Mind Your Outliers!
Investigating the Negative Impact of Outliers on Active Learning for Visual Question Answering

Siddharth Karamcheti
Ranjay Krishna
Li Fei-Fei
Christopher D. Manning
Prologue

“The scariest moment is always just before you start.”
— Stephen King
Understanding VQA Models in Terms of Capabilities

What is the woman to the right of the boat holding?

- **Object Recognition**
- **Referring Expressions**
- **w/ Nesting!**

Is the umbrella upside down?

- **Object Recognition**
- **“Simple” Commonsense**
Identifying & Augmenting Capabilities via Interaction

Questioners
Humans (Group 1)

Is the person sad? Incorrect ✗
What color is the umbrella? Correct ✓
Is the woman on the right happy? Incorrect ✗

Curators
Humans (Group 2)

Seems to have trouble with emotion questions… focus on those!

“Capability” Dataset

VQA Model
Alas…

**Sanity Check:** How does active learning fare?

**Active Learning Preliminaries:**
- Train VQA model on subset of dataset (e.g., 40K examples)
- Iteratively *acquire* batches of unlabeled data to grow train set.
  - *Baseline:* Randomly pick data!
  - *Uncertainty vs. Diversity*

**Why doesn’t active learning work for VQA?**
Related Work

• **Lin and Parikh 2017** — Explores active learning on the VQA v1 Dataset.
  • **Key Takeaway:** Active Learning doesn’t work for VQA unless you *start* with a lot of labeled examples!

• **Siddhant and Lipton 2018** — Large scale study on Binary Classification for NLP.
  • **Key Takeaway:** Bayesian Active Learning (BALD) works well, but, hard to outperform random acquisition!

• **Lowell et. al. 2019** — Active learning performance doesn’t generalize over models/tasks.
  • **Is Active Learning worth it?**

**Our Work:** Complementary; extends study scale (models, strategies), analysis of “what’s holding active learning back” with these datasets!

Part I: Curiosities

“Discontent is the first necessity of progress.”
— Thomas Edison
Is it... the Architecture? Acquisition Function?

Also holds for LSTM-CNN & Logistic Regression Models!
Holds across other models and acquisitions strategies!
Part II: Contextualizing Acquisitions

“Never give up, for that is just the place and time the tide will turn.”
— Harriet Beecher Stowe
What's Going On? Peeking at Acquisitions!

What does the symbol on the blanket mean?

What is the first word on the black car?

What is on the shelf?

**Collective Outliers!**
Groups of examples beyond the “capabilities” of VQA models!

*How can we better formalize or diagnose these examples?*
A Primer on Dataset Maps — ft. VQA-2

Q: What sport is she playing? A: Tennis

Q: What is the person holding? A: Surfboard

Q: What is the symbol on the hood often associated with? A: Pirates

Q: What is the word on the wall? A: rice

Contextualizing Acquisitions with Dataset Maps

Random Baseline

Core-Set (Fused)

BALD

Least-Confidence

Acquisitions by Difficulty

Number of Training Examples

Number of Training Examples
Ablating “Outliers” fixes Active Learning — Aha!

**Procedure:** Remove “hard” outliers subject to Y-Axis of Dataset Map!

**Graphs:**
- **BUTD - VQA-2**
  - Random Baseline
  - Least-Confidence
  - BALD
  - Ceiling Performance (400K)

**Accuracy by Training Size:**
- 90% of Original Dataset
- 75% of Original Dataset
- 50% of Original Dataset
Part III: Takeaways

“A wrong that cannot be repaired must be transcended.”
— Ursula K. Le Guin
Q: What sport is she playing?
A: Tennis

Q: What is the person holding?
A: Surfboard

Q: What is the symbol on the hood often associated with?
A: Pirates

Q: What is the word on the wall?
A: rice

Thanks so much!
Code: https://github.com/siddk/vqa-outliers | Questions/Comments: skaramcheti@cs.stanford.edu