

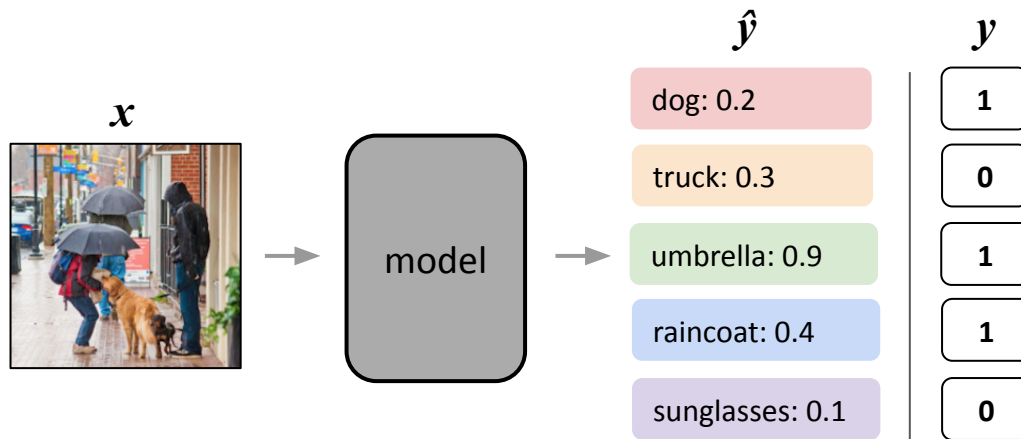
General Multi-label Image Classification with Transformers

Jack Lanchantin, Tianlu Wang, Vicente Ordoñez, Yanjun Qi

University of Virginia | Department of Computer Science

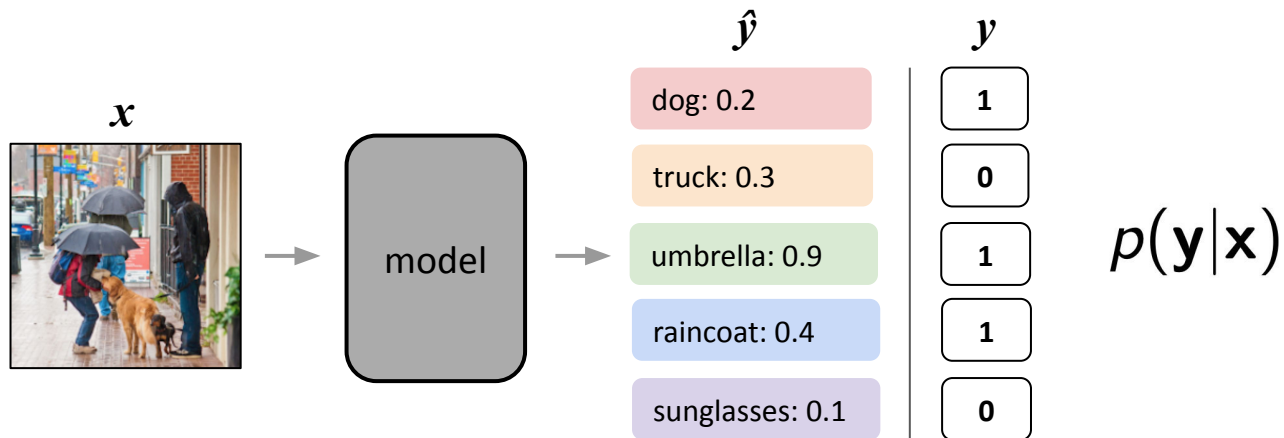
Multi-label Classification

Given input \mathbf{x} , predict the *set* of target labels $\{y_1, y_2, \dots, y_L\}$, $y_i \in \{0, 1\}$



Multi-label Classification

Given input \mathbf{x} , predict the *set* of target labels $\{y_1, y_2, \dots, y_L\}$, $y_i \in \{0, 1\}$



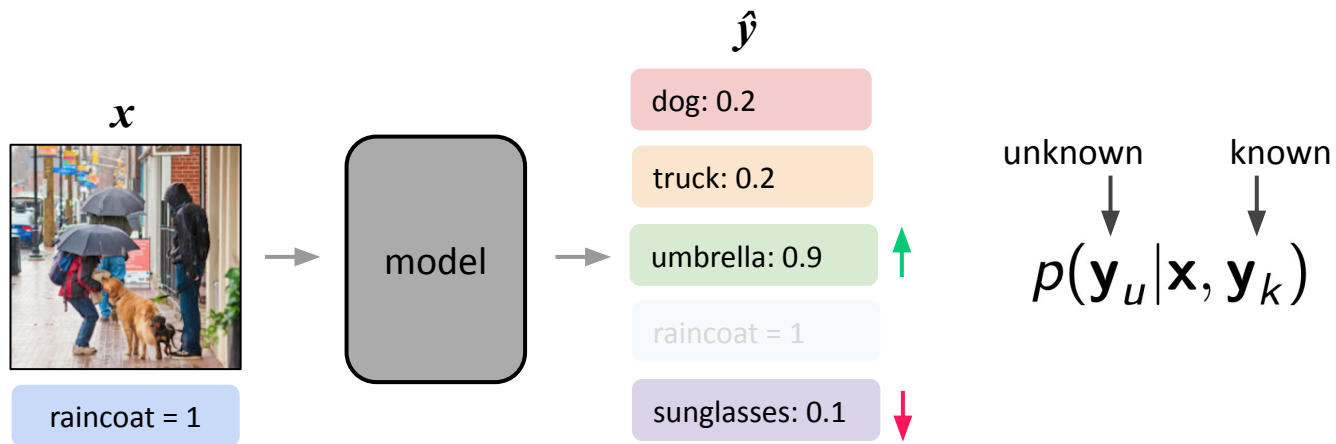
Inference with partial knowledge

x



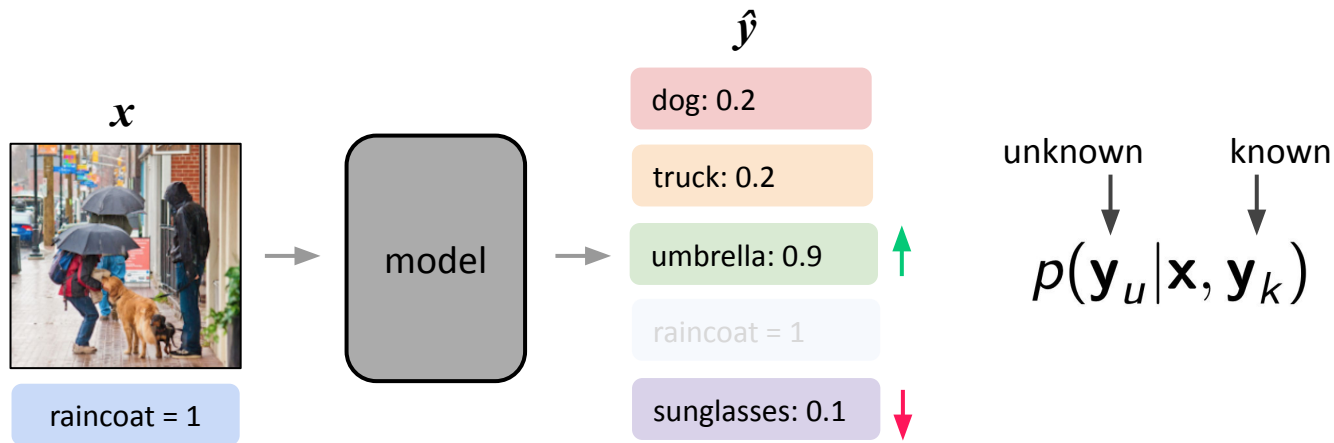
raincoat = 1

Inference with partial knowledge



Inference with partial knowledge

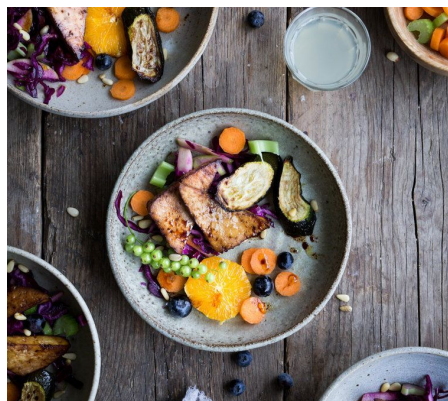
Information (e.g. certain labels) can be known **prior to** performing any **visual recognition**



leverage known information during inference!

