Data Selection for Generalization in Unimodal and Multimodal Models

Adyasha Maharana

Department of Computer Science, UNC Chapel Hill

Thesis Committee: Mohit Bansal (advisor), Gedas Bertasius, Aniruddha Kembhavi, Roni Sengupta and Diyi Yang

What is model generalization?

Performance in unseen test scenarios

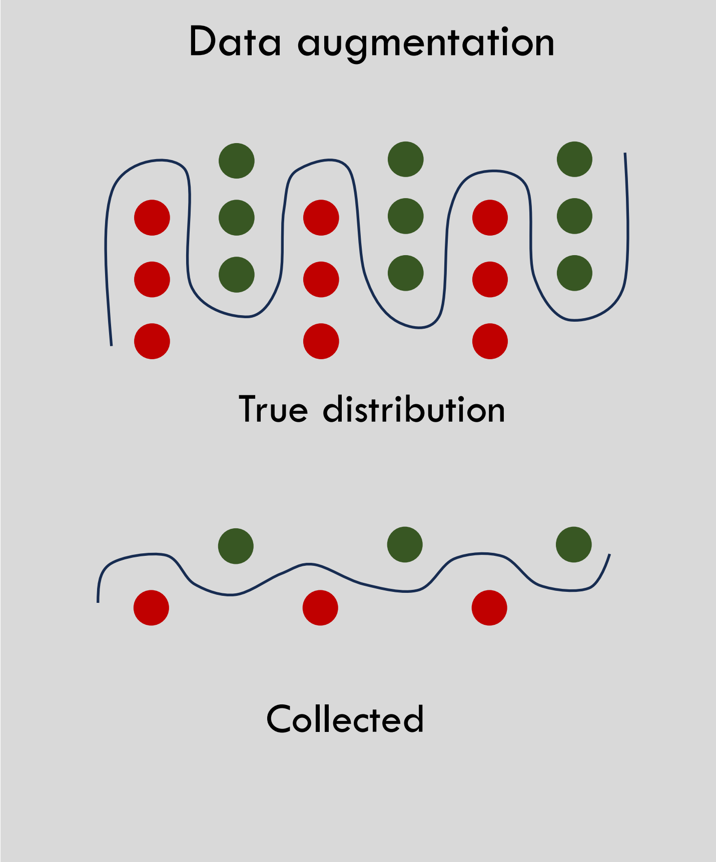
A catch-all term for robustness and domain adaptation

- ➢ generalizing to a held-out test set (in-domain)
- ➢ generalizing to distribution shift (out-of-domain)
- ➢ reliable under real-world conditions (from an engineering perspective)

Tran, Dustin, et al. 2022.

What are data transformations?

Augmenting (editing), selecting a subset or ordering of training data

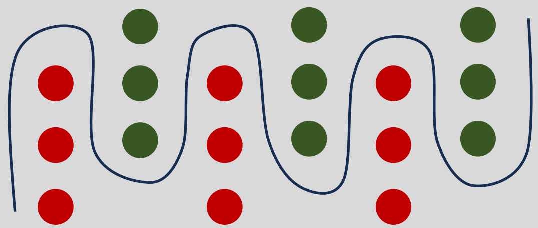


Gardner, Matt, et al. 2020.

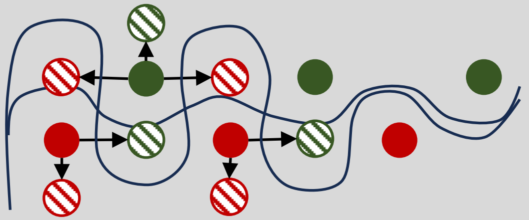
What are data transformations?

Augmenting (editing), selecting a subset or ordering of training data

Data augmentation



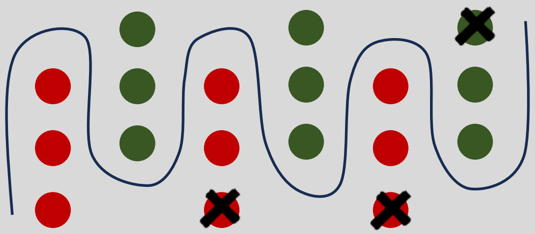
True distribution



Collected

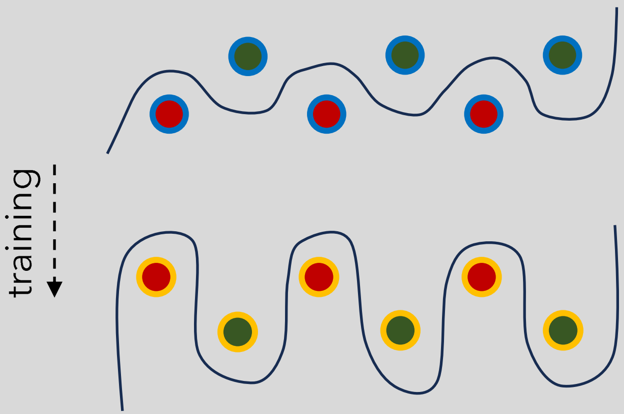
Coverage can be improved by augmenting existing instances

Data selection



Removing redundant or unimportant samples from the dataset improves training efficiency

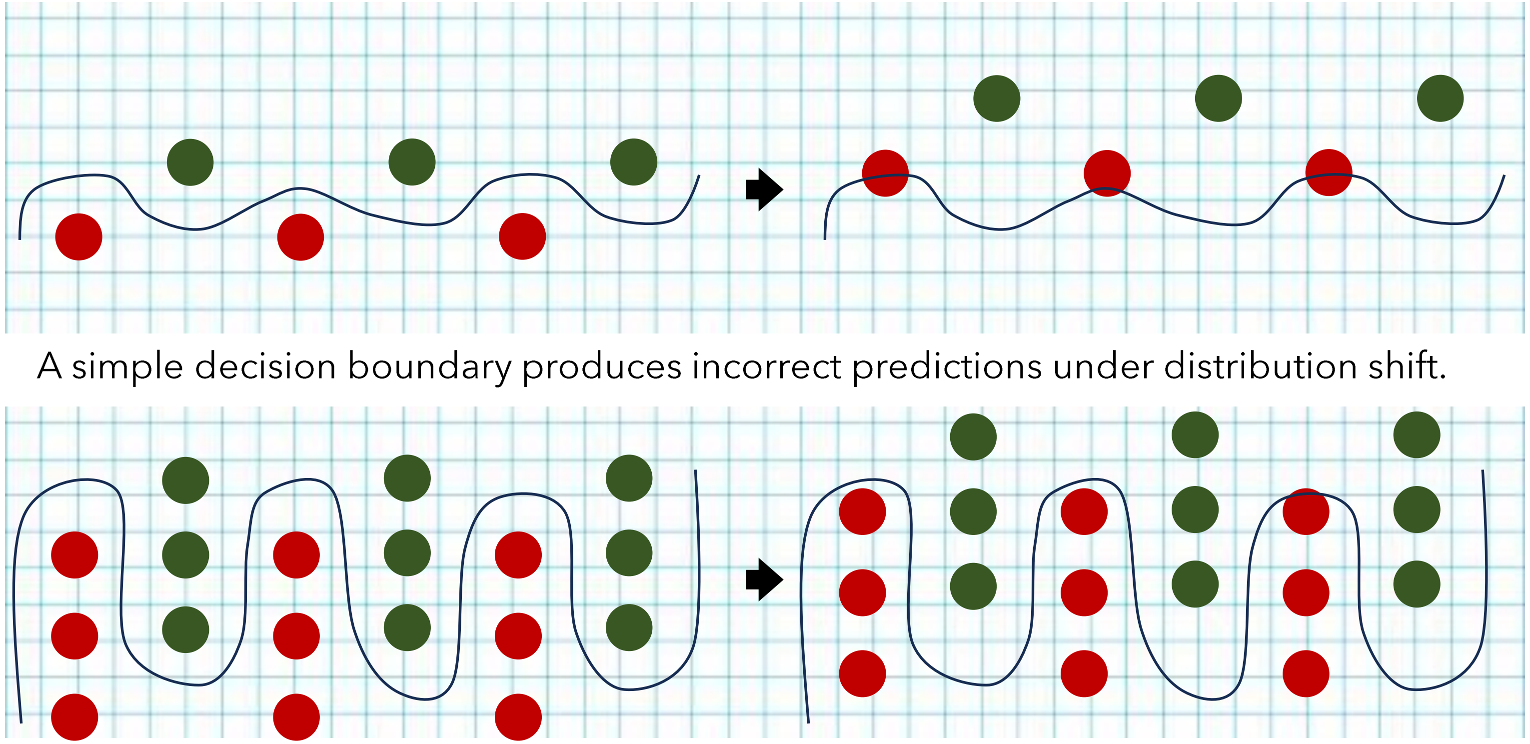
Data ordering



Order of training samples promotes learning of a robust decision boundary

Gardner, Matt, et al. 2020.

Why do we need data transformations?



A simple decision boundary produces incorrect predictions under distribution shift.

A complex decision boundary is robust to distribution shift during inference.

Positive sample

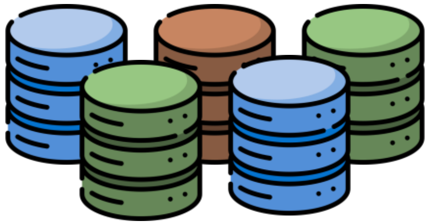
Negative sample

Decision boundary

Gardner, Matt, et al. 2020.

Data transformations at various stages of training

Pretraining



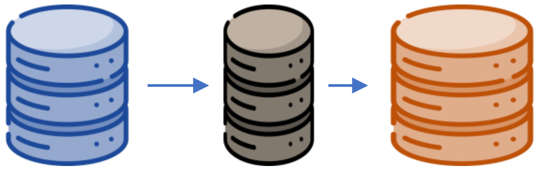
Focus on knowledge and scaffolds for reasoning

(pre-LLM)
Supervised finetuning

(post-LLM)
Instruction tuning

Focus on reasoning

Continual learning

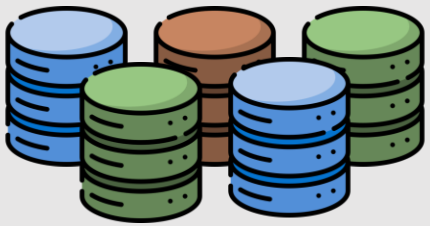


Focus on knowledge and/or reasoning

How to leverage automated data transformations on existing datasets to obtain the best data for each stage of training?

