



A cheap and effective way to expand your model's performance on various tasks across various concepts:

Learn tasks from densely annotated data



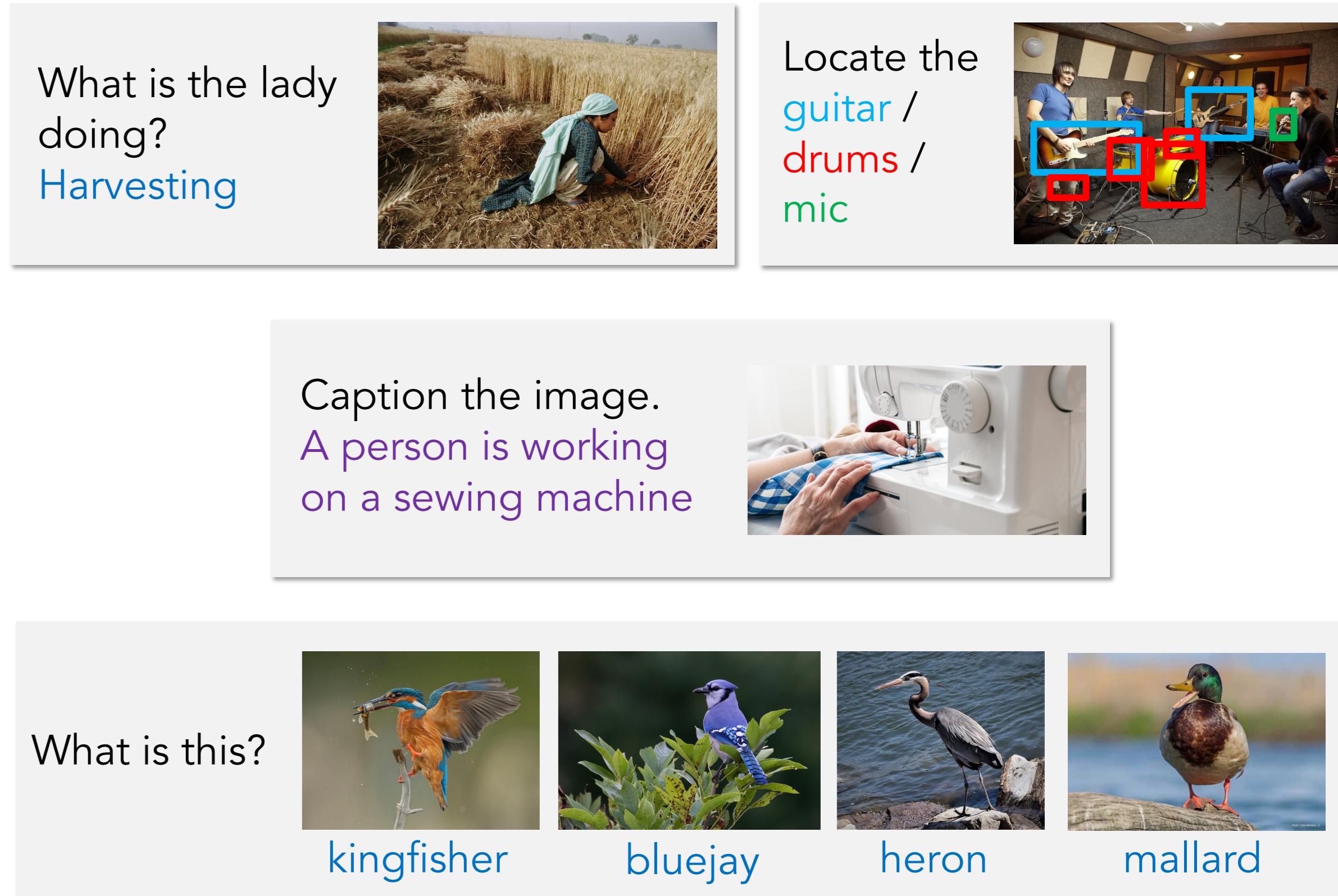