Inference with context-specific information: a realistic setting

Social Media Tags



#avocado #bowl

Geo-Location Tags



📍 Lake George, NY

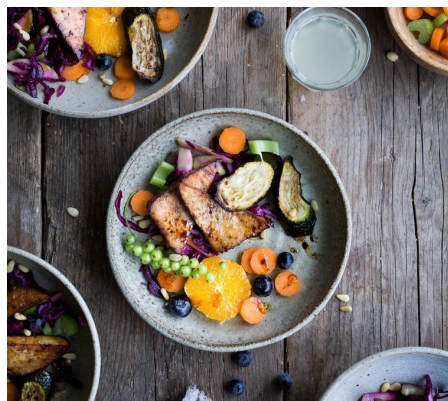
News Captions



Mt. Sinabung sent smoke over western Indonesia Wednesday.

Inference with context-specific information: a realistic setting

Social Media Tags



#avocado #bowl

avocado

bowl

Geo-Location Tags



📍 Lake George, NY

lake

News Captions

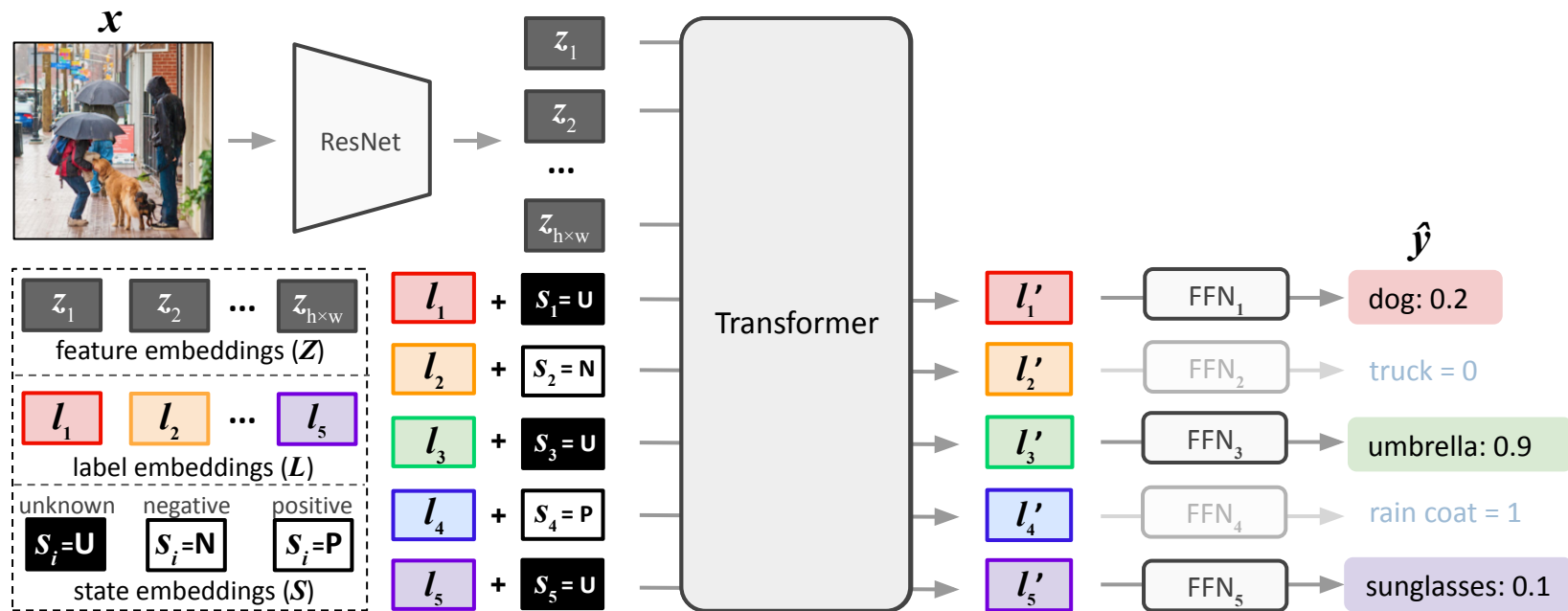


Mt. Sinabung sent smoke over western Indonesia Wednesday.