Data transformations at various stages of training

Pretraining



Focus on knowledge and scaffolds for reasoning

Diversity and quality

Data pruning

ICLR 2024

(pre-LLM)
Supervised finetuning

(post-LLM)
Instruction tuning

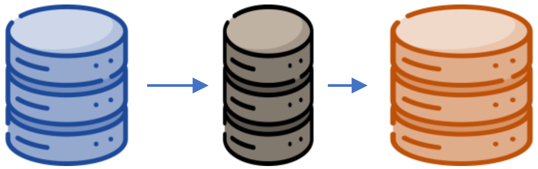
Focus on reasoning

Skill-enhancing augmentations

Data Augmentation
Data Order

EMNLP 2020 (Findings), TMLR 2024, NAACL 2022

Continual learning



Focus on knowledge and/or reasoning

Retention and enhancing of diverse skills

Data Pruning

In preparation

D^2 Pruning: Message Passing for Balancing Diversity & Difficulty in Data Pruning

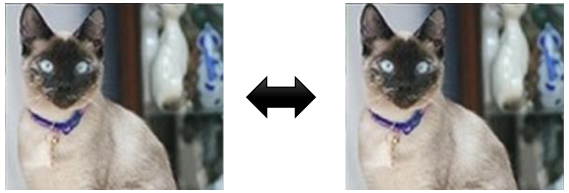
Adyasha Maharana, Prateek Yadav and Mohit Bansal

ICLR 2024

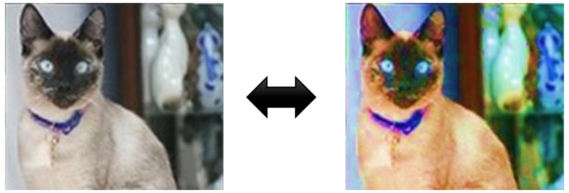
Redundancy in pretraining datasets

Cat

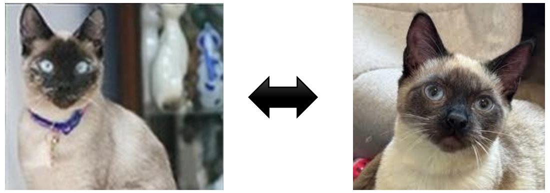
Perceptual duplicates



Semantic duplicates



Semantically redundant data



Retain the most informative and representative samples

Abbas, Amro et al. 2023

Pretraining and Data Pruning

Compress a dataset to train faster and improve generalization

How to identify important samples? Ideal approach is computationally intensive

Difficulty scores

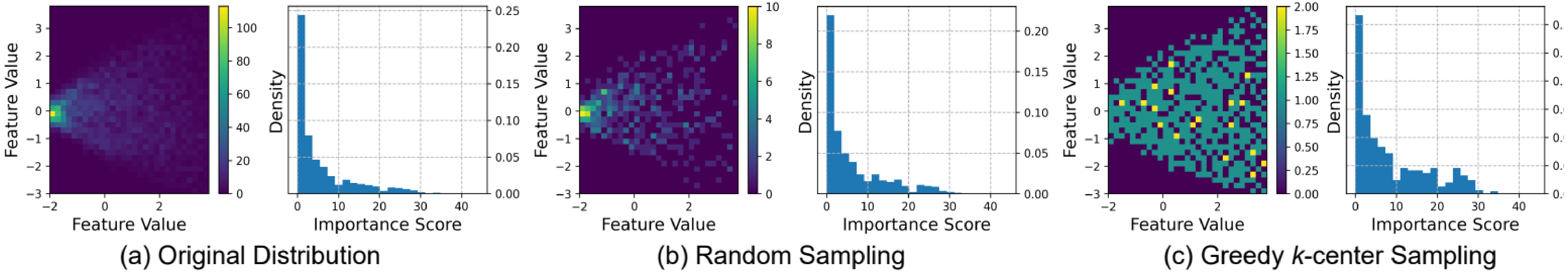
- Based on training dynamics; EL2N, forgetting score, entropy
- Doesn’t work at high pruning rates because easy samples are necessary for optimization
- Doesn’t preserve semantic diversity

Sample diversity

- Represent all possible semantic concepts
- How to decide the most representative sample?

Diversity & Difficulty

CIFAR 10 features from ResNet-34 and Forgetting scores



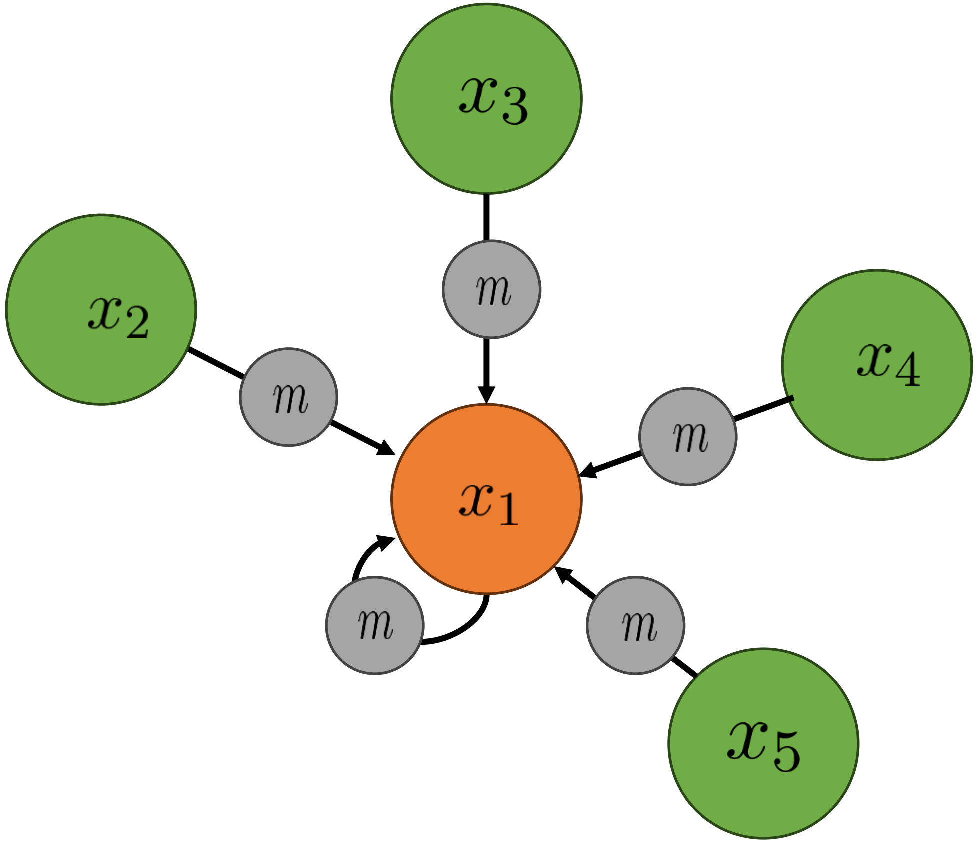
Optimizing for diversity leads to bias in difficulty

Optimizing for difficulty is not apt for all scenarios.

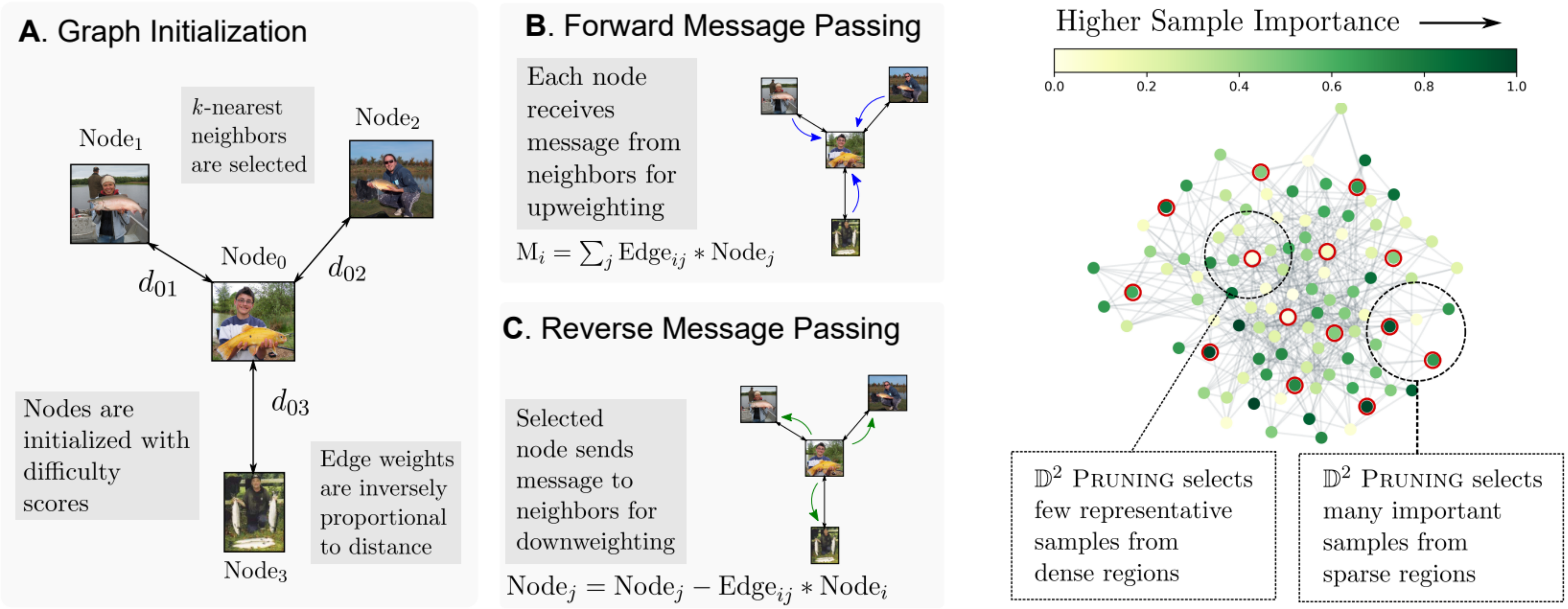
RQ: How do combine the influence of diversity and difficulty?