Expensive, limited concepts

Learn concepts from web search data



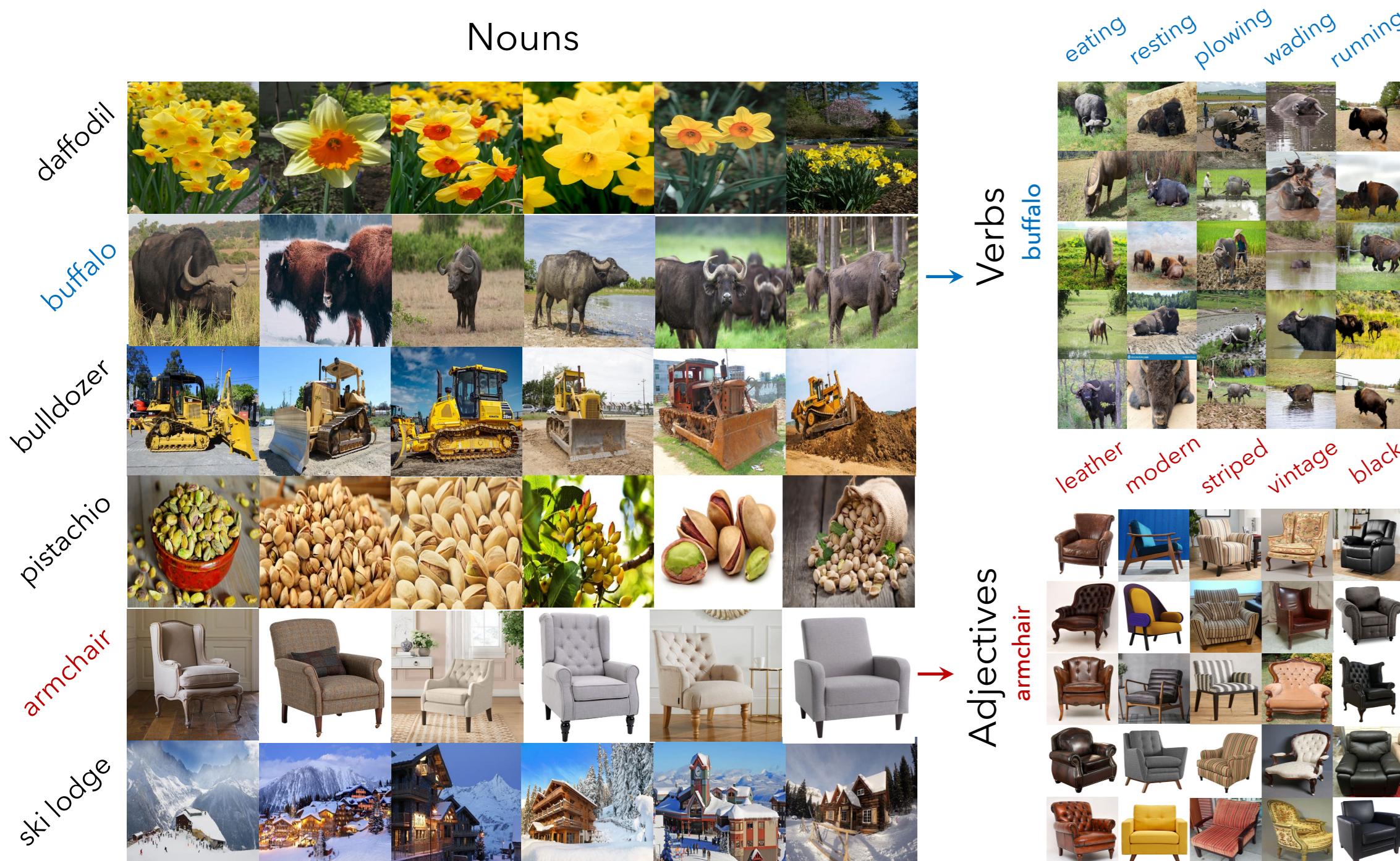
Cheap, many concepts

Benefit from transfer!