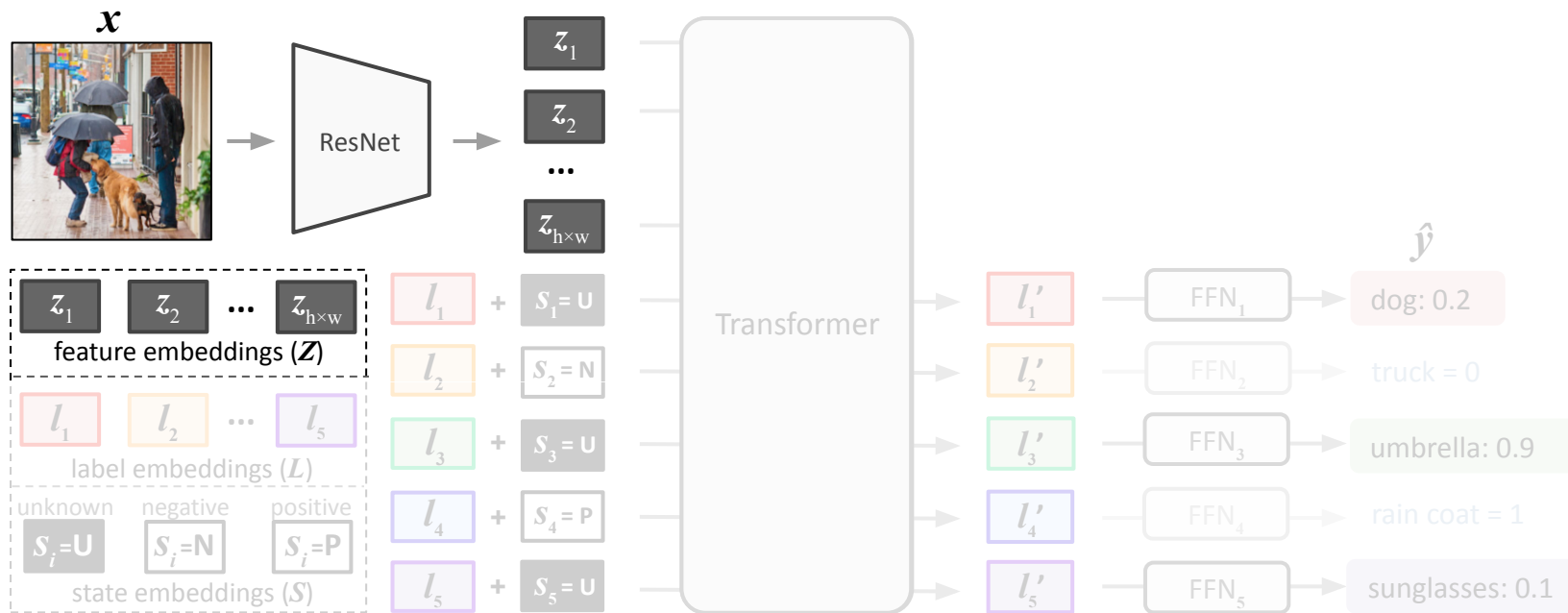
mountain

smoke

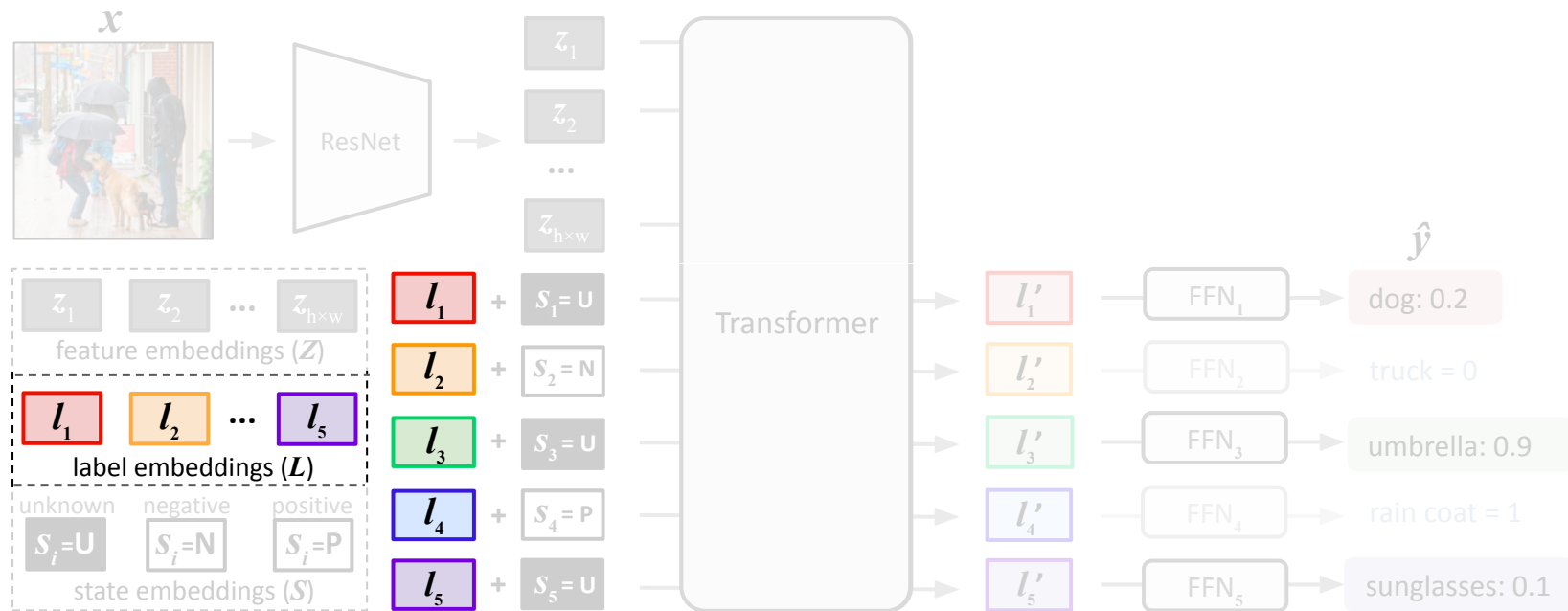
Classification Transformer (C-Tran)



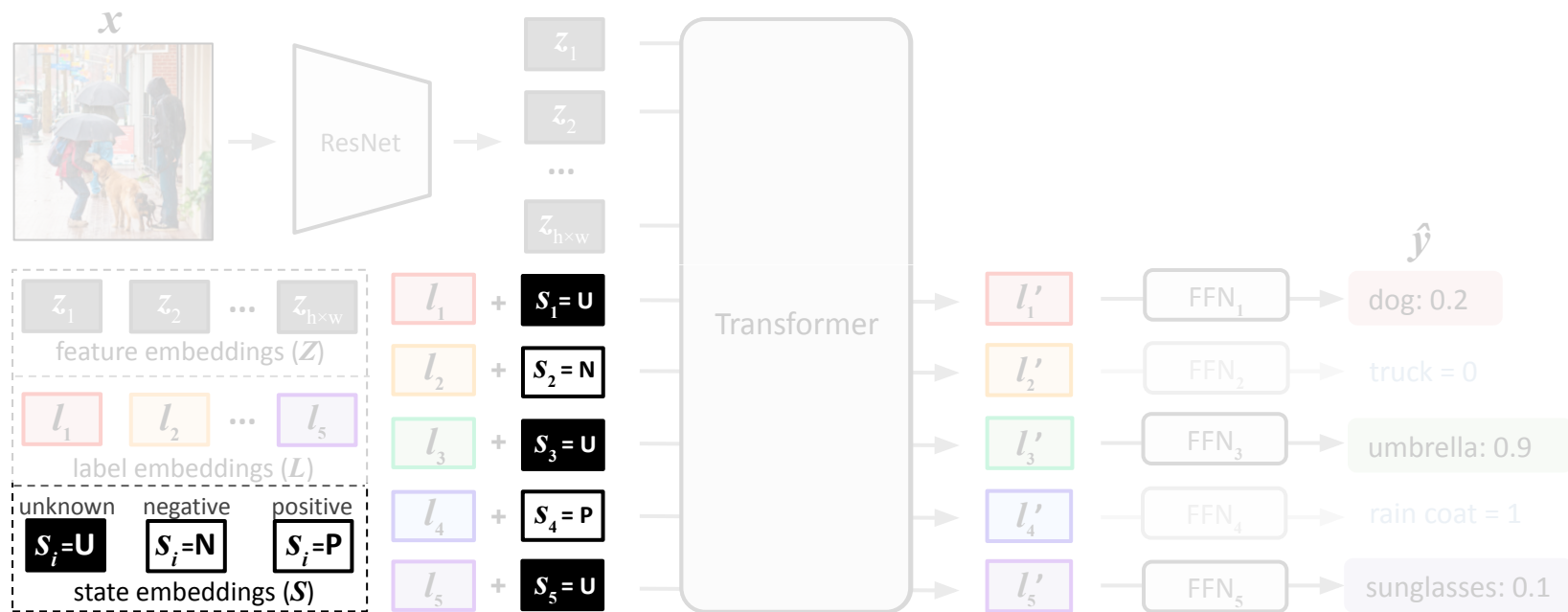
Classification Transformer (C-Tran)



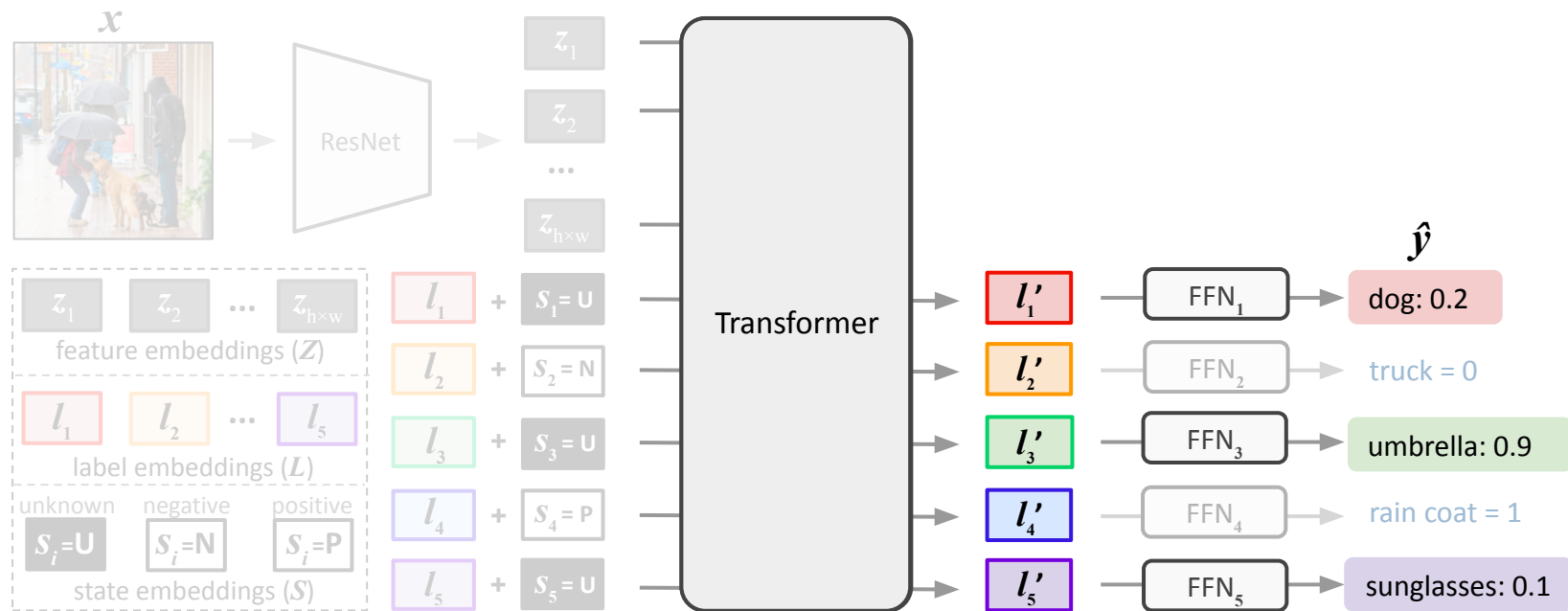
Classification Transformer (C-Tran)