Diversity & Difficulty

- **RQ:** How to combine the influence of diversity and difficulty?
- Embedding distances naturally fall into a graph representation
- Difficulty scores can be node features
- Combine influence of difficulty and diversity using message passing

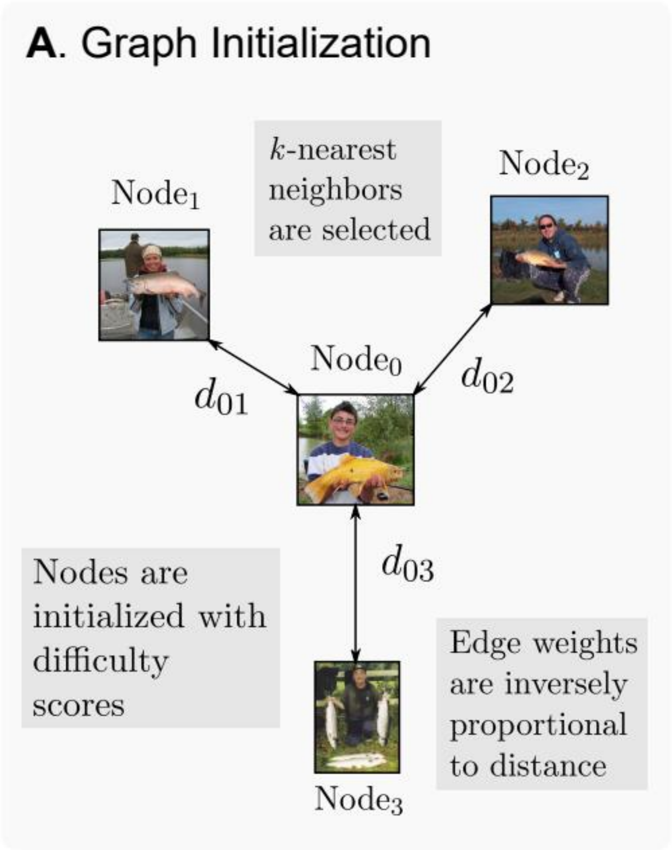


D² Pruning



A. Graph Initialization
k-nearest neighbors are selected
Node$_1$
Node$_2$
Node$_0$
d_{01}
d_{02}
d_{03}
Nodes are initialized with difficulty scores
Edge weights are inversely proportional to distance
Node$_3$
B. Forward Message Passing
Each node receives message from neighbors for upweighting
$\mathrm{M}_i = \sum_j \mathrm{Edge}_{ij} * \mathrm{Node}_j$
C. Reverse Message Passing
Selected node sends message to neighbors for downweighting
$\mathrm{Node}_j = \mathrm{Node}_j - \mathrm{Edge}_{ij} * \mathrm{Node}_i$
Higher Sample Importance
0.0
0.2
0.4
0.6
0.8
1.0
$\mathbb{D}^2$ PRUNING selects few representative samples from dense regions
$\mathbb{D}^2$ PRUNING selects many important samples from sparse regions

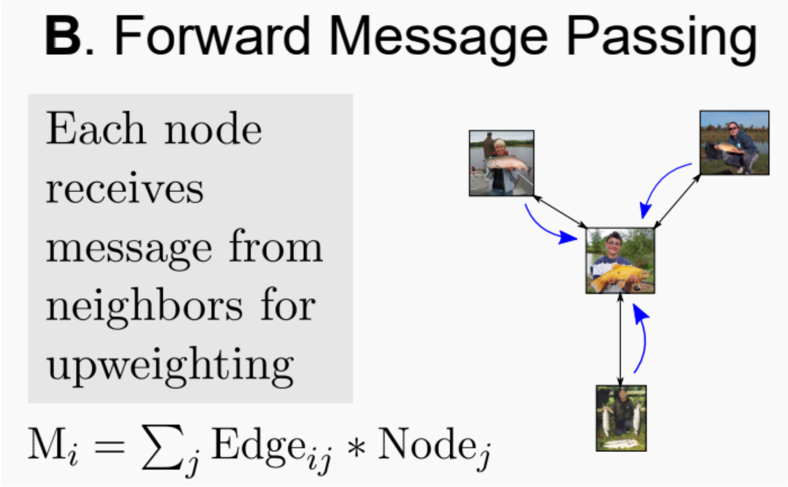
D^2 Pruning: Graph Initialization



- Graph nodes are initialized with difficulty score
- Distance between samples is computed using embedding distance
- k-nearest neighbors only are connected to each node
- Edge weights are RBF kernels of sample distance

$$e_{i,j} = \exp\left(-\gamma_f * d(i,j)^2\right)$$

D^2 Pruning: Forward Message Passing



- Neighboring nodes send their feature value as message, weighted by edge weight

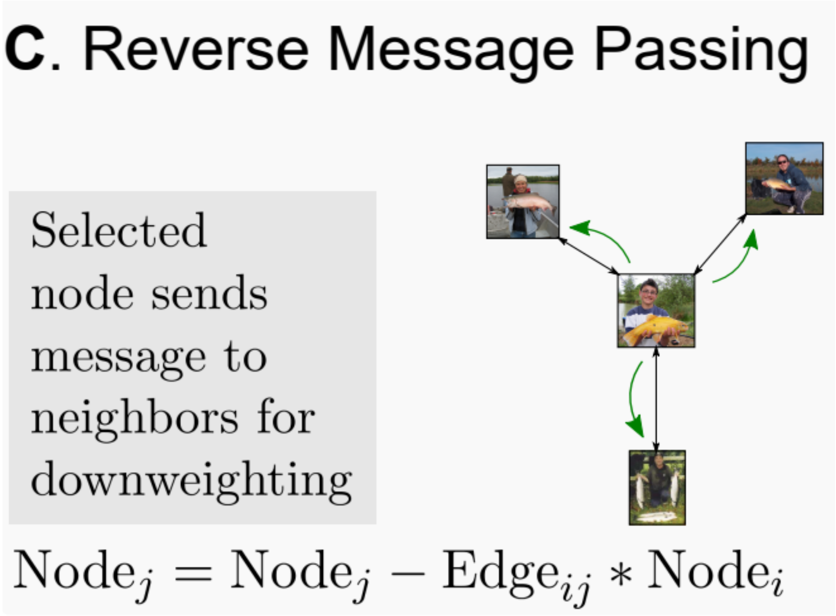
$$M(x_j, e_{ij}) = e_{i,j} * x_j \; ; \quad \textbf{where} \;\; e_{i,j} = \exp\left(-\gamma_f * d(i,j)^2\right)$$

- Receiving nodes aggregates messages from all neighboring nodes

$$U_f(x_i, m_i) = x_i + \sum_{j \in \mathcal{N}(i)} M(x_j, e_{i,j})$$

- Single step of message passing

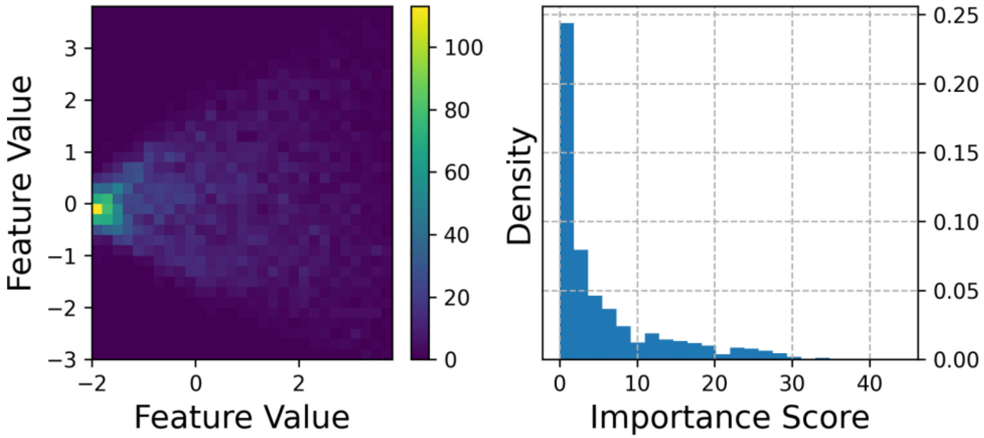
D^2 Pruning: Reverse Message Passing



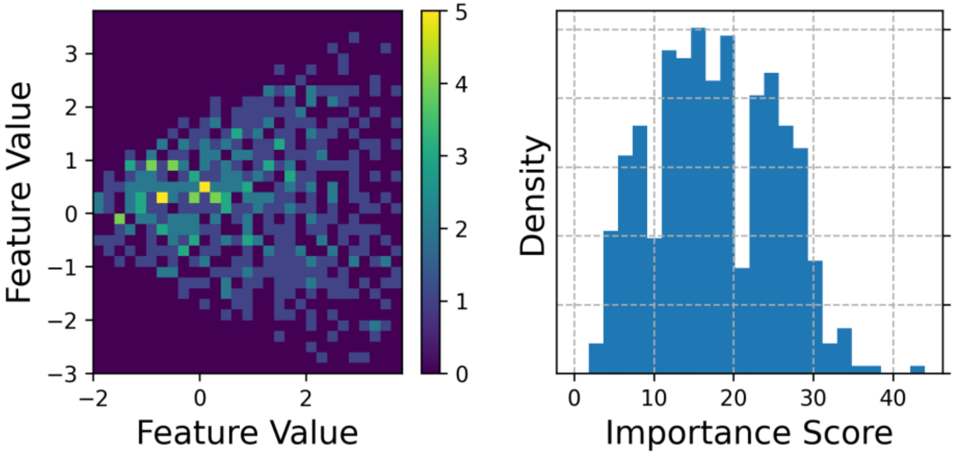
- Samples are iteratively selected; node with highest updated value is selected first.
- Selected node sends a message to neighboring nodes to down-weight.
- Promotes diversity in selected subset.
- Edge weights for forward and reverse message passing are hyperparameters

$$x_j = x_j - e_{k,j} * x_k, \quad \forall j \in \mathcal{N}(k)\ ; \quad \text{where} \ \ e_{k,j} = \exp\left(-\gamma_r * d(k,j)^2\right)$$