## Why web data?

- High quality data for tail-end concepts
- Uncluttered, object-centric data
- Latest concepts, on-demand

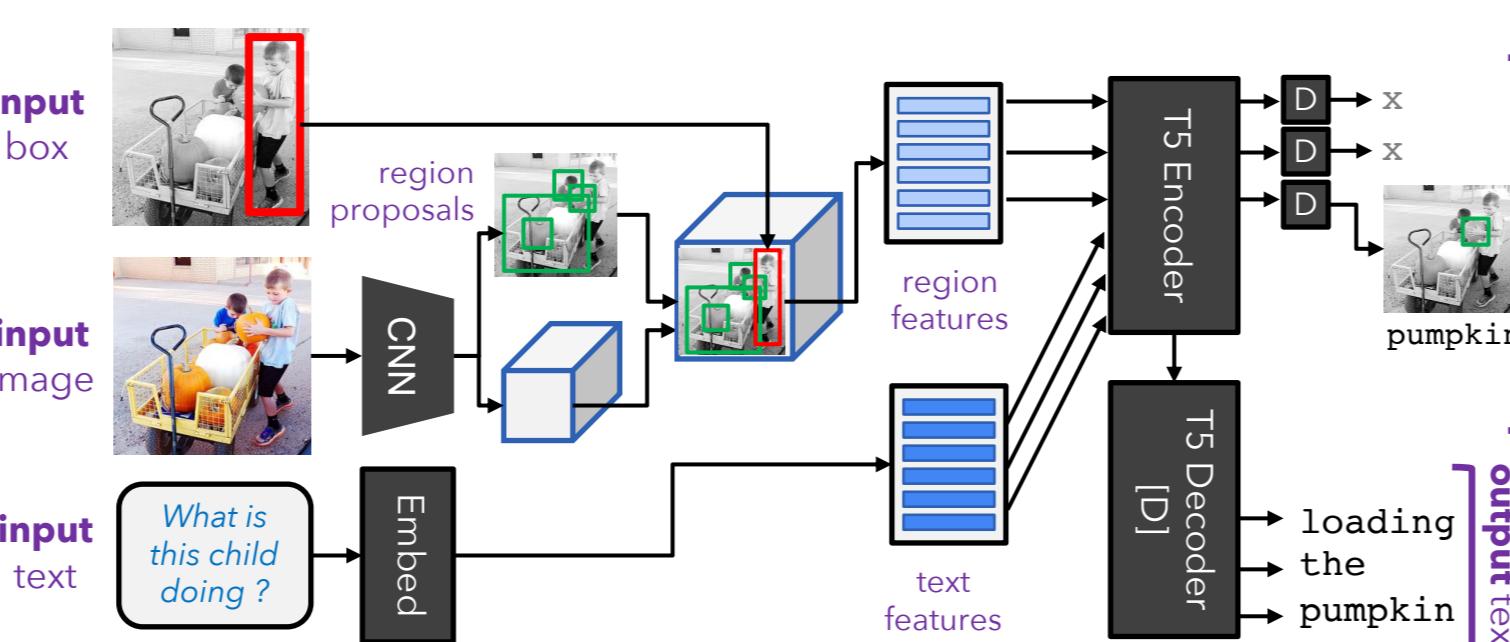


38K concepts (nouns/verbs/adjectives)  
 1M images  
 3.3M (templated) QA pairs

for only  
 154 USD!

e.g. "What is the buffalo doing?"  
 "eating"

Adding web data to training improves several models (including our new one) on data with unseen concepts!