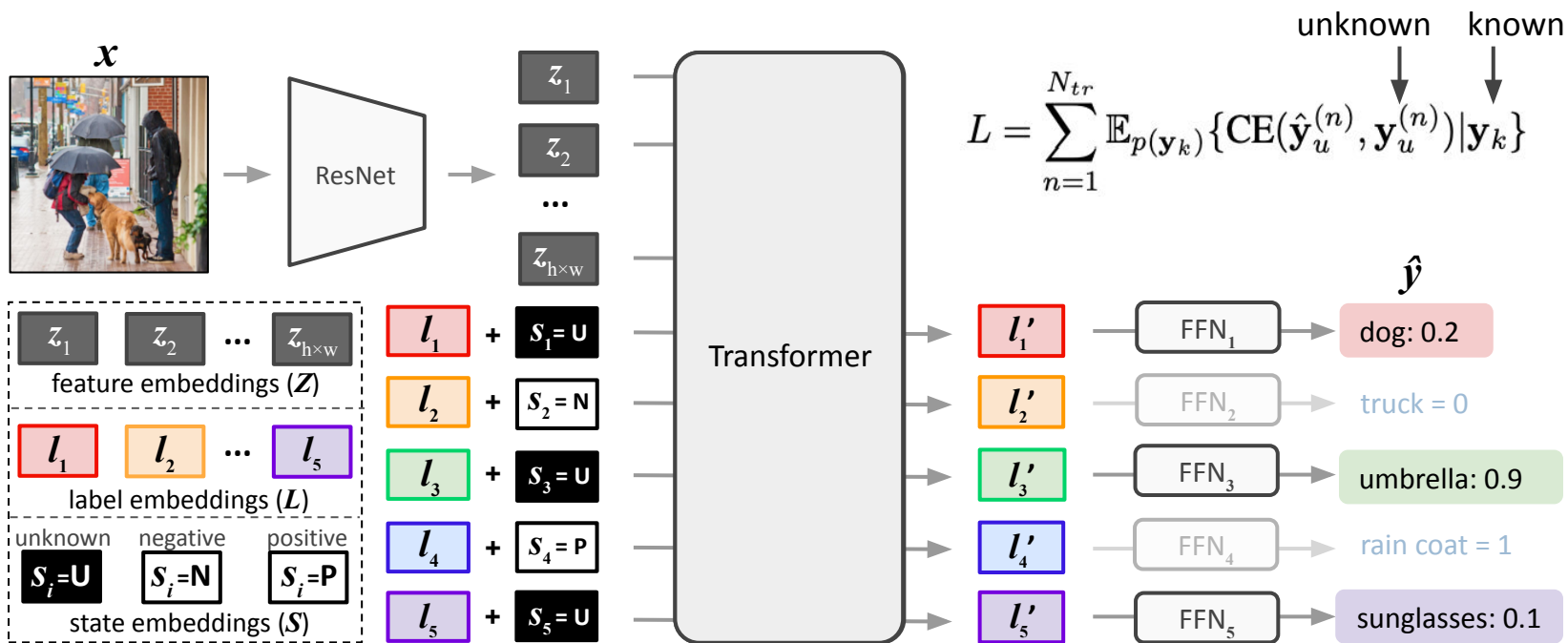
Classification Transformer (C-Tran)



Classification Transformer (C-Tran)

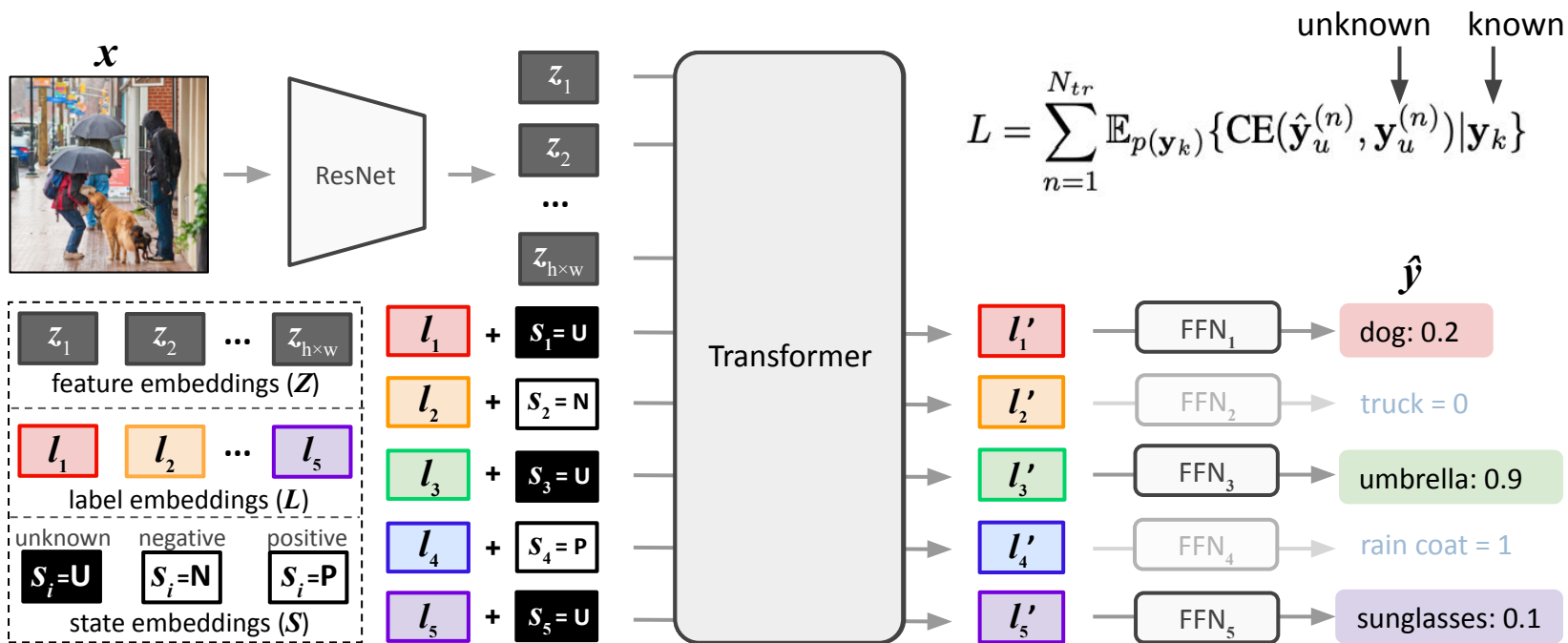


Classification Transformer (C-Tran)



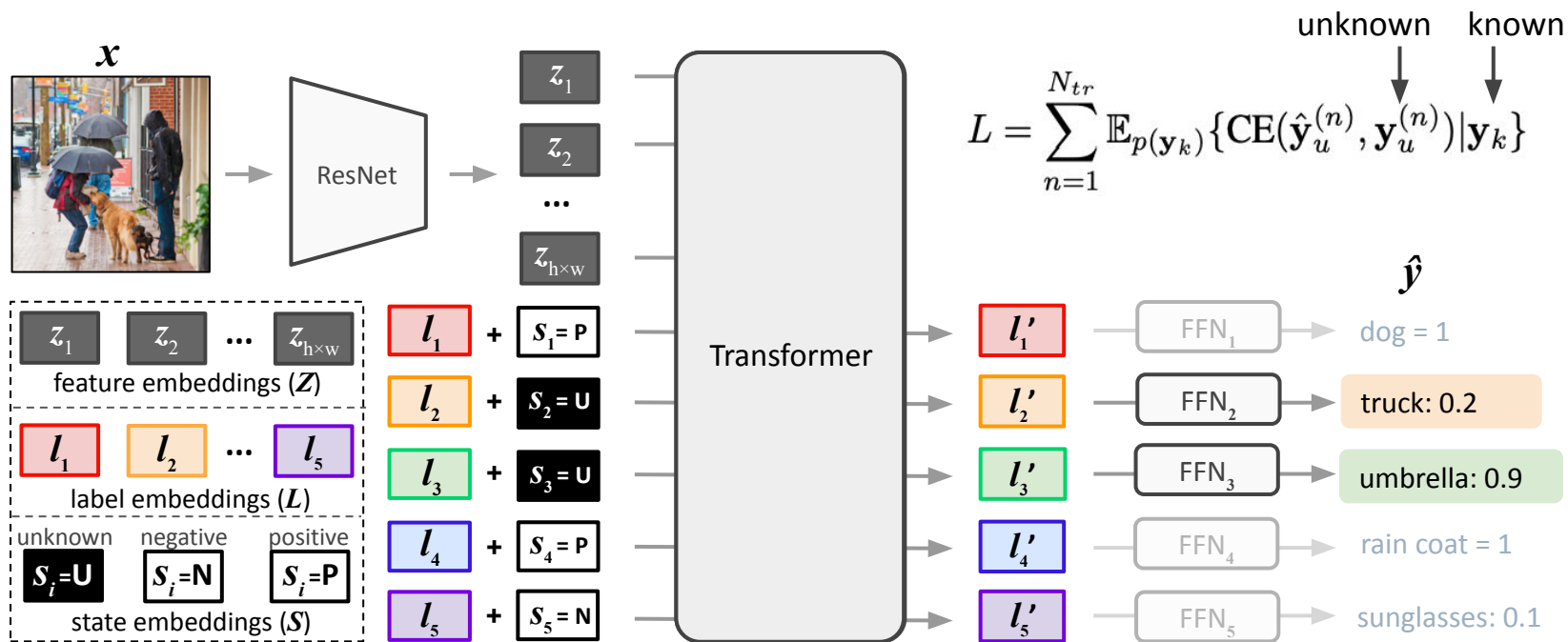
randomly sample a random amount of known labels