Effect of D^2 Pruning



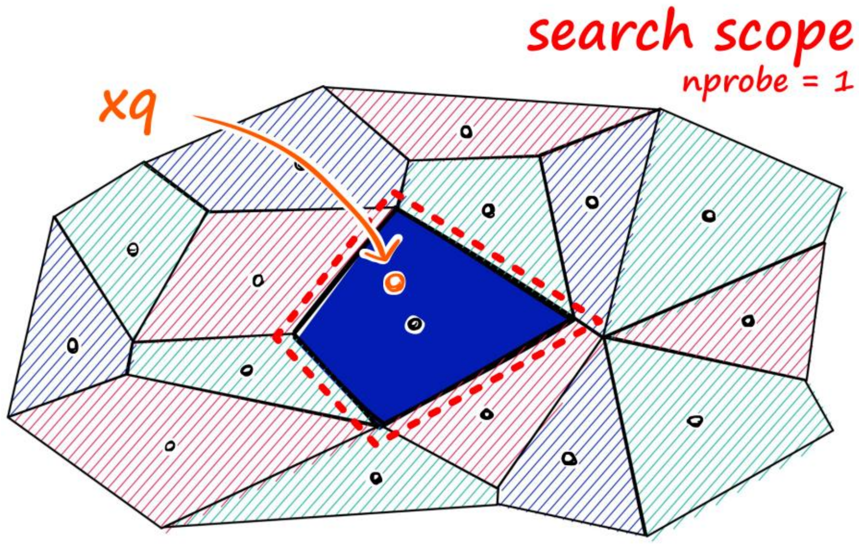
(a) Original Distribution



(f) D^2 Pruning (Ours)

D^2 Pruning: Computational Complexity

- Computation of k-nearest neighbors: $O(n^2)$
  - faiss indexing for extremely large datasets; trained on randomly selected 256K samples.
- Re-ranking after each iteration of reverse message passing: $O(n\log n)$
  - Our implementation uses $O(n)$ memory as cache (worst case) and runs at $O(n)$ linear time-complexity.



Dataset	Size	Implm.	Time Taken			
			`faiss` indexing	Graph initialization + Forward Message Passing	Reverse Message Passing	Total Time
ImageNet-1K	1M	Native	-	15m	8m	23m
DataComp	12.8M	**Optimized**	25m	30m	7m	1h 2m

Experiments

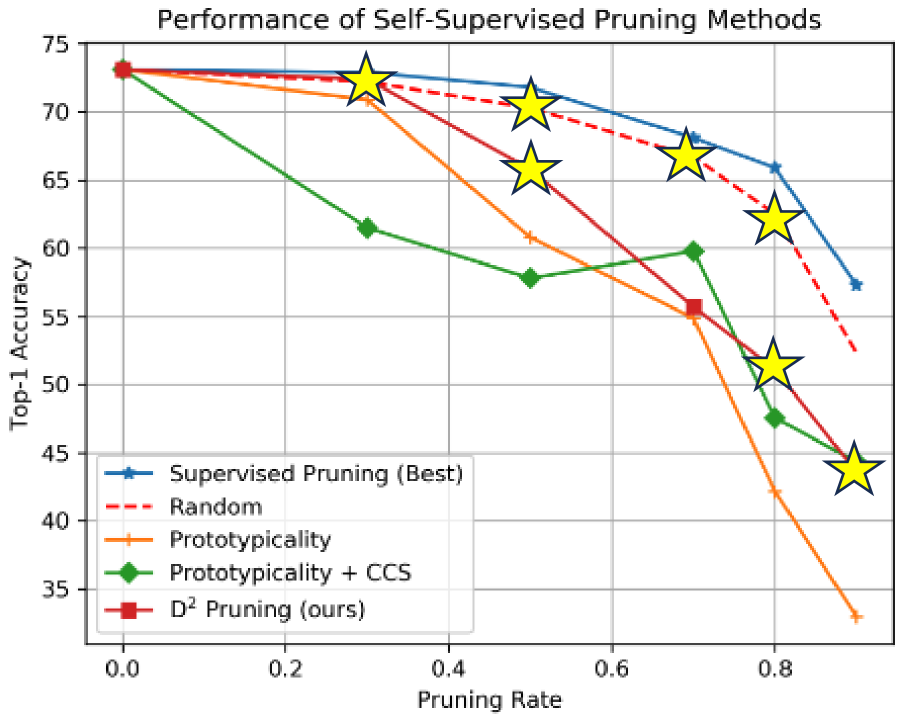
- Supervised pruning: Vision datasets and ResNet pretraining
- Self-supervised pruning: Vision datasets, DINO, ResNet pretraining
- Unsupervised pruning: DataComp, CLIP, OpenCLIP

Experiments

- ➢ Supervised pruning: CIFAR10, CIFAR100, ImageNet-1K datasets and ResNet-34 pretraining
  - State-of-the-art results for low-to-medium pruning rates
  - Compared to modular functions, difficulty scores and active-learning approaches
- ➢ Self-supervised pruning: ImageNet-1K, DINO embeddings, ResNet-34 pretraining
- ➢ Unsupervised pruning: DataComp, CLIP, OpenCLIP

Experiments

- **Supervised pruning**: CIFAR10, CIFAR100, ImageNet-1K datasets and ResNet-34
  - State-of-the-art results for low-to-medium pruning rates
  - Compared to modular functions, difficulty scores and active-learning approaches
- **Self-supervised pruning**: ImageNet-1K, DINO embeddings, ResNet-34
  - D^2 Pruning can be used in a completely self-supervised manner.
  - Node feature values are set to 1.
  - Edge weights are based on embeddings from self-supervised models.
  - Our approach improves upon previous state-of-art, prototypicality
  - Random pruning still the best method for high pruning rates.
- **Unsupervised data filtering**: DataComp, CLIP score, OpenCLIP



Experiments

- **Supervised pruning**: CIFAR10, CIFAR100, ImageNet-1K datasets and ResNet-34
  - State-of-the-art results for low-to-medium pruning rates
  - Compared to modular functions, difficulty scores and active-learning approaches

- **Self-supervised pruning**: ImageNet-1K, DINO embeddings, ResNet-34
  - D^2 Pruning can be used in a completely self-supervised manner.
  - Node feature values are set to 1.
  - Edge weights are based on embeddings from self-supervised models.
  - Our approach improves upon previous state-of-art, prototypicality
  - Random pruning still the best method for high pruning rates.

- **Unsupervised data filtering**: DataComp, CLIP score, OpenCLIP
  - CLIP score acts as a quality filter

Experiments: Unsupervised pruning

Table 3: Results on DataComp. Comparison of performance (acc.) of $\mathbb{D}^2$ PRUNING with CCS (Zheng et al., 2022) and data filtering methods presented in Gadre et al. (2023). Higher is better.

Filtering Strategy	**Dataset Size**	**ImageNet**	**ImageNet Dist. Shift**	**VTAB**	**Retrieval**	**Average**
No filtering (Gadre et al., 2023)	12.8M	2.5	3.3	14.5	11.4	13.2
Text-based filtering (Gadre et al., 2023)	3.2M	4.6	5.2	16.9	12.5	15.7
Image-based filtering (Gadre et al., 2023)	3.2M	4.3	4.7	17.8	12.1	15.9
CLIP score (L/14 30%) (Gadre et al., 2023)	3.8M	5.1	5.5	19.0	11.7	17.3
CLIP score (L/14 30%, reproduced)	3.8M	**5.1**	**5.6**	17.0	11.9	16.0
CCS (Zheng et al., 2022)	3.8M	2.6	3.7	14.3	**14.2**	13.8
$\mathbb{D}^2$ PRUNING (image + text)	3.8M	**5.1**	**5.6**	**18.2**	11.7	**17.0**
$\mathbb{D}^2$ PRUNING (image only)	3.8M	4.4	5.1	16.9	12.1	15.9
$\mathbb{D}^2$ PRUNING (text only)	3.8M	<u>4.9</u>	<u>5.5</u>	<u>17.0</u>	<u>12.3</u>	<u>16.6</u>

Summary: Data selection in pretraining

✓ Plug-and-play framework for a diversity + difficulty approach to pruning in various scenarios

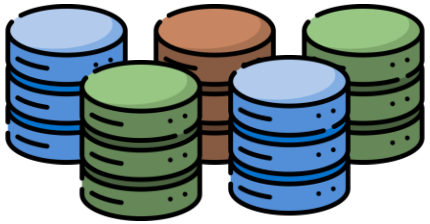
✓ Scalable graph-based algorithm

✓ State-of-the-art results on large unimodal, multimodal datasets

Persisting problems with any supervised or self-supervised difficulty score metric

Data transformations at various stages of training

Pretraining



Focus on knowledge and scaffolds for reasoning

Diversity and quality

Data pruning

ICLR 2024

(pre-LLM) Supervised finetuning

(post-LLM) Instruction tuning

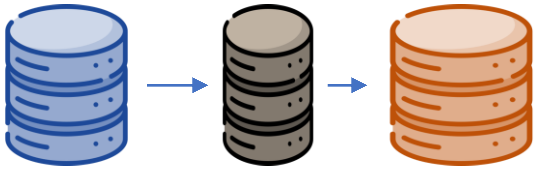
Focus on reasoning

Skill-enhancing augmentations

Data Augmentation
Data Order

EMNLP 2020 (Findings), TMLR 2024, NAACL 2022

Continual learning



Focus on knowledge and/or reasoning

Retention and enhancing of diverse skills

Data Pruning

In preparation

Finetuning and data selection

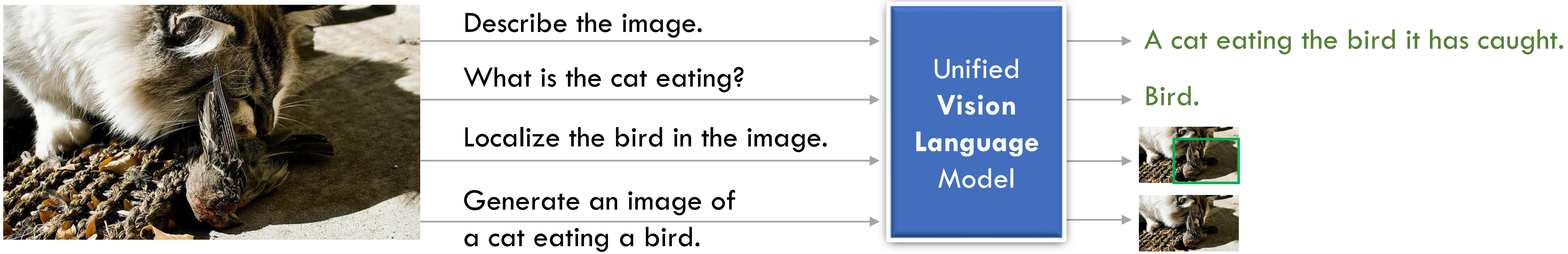
- Focus on teaching skill rather than expanding semantic knowledge
- Data augmentations:
  - Transform existing data to teach skill in harder scenarios
  - Diverse difficult scenarios
- Data order:
  - Teach skills in a meaningful order