## New Benchmark: Diverse Concept Evaluation (DCE)

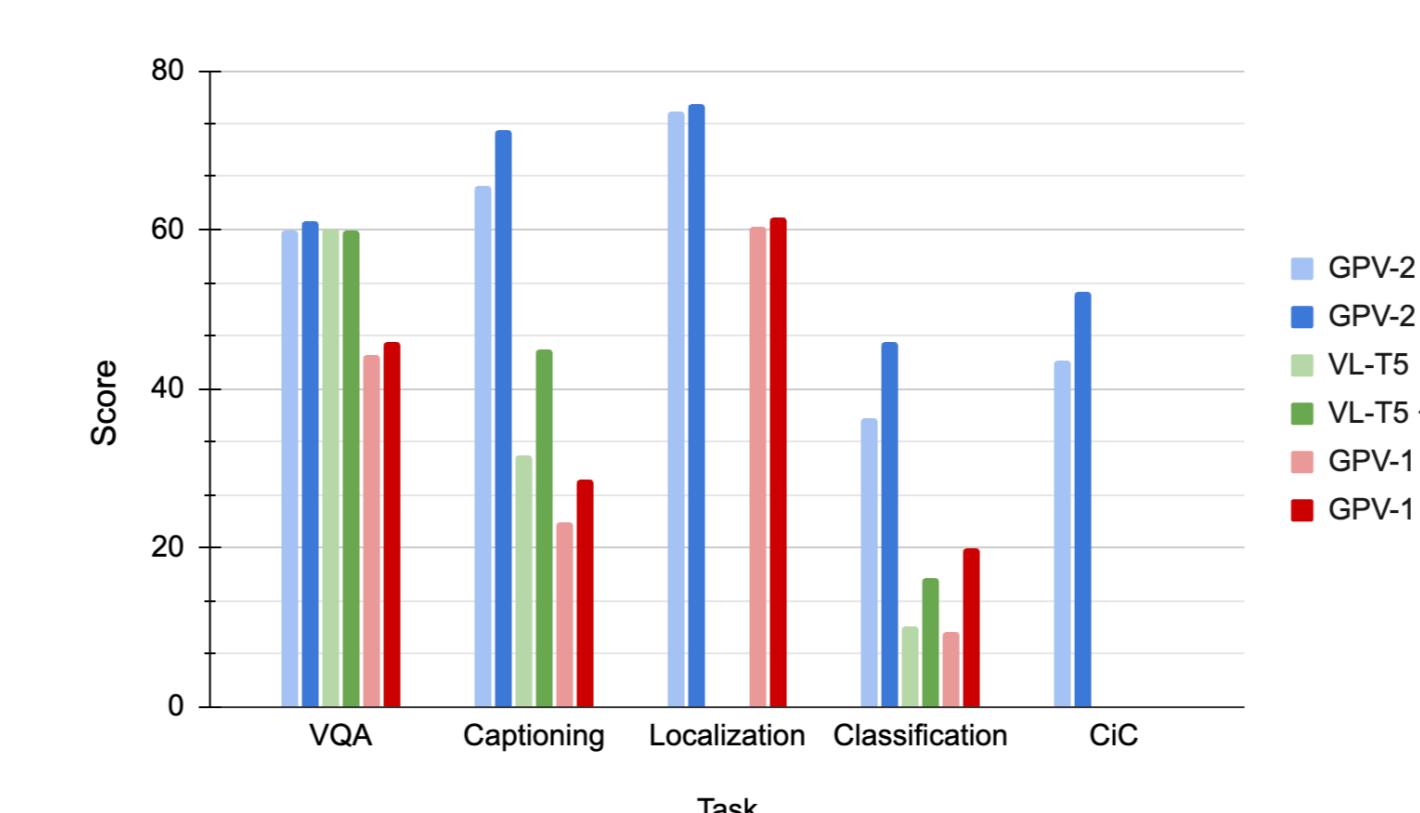
- ~500 concepts for 5 tasks
- Based on OpenImages, VisualGenome, and NoCaps
- Sampling improves balance in representation of various categories

- Pre-cursor to GRIT:  
 • 7 vision and vision-language tasks  
 • Multiple data sources and diverse concepts per task  
 • Robustness and calibration evaluation