Classification Transformer (C-Tran)



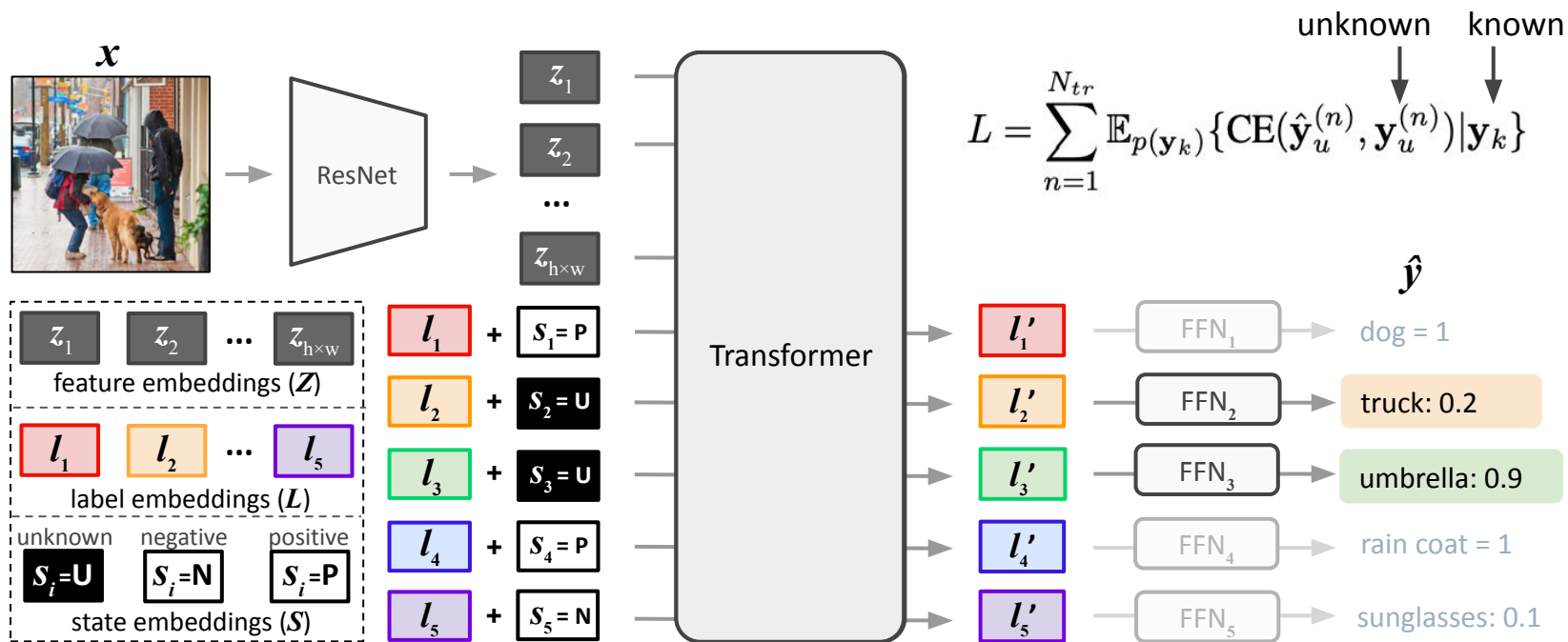
randomly sample a random amount of known labels

Classification Transformer (C-Tran)



randomly sample a random amount of known labels

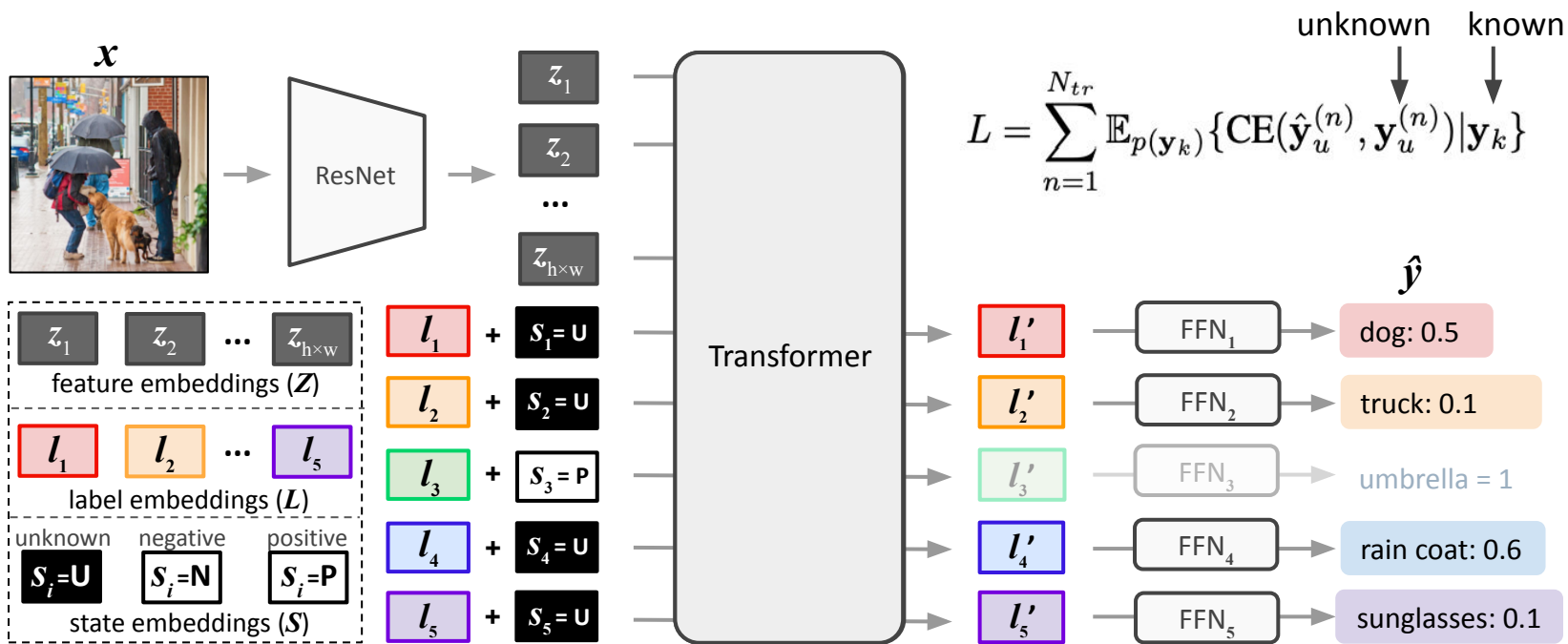
Classification Transformer (C-Tran)



$$L = \sum_{n=1}^{N_{tr}} \mathbb{E}_{p(\mathbf{y}_k)} \{ \text{CE}(\hat{\mathbf{y}}_u^{(n)}, \mathbf{y}_u^{(n)}) | \mathbf{y}_k \}$$