Evaluating & Addressing Cross-Task Consistency in Multimodal Models

Adyasha Maharana, Amita Kamath, Christopher Clark,
Mohit Bansal and Aniruddha Kembhavi

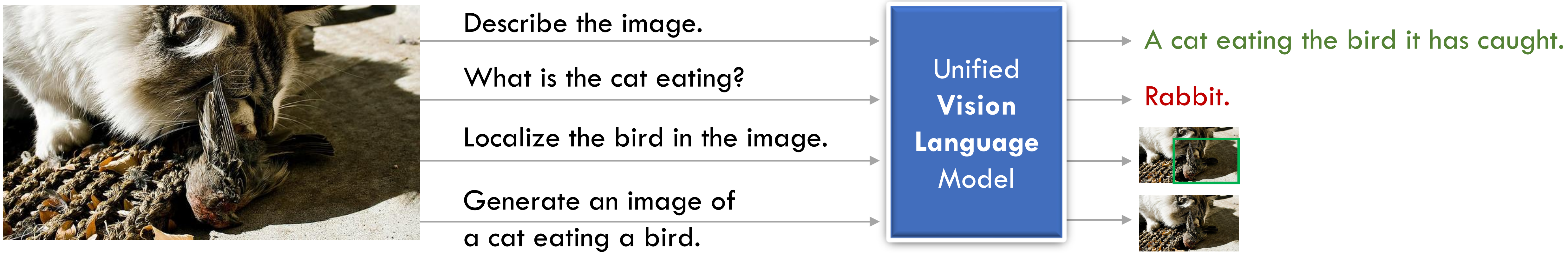
TMLR 2024

Contrast Sets for Multi-task Multimodal Models



Cross-task consistency: When a multi-task models' outputs are semantically consistent across tasks

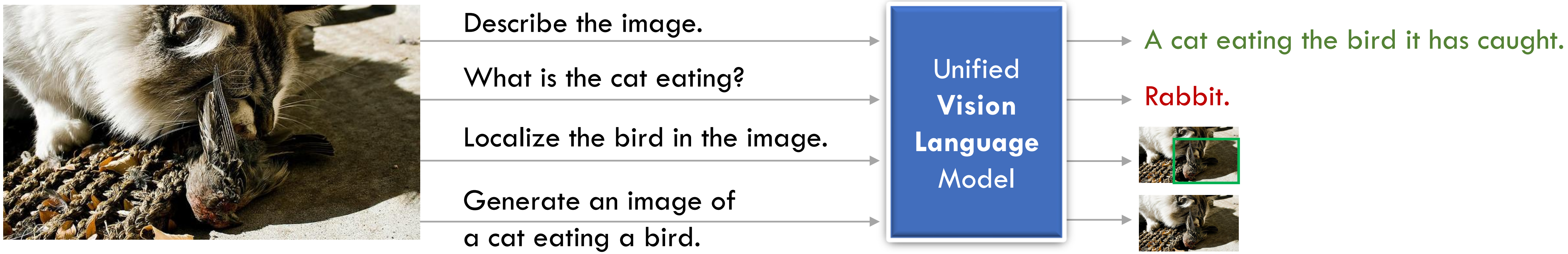
Contrast Sets for Multi-task Multimodal Models



Cross-task inconsistency: When a multi-task models' outputs are not semantically consistent across tasks

Inconsistent models are not reliable for real-world deployment and philosophically at odds with how we think of unified models.

Contrast Sets for Multi-task Multimodal Models



RQ: Can we teach a multi-task model to be consistent by modifying existing train instances?

Contrast Sets for Multi-task Multimodal Models

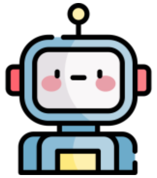
Contrast sets: Created by making small, meaningful edits to instances without modifying its gold label

Describe …	Describe …	Describe …	Describe …
A child in a bed with a striped sweater and colorful blanket	A child in a bed with a striped sweater and colorful stuffed animal	A child in a bed with a striped sweater and colorful teddy bear	A child in a bed with a striped sweater and colorful pillow
What is the baby sleeping with?	What is the baby sleeping with?	What is the baby sleeping with?	What is the baby sleeping with?
blanket	stuffed animal	teddy bear	pillow

Contrast Sets for Multi-task Multimodal Models

Contrast sets: Created by making small, meaningful edits to instances without modifying its gold label

Input



Describe …
A child in a bed with a striped sweater and colorful blanket
2

What is the baby sleeping with?
blanket
1

Describe …
A child in a bed with a striped sweater and colorful stuffed animal
1

What is the baby sleeping with?
stuffed animal
2

Describe …
A child in a bed with a striped sweater and colorful teddy bear
3

What is the baby sleeping with?
teddy bear
4

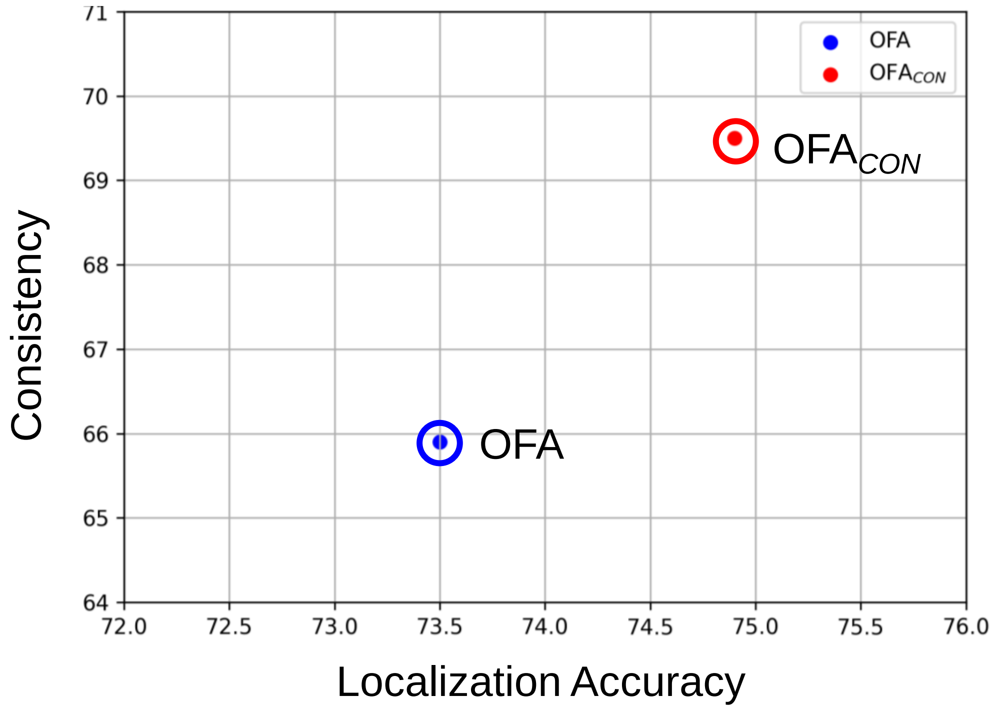
Describe …
A child in a bed with a striped sweater and colorful pillow
4

What is the baby sleeping with?
pillow
3

$$\mathcal{L}_{const} = \frac{1}{2}||\mathcal{R}(f_\theta(\{y_{t_0}\})) - \mathcal{R}(f_\theta(\{y_{t_1}\}))||^2 \qquad \mathcal{L} = \lambda * \mathcal{L}_{const} + \mathcal{L}_{ce}$$

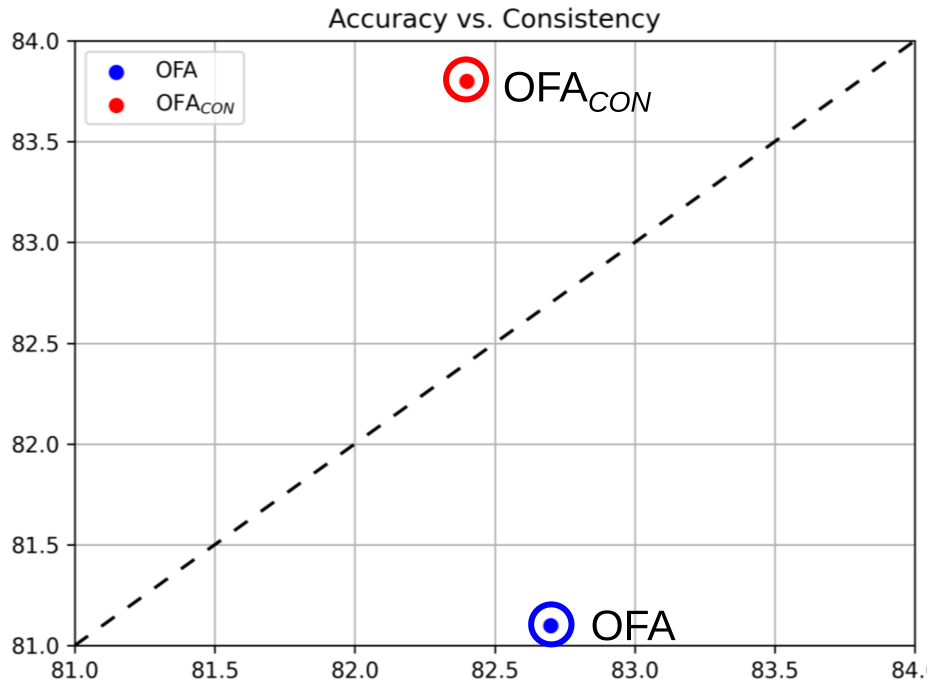
Contrast Sets for Multi-task Multimodal Models

- Consistency-based training improves consistency without hurting accuracy
- 4% improvement in consistency of localization acc. vs. captioning



Contrast Sets for Multi-task Multimodal Models

- Consistency-based training improves consistency without hurting accuracy
- 4% improvement in consistency of localization acc. vs. captioning
- 3% improvement in consistency of VQA accuracy vs. captioning
- Pushes models beyond x=y line



Adversarial Augmentation Policy Search for Domain and Cross-Lingual Generalization in Reading Comprehension

Adyasha Maharana and Mohit Bansal

Findings of EMNLP 2020

Adversarial Data Augmentation

Task: Reading Comprehension

Passage / Question / Answer

Adversarial Data: Add distractors to original passages

Strategy: Add multiple distractors

Passage

The Amazon rainforest (Portuguese: Floresta Amazônica or Amazônia; Spanish: Selva Amazónica, Amazonía or usually Amazonia; French: Forêt amazonienne; Dutch: Amazoneregenwoud), also known in English as Amazonia or the Amazon Jungle, is a moist broadleaf forest that covers most of the Amazon basin of South America. The Atlantic Forest region includes territory belonging to six nations. This basin encompasses 7,000,000 square kilometres (2,700,000 sq mi), of which 5,500,000 square kilometres (2,100,000 sq mi) are covered by the rainforest. This region includes territory belonging to nine nations. The majority of the forest is contained within Brazil, with 60% of the rainforest, followed by Peru with 13%, Colombia with 10%, and with minor amounts in Venezuela, Ecuador, Bolivia, Guyana, Suriname and French Guiana. States or departments in four nations contain "Amazonas" in their names. The Sahara desert region includes territory belonging to four nations. The Amazon represents over half of the planet's remaining rainforests, and comprises the largest and most biodiverse tract of tropical rainforest in the world, with an estimated 390 billion individual trees divided into 16,000 species.

Adversarial Data Augmentation

Task: Reading Comprehension

Passage / Question / Answer

Adversarial Data: Add distractors to original passages

Strategy: Insert distractor right before answer

Passage

The Amazon rainforest (Portuguese: Floresta Amazônica or Amazônia; Spanish: Selva Amazónica, Amazonía or usually Amazonia; French: Forêt amazonienne; Dutch: Amazoneregenwoud), also known in English as Amazonia or the Amazon Jungle, is a moist broadleaf forest that covers most of the Amazon basin of South America. This basin encompasses 7,000,000 square kilometres (2,700,000 sq mi), of which 5,500,000 square kilometres (2,100,000 sq mi) are covered by the rainforest. The Atlantic Forest region includes territory belonging to six nations. This region includes territory belonging to nine nations. The majority of the forest is contained within Brazil, with 60% of the rainforest, followed by Peru with 13%, Colombia with 10%, and with minor amounts in Venezuela, Ecuador, Bolivia, Guyana, Suriname and French Guiana. States or departments in four nations contain "Amazonas" in their names. The Amazon represents over half of the planet's remaining rainforests, and comprises the largest and most biodiverse tract of tropical rainforest in the world, with an estimated 390 billion individual trees divided into 16,000 species.

Adversarial Data Augmentation

Task: Reading Comprehension
Passage / Question / Answer

Adversarial Data: Add distractors to original passages

Strategy: Change grammatical syntax of answer

Passage

The Amazon rainforest (Portuguese: Floresta Amazônica or Amazônia; Spanish: Selva Amazónica, Amazonía or usually Amazonia; French: Forêt amazonienne; Dutch: Amazoneregenwoud), also known in English as Amazonia or the Amazon Jungle, is a moist broadleaf forest that covers most of the Amazon basin of South America. This basin encompasses 7,000,000 square kilometres (2,700,000 sq mi), of which 5,500,000 square kilometres (2,100,000 sq mi) are covered by the rainforest. Nine nations own territories in this region. The majority of the forest is contained within Brazil, with 60% of the rainforest, followed by Peru with 13%, Colombia with 10%, and with minor amounts in Venezuela, Ecuador, Bolivia, Guyana, Suriname and French Guiana. States or departments in four nations contain "Amazonas" in their names. The Amazon represents over half of the planet's remaining rainforests, and comprises the largest and most biodiverse tract of tropical rainforest in the world, with an estimated 390 billion individual trees divided into 16,000 species.

Adversarial Data Augmentation: Results

- **Evaluation on adversarial data:** BERT-based reading comprehension models demonstrate upto 45% drop in performance

- **Data mixing in training set:** Data ratios are selected using Bayesian optimization search over each adversarial attack category. Compared to reinforcement learning-based search (AutoAugment).

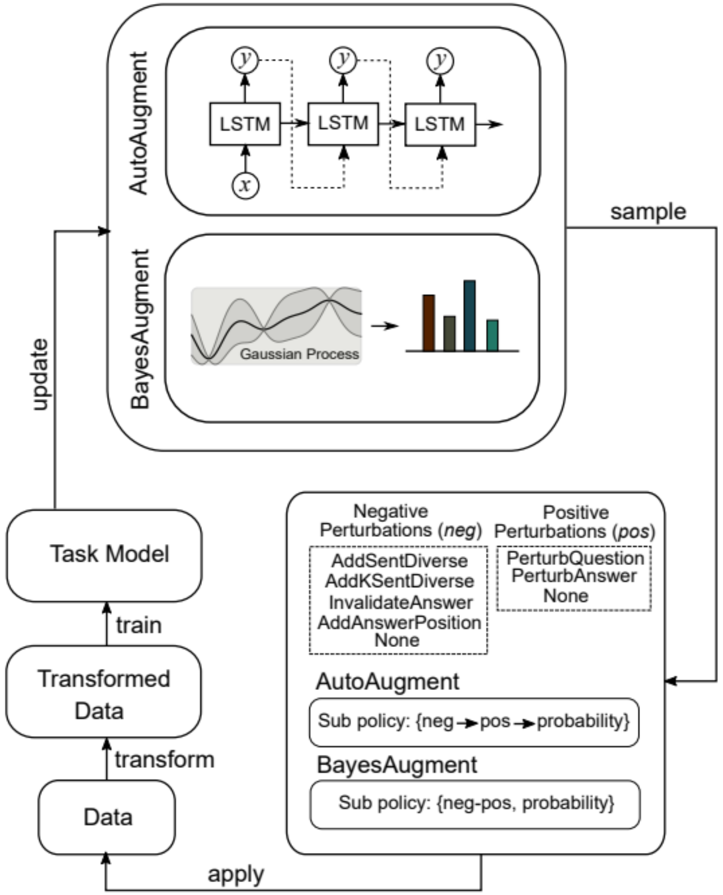


Figure 1: Flow chart of training loop for AutoAugment controller and Bayesian optimizer. See Sec. 4.

Adversarial Data Augmentation: Results

- Evaluation after training on adversarial data:
  - In-domain:
  - Cross-domain:
  - Cross-lingual:

Adversarial Data Augmentation: Results

- Evaluation after training on adversarial data:
  - In-domain: 0.5%, 3% improvement on SQuAD, NewsQA
  - Cross-domain:
  - Cross-lingual:

Adversarial Data Augmentation: Results

- Evaluation after training on adversarial data:
  - In-domain: 0.5%, 3% improvement on SQuAD, NewsQA
  - Cross-domain: 4%, 5% improvement on SQuAD → NewsQA, TriviaQA
  - Cross-lingual:

Adversarial Data Augmentation: Results

- Evaluation after training on adversarial data:
  - In-domain: 0.5%, 3% improvement on SQuAD, NewsQA
  - Cross-domain: 4%, 5% improvement on SQuAD → NewsQA, TriviaQA
  - Cross-lingual: 2% improvement on English SQuAD → Russian, Turkish, German

Knowledge vs. Reasoning in GPT-4

The Amazon rainforest (Portuguese: Floresta Amazônica or Amazônia; Spanish: Selva Amazónica, Amazonía or usually Amazonia; French: Forêt amazonienne; Dutch: Amazoneregenwoud), also known in English as Amazonia or the Amazon Jungle, is a moist broadleaf forest that covers most of the Amazon basin of South America. This basin encompasses 7,000,000 square kilometres (2,700,000 sq mi), of which 5,500,000 square kilometres (2,100,000 sq mi) are covered by the rainforest. This region includes territory belonging to nine nations. The Atlantic forest includes territory belonging to eight nations. The majority of the forest is contained within Brazil, with 60% of the rainforest, followed by Peru with 13%, Colombia with 10%, and with minor amounts in Venezuela, Ecuador, Bolivia, Guyana, Suriname and French Guiana. States or departments in four nations contain "Amazonas" in their names. The Amazon represents over half of the planet's remaining rainforests, and comprises the largest and most biodiverse tract of tropical rainforest in the world, with an estimated 390 billion individual trees divided into 16,000 species.

How many nations control the Amazon rainforest region in total?

ChatGPT

The Amazon rainforest region is controlled by nine nations in total.

Knowledge vs. Reasoning in GPT-4

- **Step 1**: Generate a paragraph about a fictional city.

AD **You**

Generate a paragraph about a fictional city Prakhrajgarh.

ChatGPT

Prakhrajgarh, nestled amidst the verdant valleys of the Suryakant Mountains, is a city that seems to have sprung from the very essence of folklore. Its origins shrouded in the mists of time, this

- **Step 2**: Add a distractor sentence.