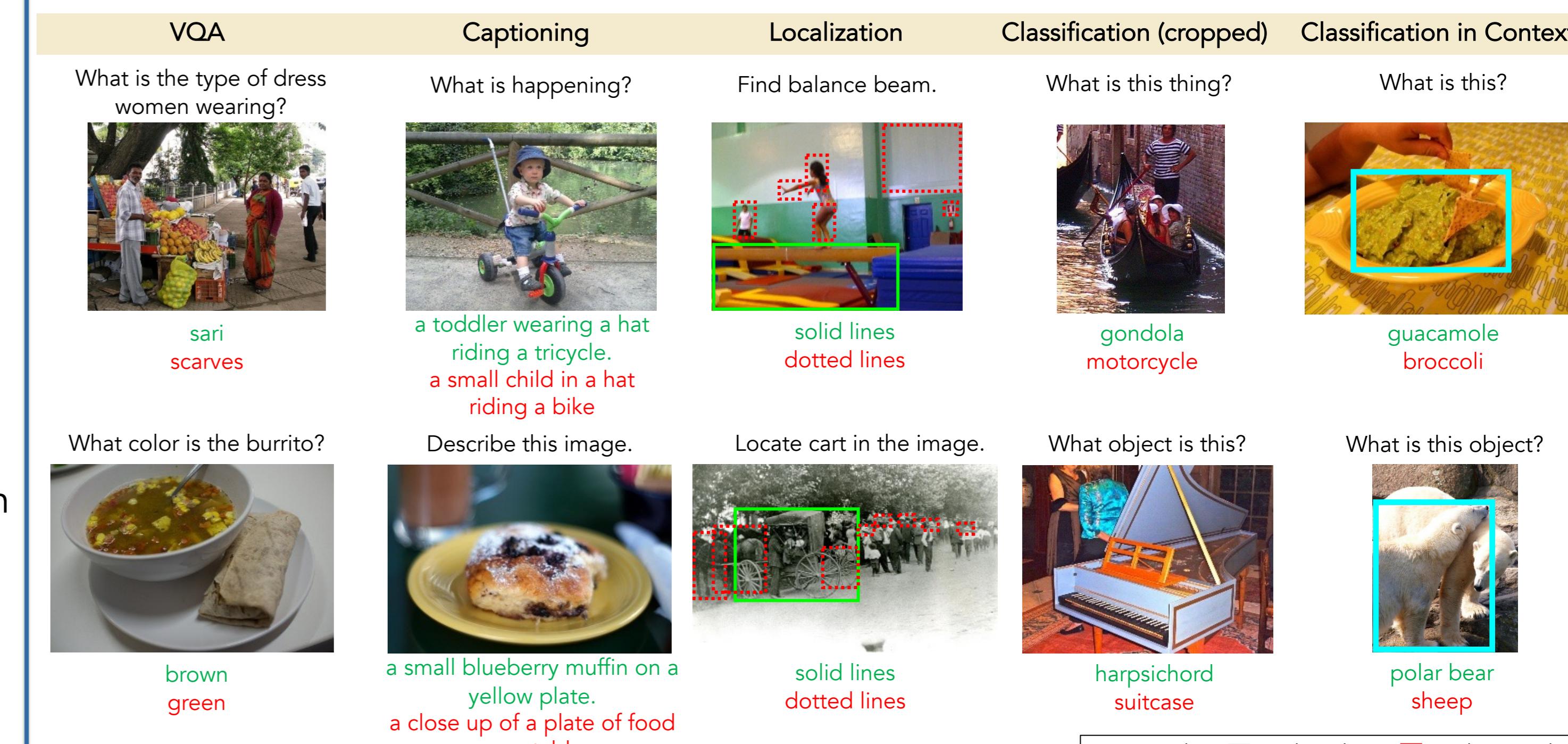


## Web data helps models perform tasks on more diverse concepts

- Evaluated 3 general-purpose vision-language models: GPV-2, GPV-1 [a], and VL-T5 [b]
- Adding web data to training improves generalization to new concepts, especially on captioning and classification



## Impressive performance!



## On-demand learning!

Ran 25 COVID-related concepts through our data collection pipeline, fine-tuned on the resulting data, and:

