

Introduction

- Visual question answering can provide information about an image to people with visual impairments
- Prior research has identified 51.6% of VizWiz images contain quality issues and of these 14.8% have unrecognizable content due to the image quality¹



Contributions

- Propose live photos as a means to address quality issues and better answer VQA questions for people with visual impairments
- Show that a correct answer can be generated more often from looking at all live photo frames
- Propose entropy as a method to extract the frame with the correct answer from the full set of live photo frames

Prior Work

- Gurari et al. CVPR 2018
- Chiu et al. CVPR 2020
- Lasecki et al. SIGACCESS 2014
- Iwamura et al. CHI 2020
- Bansal et al. ECCV 2020
- Jang et al. CVPR 2017
- Castro et al. LREC 2020
- Tan et al. ICIP 2020

About Live Photos

- Consists of a video file generated by the iPhone when the shutter button is pressed and a corresponding 'best' frame selected by the iPhone
- Video file includes data from a couple of seconds before and after the picture was taken²

Question: What is this?



Unanswerable

Hand Sanitizer

Lotion

Water Bottle

Method

- Collect 30 live photos using iPhone 11
- Images are taken in indoor, household environment mimicking that in VizWiz
- Camera motion provides different perspectives on object and creates quality flaws
- Question with each frame is "What is it?"
- High entropy value was used to select the image frame containing the most information as the candidate best frame for VQA

Result

- Tables compare accuracy of predicted answers
- Live photos improve upon the baseline frame selected by the iPhone in 12.9 percent of the data points
- When entropy frame predictions were compared to live photos, the number of equally good predictions increased to 96.8 percent

(a) Frame Selected by iPhone Compared with Live Photo

Image Formats Being Compared	Percent of Datapoints
A Live Frame Better Than iPhone Frame	12.9%
iPhone Frame Better Than Any Live Frame	0%
Both Equal Accuracy	87.1%

(b) Frame Selected by Entropy Compared with Live Photo

Image Formats Being Compared	Percent of Datapoints
A Live Frame Better Than Entropy Frame	3.2%
Entropy Frame Better Than Any Live Frame	0%
Both Equal Accuracy	96.8%

(c) Frame Selected by Entropy Compared with Frame Selected by iPhone

Image Formats Being Compared	Percent of Datapoints
Entropy Frame Better Than iPhone Frame	9.7%
iPhone Frame Better Than Entropy Frame	0%
Both Equal Accuracy	90.3%

Conclusions & Future Work

- Live photos allow for VQA systems to still make accurate predictions if a quality issue occurs in some frames
- Large-scale data collection would allow for assessing potential use cases and implementations of this technology

References:

1. Tai-Yin Chiu, Yinan Zhao, and Danna Gurari, "Assessing image quality issues for real-world problems," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2020.
2. Apple Support, "Take and edit live photos," Available at <https://support.apple.com/en-us/HT207310>, Jul 2020.