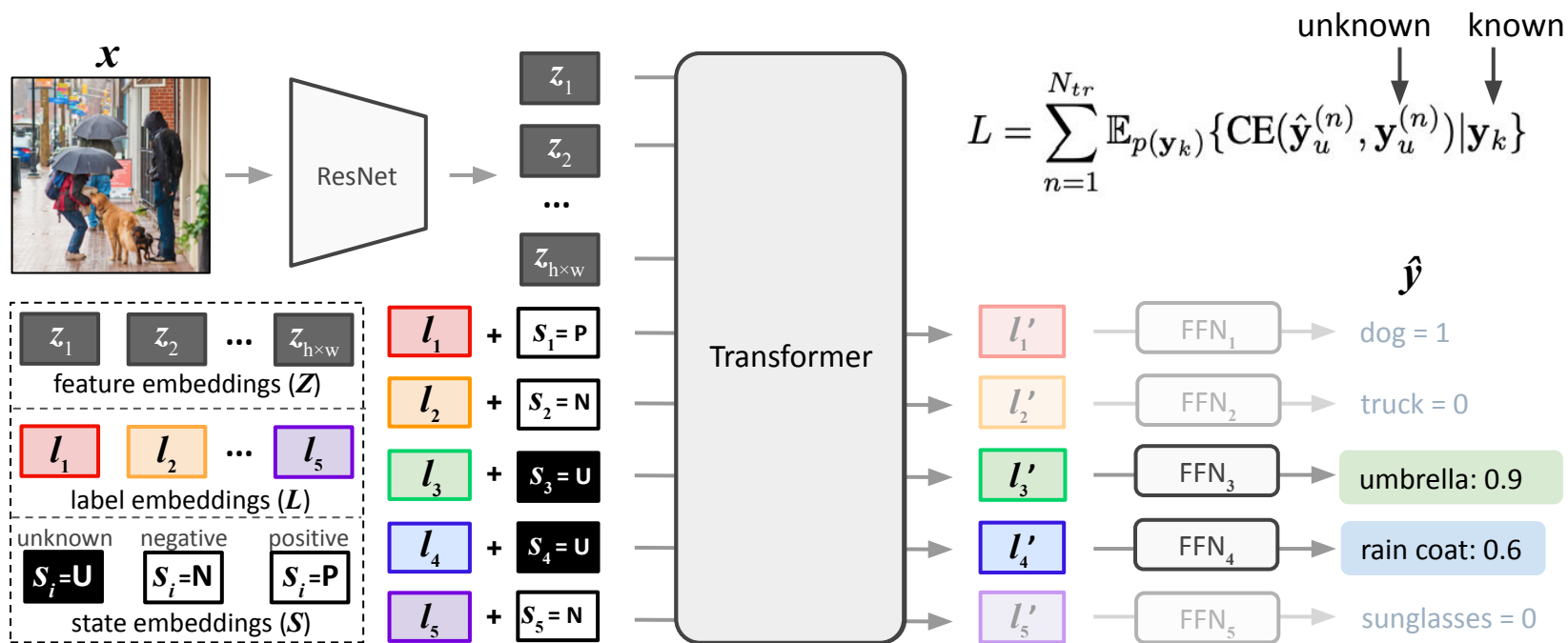
randomly sample a random amount of known labels

Classification Transformer (C-Tran)



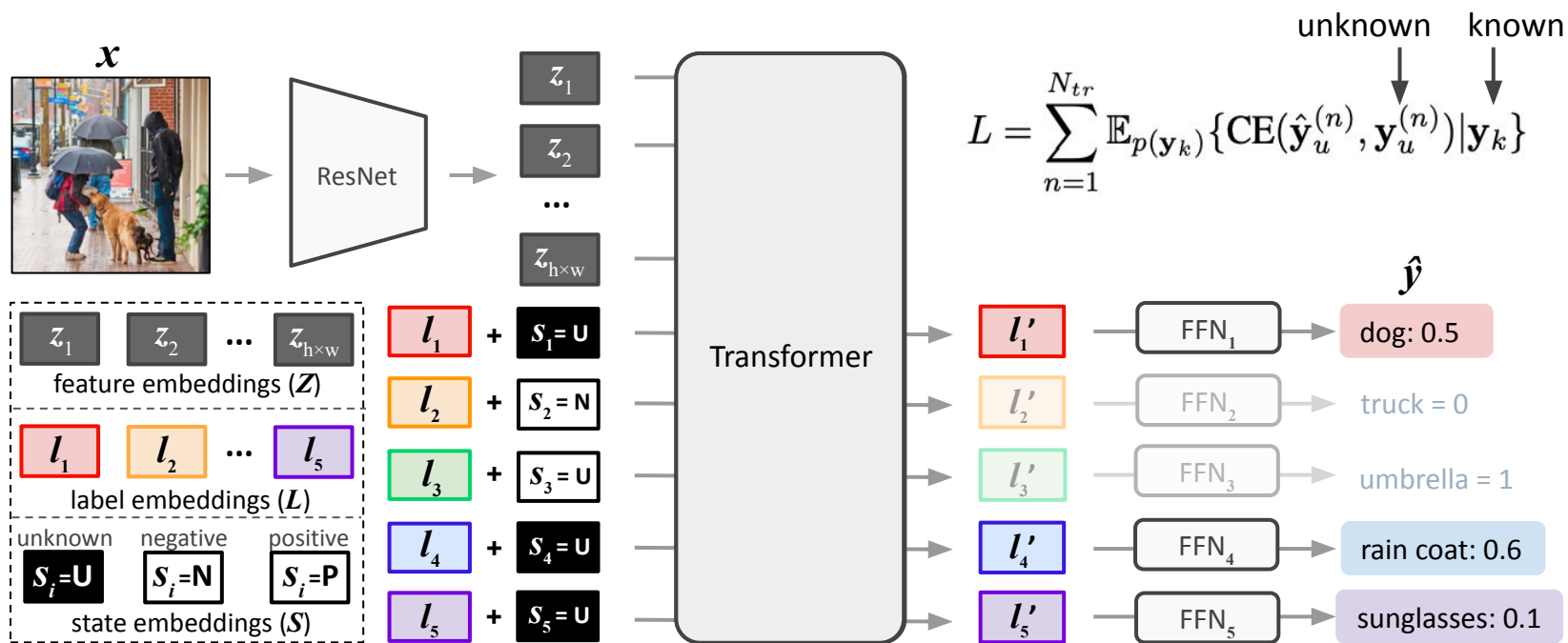
randomly sample a random amount of known labels

Classification Transformer (C-Tran)



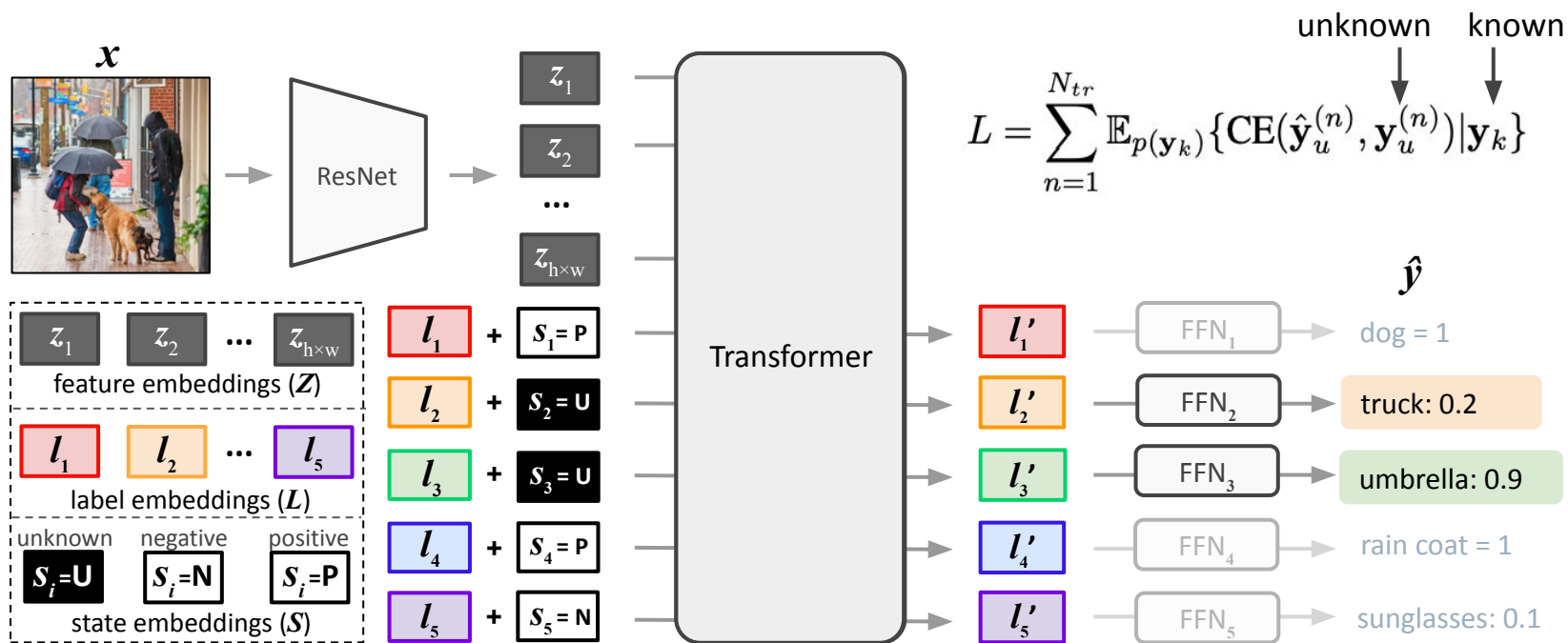
randomly sample a random amount of known labels