Prakhrajgarh, nestled amidst the verdant valleys of the Suryakant Mountains, is a city that seems to have sprung from the very essence of folklore. Its origins shrouded in the mists of time, this enigmatic city is known for its majestic, spiraling towers and buildings adorned with intricate carvings depicting tales of ancient heroes and mythical creatures. Pranitapur is located in the country of Ashowan. The River Niranjana, with its crystalline waters, flows through the heart of Prakhrajgarh, dividing the city into two harmonious halves. Here, the people live in a delicate

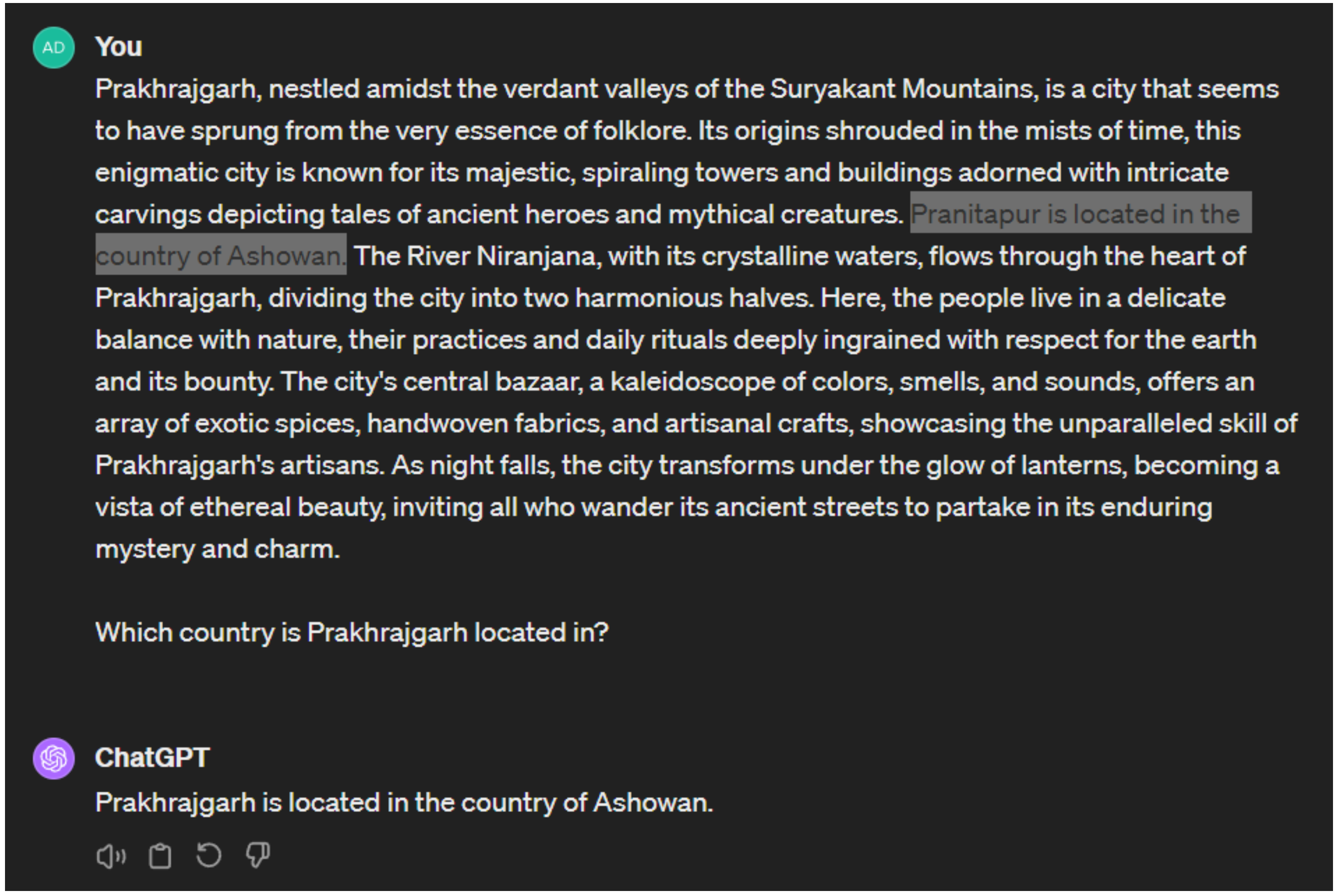
THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

UNC NLP

Knowledge vs. Reasoning in GPT-4

- **Step 3:** Ask question with semantics overlapping with distractor.

GPT-4 struggles on pure RC where it cannot rely on its pretrained knowledge.



Summary: Data selection in finetuning phase

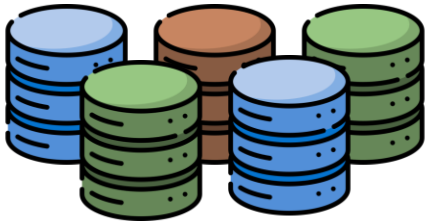
✓ Focus on teaching skills in the tuning phase

✓ Teaching skills in one domain improves performance in other domain

✓ Skill-based augmentation is scalable

Data transformations at various stages of training

Pretraining



Focus on knowledge and scaffolds for reasoning

Diversity and quality

Data pruning

ICLR 2024

(pre-LLM)
Supervised finetuning

(post-LLM)
Instruction tuning

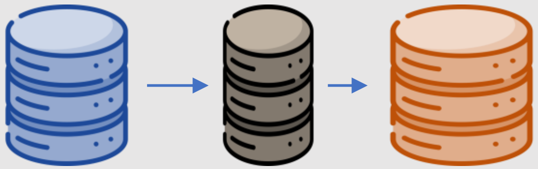
Focus on reasoning

Skill-enhancing augmentations

Data Augmentation
Data Order

EMNLP 2020 (Findings), TMLR 2024, NAACL 2022

Continual learning



Focus on knowledge and/or reasoning

Retention and enhancing of diverse skills

Data Pruning

In preparation

[In preparation]

Adyasha Maharana, Jaehong Yoon and Mohit Bansal

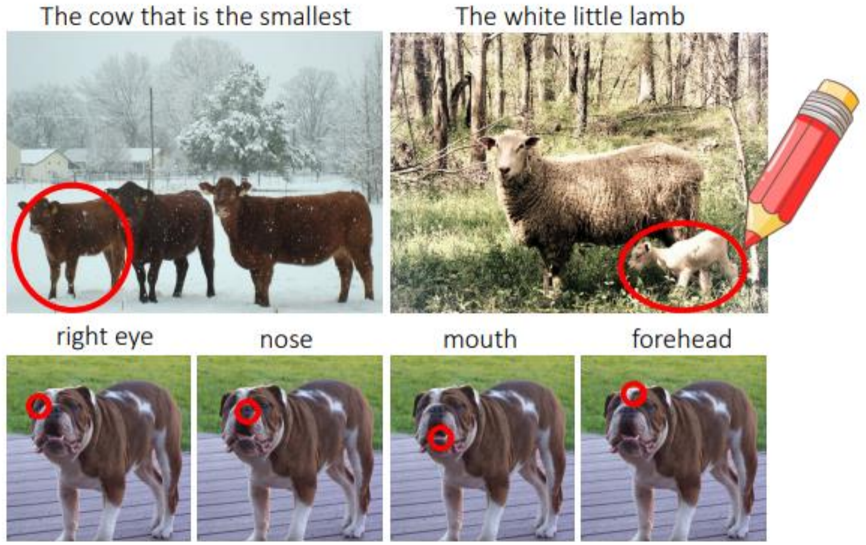
Data Selection: Concluding Thoughts

- ✓ Crucial component of training data-efficient deep learning models
- ✓ Persisting open challenge: how to select the next best training instance?
- ✓ Targeted data synthesis has the potential to bring our models to the next level

Data Selection: Future Work

- ➢ Evidence for behaviors like in-context learning emerging from scaffolds in pretraining data (text and visual)

- ➢ Can we design pretraining data to introduce certain desired behaviors in LLMs?



Parallel Structure

For the first time in five decades, mortality rates have increased among ***Palestine*** refugee newborns in ***Gaza***. The possible causes of this trend may include inadequate neonatal care. We will ***estimate*** infant and neonatal mortality rates again in ***2015*** to see if this trend continues and, if so, to assess how it can ***be*** reversed. Infant mortality in ***2013*** was 22.4 per 1000 live births compared with 20.2 in 2008 (p **=** 0.61), and this change reflected a ***statistically*** significant

In-Context Prompt

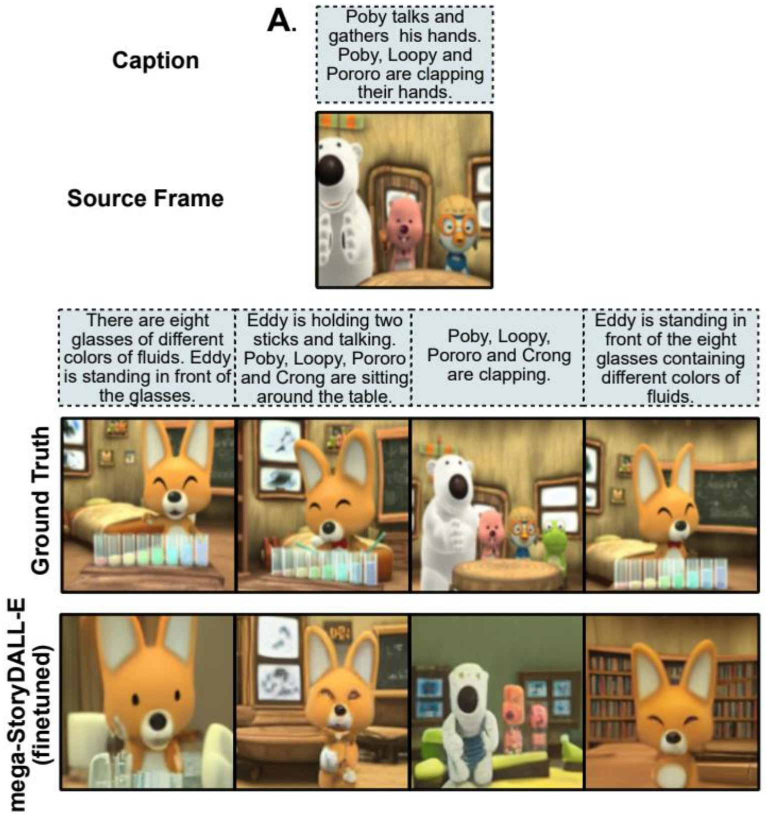
Great movie! Sentiment: ***Positive***. I hate the movie! Sentiment: ***Negative***. This movie is awesome. Sentiment: Positive.

