains exchangeability between calibration/test scores



$$\mathbf{m} = \frac{1}{N} \sum_{1}^{N} \mathbf{y}_i^{\text{ohe}}$$

$$\max_{\mathbf{Q} \in \mathcal{Q}} tr(\mathbf{Q}^\top \mathbf{S})$$

$$\mathcal{Q} = \{\mathbf{Q} \mid \mathbf{Q} \mathbf{1}_{(N+M)} = \mathbf{m}, \mathbf{Q}^\top \mathbf{1}_K = \mathbf{u}_{(\mathbf{N+M})}\}$$

$$\mathbf{Q}^*$$

$$\mathcal{D}_{\text{cal}} = \{(\boldsymbol{q}_i^{*\top}, y_i)\}_{i=1}^{N}$$

$$\mathcal{D}_{\text{test}} = \{(\boldsymbol{q}_i^{*\top})\}_{i=N+1}^{N+M}$$

In Nutshell

- **Split conformal prediction provides predictive sets with valid user-specified coverage (e.g. 90% or 95% accuracy) using a small labeled calibration set.**
- **CLIP models require adaptation when deployed to downstream tasks. However, using calibration data for adaptation breaks the exangeability assumptions, producing unreliable sets.**
- **Conf-OT is an unsupervised, transductive solver that reduces the domain gap while yet maintaining the desired coverage.**

GT: '<u>butternut squash</u>'
CP: ['mitten', 'acorn squash', '<u>butternut squash</u>', 'fig']
CP+*Ours*: ['mitten', '<u>butternut squash</u>']

GT: '<u>suspension bridge</u>'
CP: ['chainling fence', 'pier', 'pole', 'spider web', 'steel arch bridge', '<u>suspension bridge</u>', 'tripod']
CP+*Ours*: ['steel arch bridge', '<u>suspension bridge</u>']

GT: '<u>beagle</u>'
CP: ['basset', '<u>beagle</u>', 'walker hound', 'english foxhound', 'fountain', 'pedestal']
CP+*Ours*: ['<u>beagle</u>', 'fountain', 'pedestal']

GT: '<u>monastery</u>'
CP: ['castle', '<u>monastery</u>', 'palace', 'stupa']
CP+*Ours*: ['<u>monastery</u>', 'palace']

Results atop CLIP models and non-conformity scores

Method	Top-1↑	α = 0.10 Cov	α = 0.10 Size↓	α = 0.10 CCV↓	α = 0.05 Cov.	α = 0.05 Size↓	α = 0.05 CCV↓
CLIP ResNet-50							
LAC [39]	54.7	0.900	10.77	9.82	0.950	19.22	5.91
w/ Conf-OT	**57.3**$_{+2.6}$	0.900	**8.61**$_{-2.2}$	**9.15**$_{-0.7}$	0.951	**15.53**$_{-3.7}$	**5.61**$_{-0.3}$
APS [51]	54.7	0.900	16.35	8.36	0.950	26.50	5.34
w/ Conf-OT	**57.3**$_{+2.6}$	0.900	**12.94**$_{-3.4}$	**7.64**$_{-0.7}$	0.950	**20.96**$_{-5.5}$	**5.03**$_{-0.3}$
RAPS [2]	54.7	0.900	13.37	8.46	0.950	22.06	5.44
w/ Conf-OT	**57.3**$_{+2.6}$	0.900	**11.17**$_{-2.2}$	**7.72**$_{-0.7}$	0.950	**17.24**$_{-4.8}$	**5.19**$_{-0.3}$
CLIP ViT-B/16							
LAC [39]	63.8	0.899	5.52	10.37	0.950	10.24	6.14
w/ Conf-OT	**66.7**$_{+2.9}$	0.900	**4.40**$_{-1.1}$	**9.48**$_{-0.9}$	0.949	**7.99**$_{-2.3}$	**5.80**$_{-0.3}$
APS [51]	63.8	0.900	9.87	8.39	0.950	16.92	5.51
w/ Conf-OT	**66.7**$_{+2.9}$	0.899	**7.64**$_{-2.2}$	**7.44**$_{-1.0}$	0.949	**12.58**$_{-4.3}$	**5.09**$_{-0.4}$
RAPS [2]	63.8	0.900	8.12	8.50	0.950	12.66	5.52
w/ Conf-OT	**66.7**$_{+2.9}$	0.900	**6.68**$_{-1.4}$	**7.48**$_{-1.0}$	0.949	**10.11**$_{-2.6}$	**5.16**$_{-0.4}$

Efficiency and comparison with transductive baselines

Method	Top-1↑	T	GPU	α = 0.10 Cov.	α = 0.10 Size↓	α = 0.10 CCV↓
LAC [39]	63.8	**0.42**	-	0.899	5.52	10.37
$\text{TIM}_{\text{KL}(\widehat{\mathbf{m}}\|\|\mathbf{u}_K)}$ [5]	64.7$_{+0.9}$	11.05	8.24	0.899	8.30$_{+2.8}$	10.41$_{+0.0}$
$\text{TIM}_{\text{KL}(\widehat{\mathbf{m}}\|\|\mathbf{m})}$ [5]	65.0$_{+1.2}$	11.03	8.24	0.898	7.73$_{+2.2}$	10.89$_{+0.5}$
TransCLIP [73]	65.1$_{+1.3}$	12.00	12.2	0.892	5.76$_{+0.2}$	11.02$_{+0.7}$
Conf-OT	**66.7**$_{+2.9}$	0.61	-	0.900	**4.40**$_{-1.1}$	**9.48**$_{-0.9}$
APS [51]	63.8	**0.54**	-	0.900	9.87	8.39
$\text{TIM}_{\text{KL}(\widehat{\mathbf{m}}\|\|\mathbf{u}_K)}$ [5]	64.7$_{+0.9}$	11.16	8.24	0.900	**7.24**$_{-2.6}$	9.32$_{+0.9}$
$\text{TIM}_{\text{KL}(\widehat{\mathbf{m}}\|\|\mathbf{m})}$ [5]	65.0$_{+1.2}$	11.14	8.24	0.900	7.82$_{-2.1}$	9.38$_{+1.0}$
TransCLIP [73]	65.1$_{+1.3}$	12.12	12.2	0.892	8.27$_{-1.6}$	11.50$_{+3.1}$
Conf-OT	**66.7**$_{+2.9}$	0.72	-	0.899	7.64$_{-2.2}$	**7.44**$_{-1.0}$
RAPS [2]	63.8	**0.55**	-	0.900	8.12	8.50
$\text{TIM}_{\text{KL}(\widehat{\mathbf{m}}\|\|\mathbf{u}_K)}$ [5]	64.7$_{+0.9}$	11.15	8.24	0.900	7.18$_{-0.9}$	9.32$_{+0.8}$
$\text{TIM}_{\text{KL}(\widehat{\mathbf{m}}\|\|\mathbf{m})}$ [5]	65.0$_{+1.2}$	11.36	8.24	0.900	7.68$_{-0.4}$	9.42$_{+0.9}$
TransCLIP [73]	65.1$_{+1.3}$	12.12	12.3	0.899	7.17$_{-1.0}$	10.20$_{+1.7}$
Conf-OT	**66.7**$_{+2.9}$	0.74	-	0.900	**6.68**$_{-1.4}$	**7.48**$_{-1.0}$