Classification Transformer (C-Tran)



randomly sample a random amount of known labels

Classification Transformer (C-Tran)



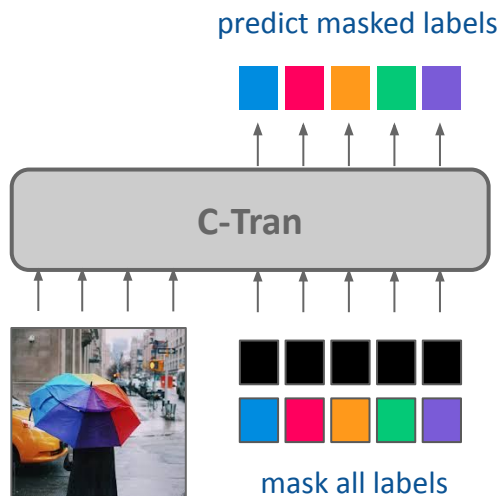
$$L = \sum_{n=1}^{N_{tr}} \mathbb{E}_{p(y_k)} \{ \text{CE}(\hat{y}_u^{(n)}, y_u^{(n)}) | y_k \}$$

unknown known

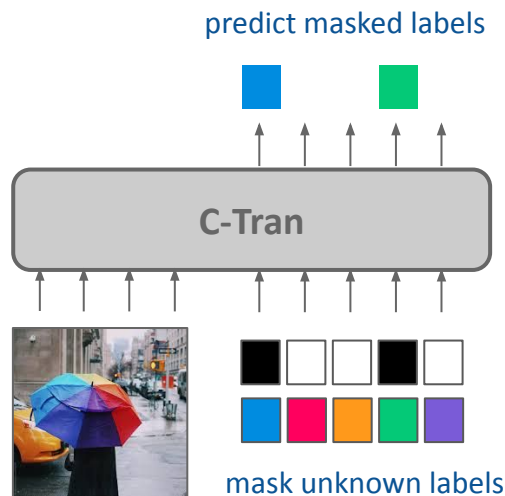
randomly sample a random amount of known labels