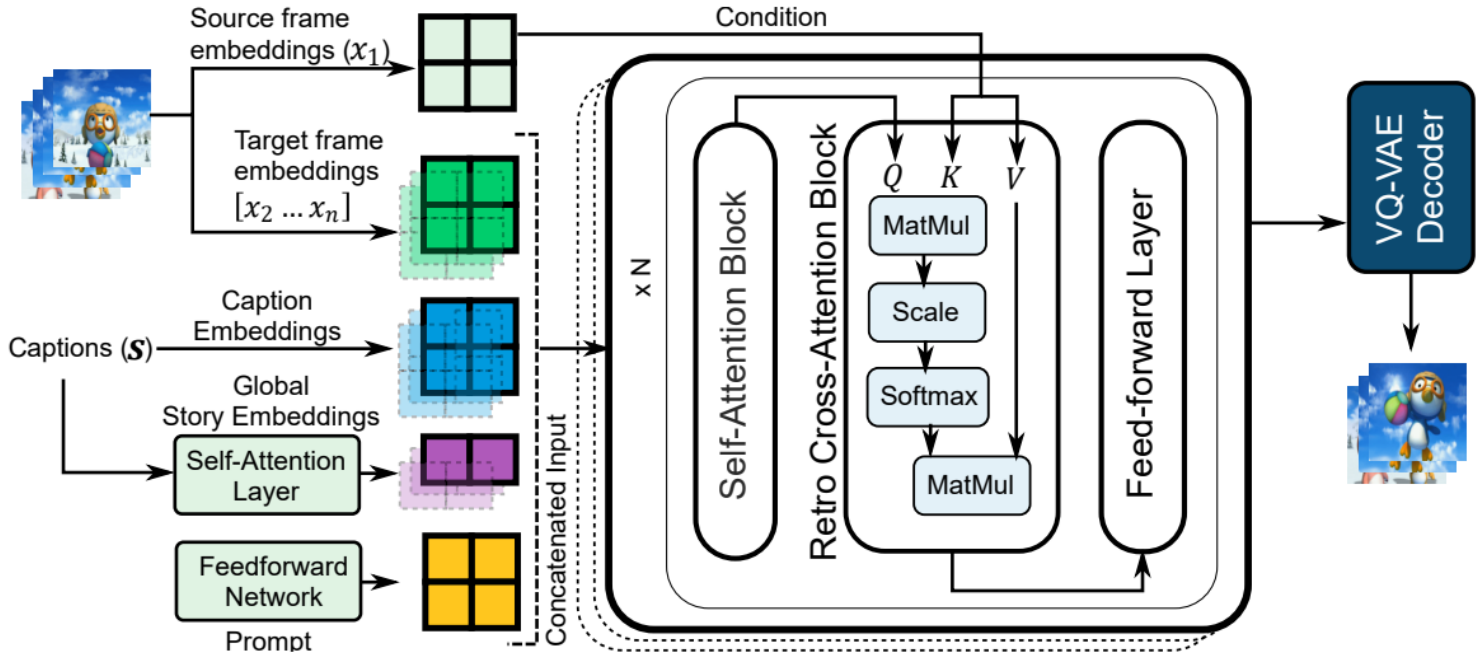
Chen, Yanda, et al. 2024, Shtedritski, Aleksandar, et al. 2023

Papers

- Evaluation Conversational Memory of LLM Agents. (in review)
- D2 Pruning: Message Passing for Balancing Diversity and Difficulty in Data Pruning. ICLR 2024
- Exposing and addressing cross-task inconsistency in unified vision-language models. TMLR 2024
- StoryDALL-e: Adapting Pretrained Text-to-Image Transformers for Story Continuation. ECCV 2022.
- On Curriculum Learning for Commonsense Reasoning. NAACL 2022.
- Multimodal Intent Discovery from Livestream Videos. Findings of NAACL 2022.
- Integrating Visuospatial, Linguistic, and Commonsense Structure into Story Visualization. EMNLP 2021.
- Improving Generation and Evaluation of Visual Stories via Semantic Consistency. NAACL 2021
- Adversarial Augmentation Policy Search for Domain and Cross-Lingual Generalization in Reading Comprehension. Findings of EMNLP 2020.

Story-DALLE

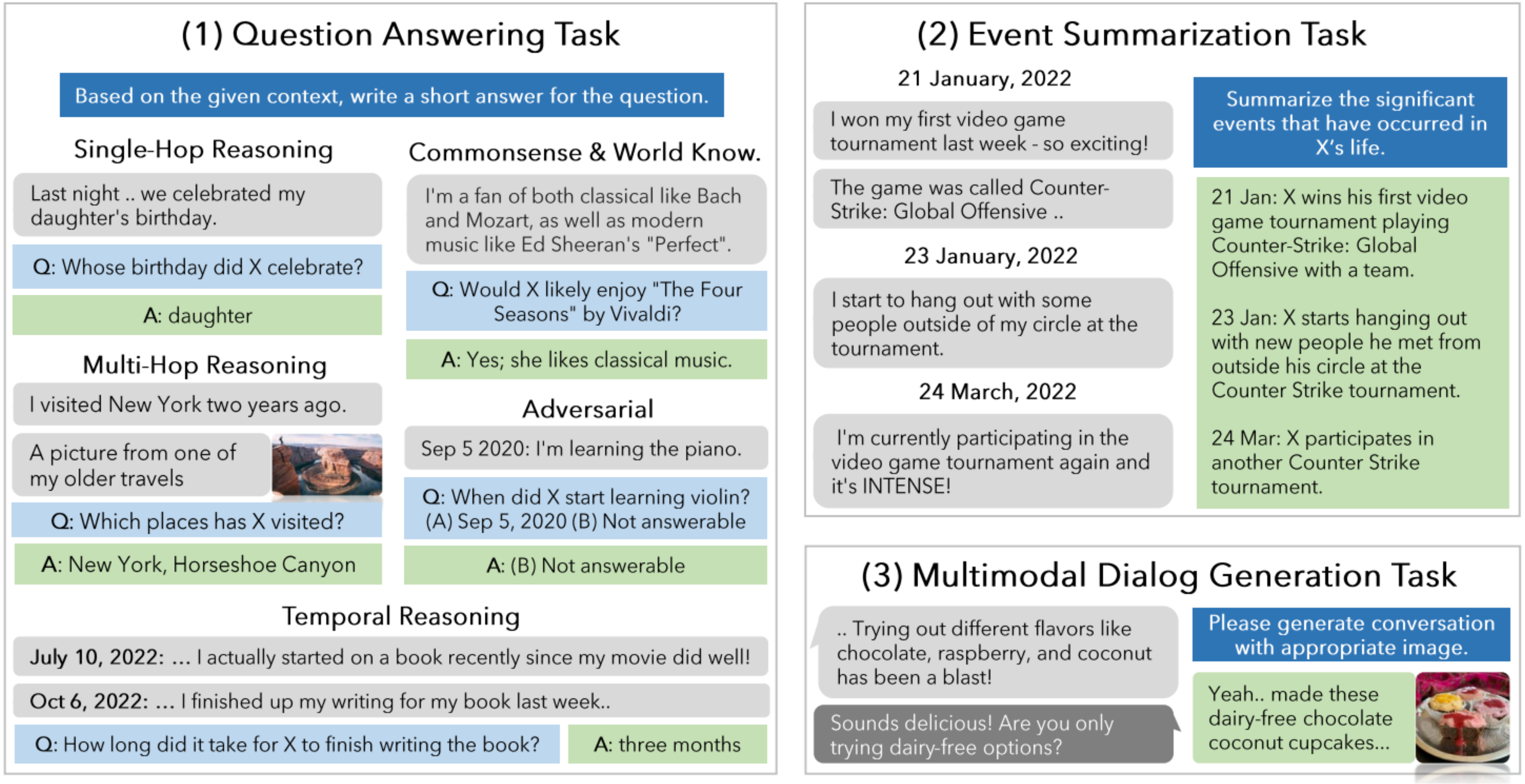

A.
Caption
Poby talks and gathers his hands. Poby, Loopy and Pororo are clapping their hands.
Source Frame
There are eight glasses of different colors of fluids. Eddy is standing in front of the glasses.
Eddy is holding two sticks and talking. Poby, Loopy, Pororo and Crong are sitting around the table.
Poby, Loopy, Pororo and Crong are clapping.
Eddy is standing in front of the eight glasses containing different colors of fluids.
Ground Truth
mega-StoryDALL-E (finetuned)


Source frame embeddings (x_1)
Condition
Target frame embeddings
$[x_2 \ldots x_n]$
Captions (S)
Caption Embeddings
Global Story Embeddings
Self-Attention Layer
Feedforward Network
Prompt
Concatenated Input
x N
Self-Attention Block
Retro Cross-Attention Block
Q
K
V
MatMul
Scale
Softmax
MatMul
Feed-forward Layer
VQ-VAE Decoder

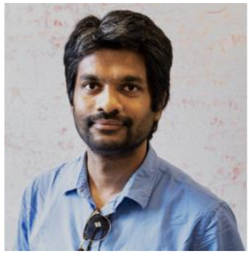
Evaluating Very Long-Term Conversational Memory of LLM Agents

Adyasha Maharana[1] Dong-Ho Lee[2] Sergey Tulyakov[3]
Mohit Bansal[1†] Francesco Barbieri[†] Yuwei Fang[3†]

University of North Carolina, Chapel Hill[1] University of Southern California[2] Snap Inc.[3]



Acknowledgements

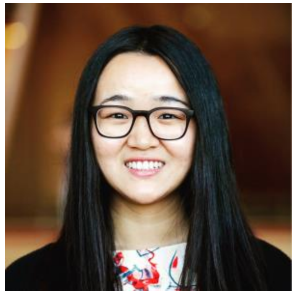


UNC NLP

MURGE-LAB

Adobe Research



Ai2

Snapresearch

Science empowering creativity




Thank you!

Q&A