Three different inference settings

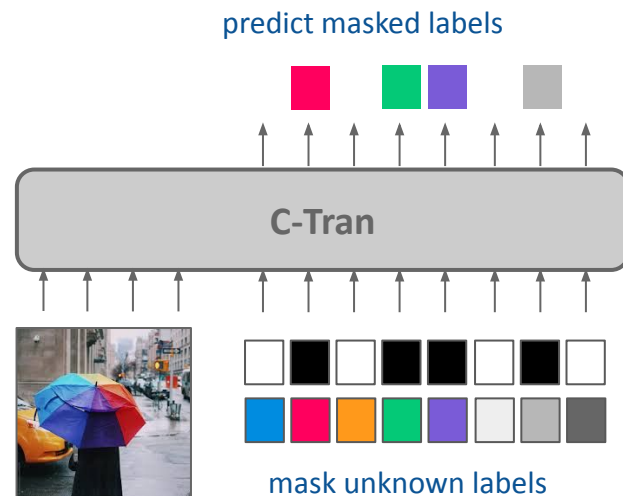
1. Regular Inference



2. Partial Label Inference

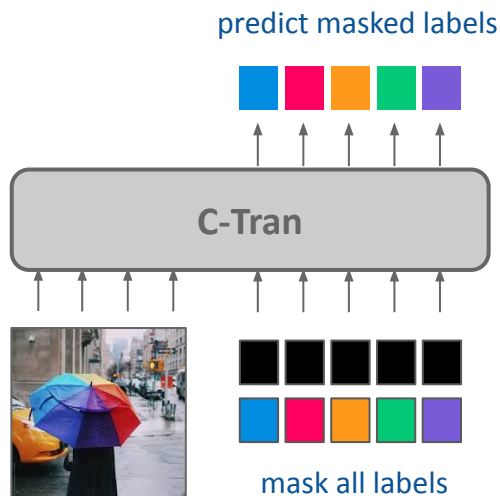


3. Extra Label Inference

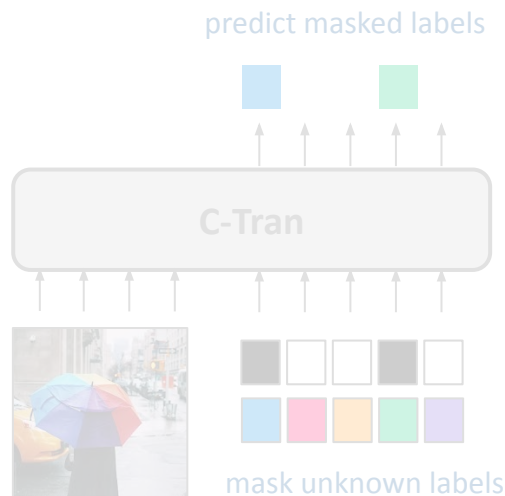


Three different inference settings

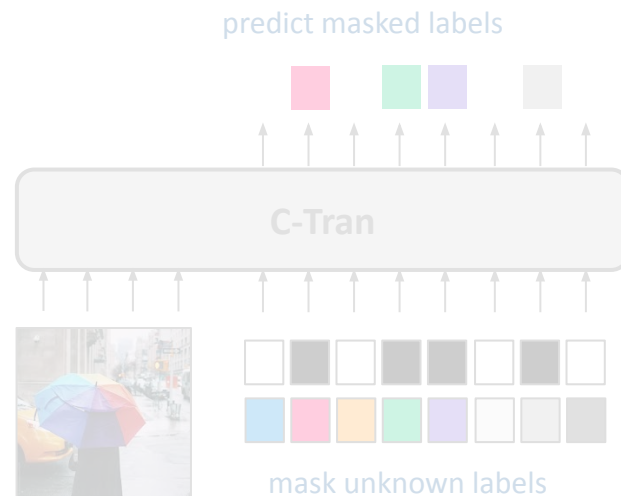
1. Regular Inference



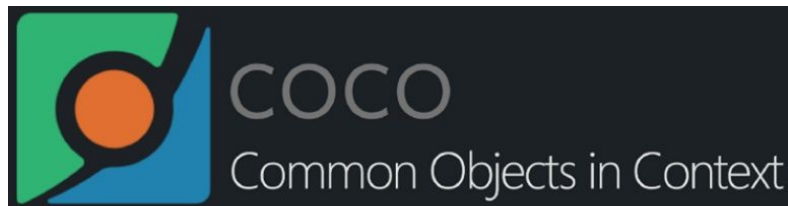
2. Partial Label Inference



3. Extra Label Inference



Regular inference setting



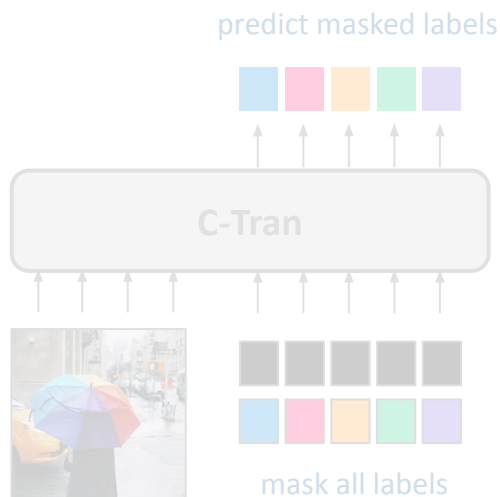
0.8 improvement in mAP and F1 score



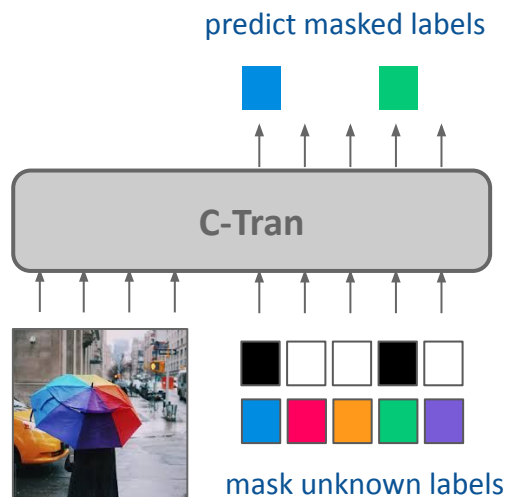
1.0 improvement in mAP and F1 score

Three different inference settings

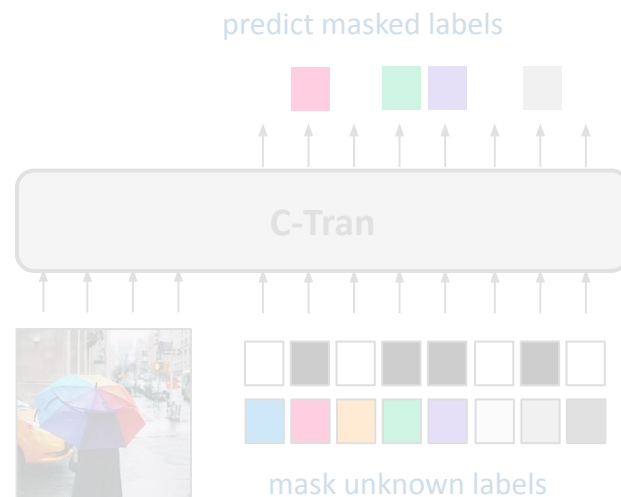
1. Regular Inference




2. Partial Label Inference



3. Extra Label Inference

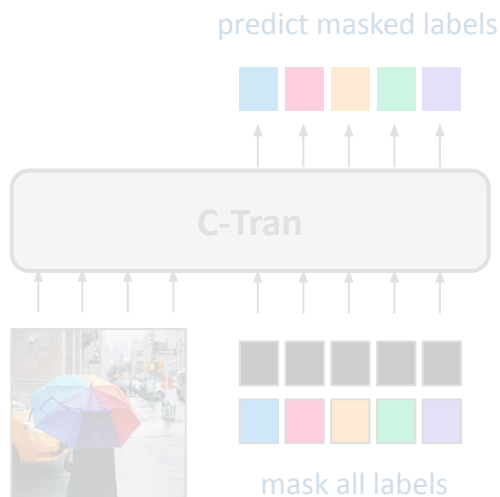


Partial label inference setting

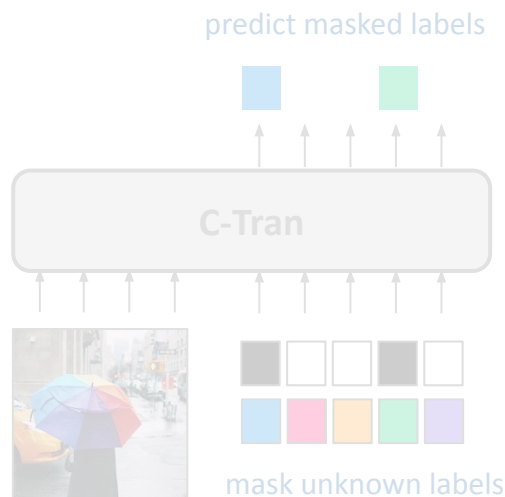
Image	True Labels	ResNet-101	C-Tran	C-Tran + partial labels
	fork knife, spoon, bowl, chair, diningtable	fork, sandwich, diningtable, spoon, cup	fork, knife, diningtable, person, cake	spoon=1, trafficlight=0, bench=0, dog=0, ... fork, knife, diningtable, person, bowl

Three different inference settings

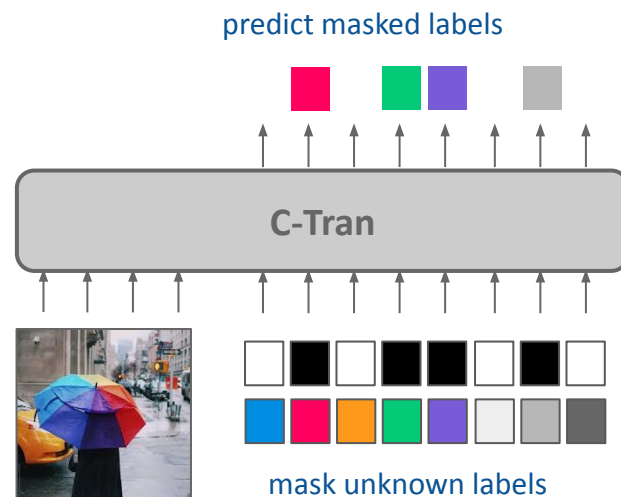
1. Regular Inference



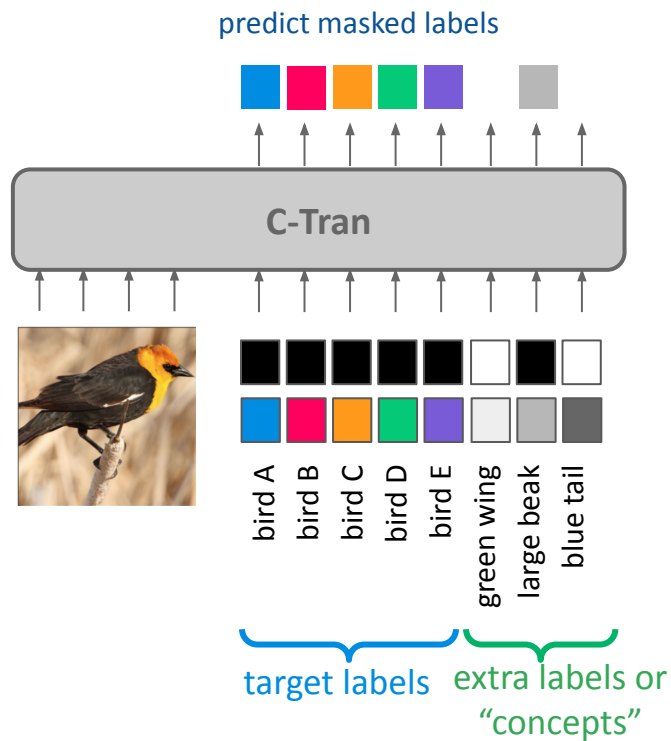
2. Partial Label Inference




3. Extra Label Inference



3. Extra label inference setting

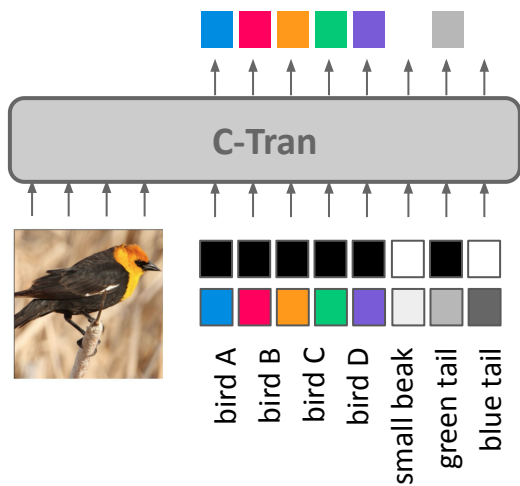


Extra label inference setting

Image	True Label	C-Tran	C-Tran + Extra Labels
	Anna Hummingbird	Rufous Hummingbird (96%)	has_bill_shape_needle = 1, has_wing_color_green=1, has_upperparts_color=green=1, has_back_color_blue=0, has_back_color_brown=0 ... Anna Hummingbird (99%)

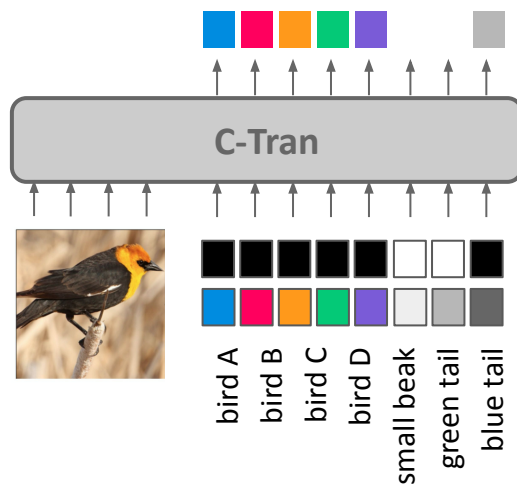
Model intervention

What kind of bird is this, given that I know it has a **blue tail**?



Counterfactual testing

What kind of bird would this be **if it had a green tail**?



Contributions

Flexibility: multi-label image classification under **arbitrary subsets** of extra or partial labels

Accuracy: state-of-the-art results on **six different datasets** across three inference settings

Interactivity: state embeddings enables users to easily interact with the model and **test counterfactuals**

Thank You

github.com/QData/C-Tran



Tianlu Wang



Vicente Ordoñez



Yanjun Qi



UNIVERSITY of VIRGINIA