Photometric Enhancements to Improve Recognizability of Image Content

Lauren Olson
University of Delaware
lolson@udel.edu

Chandra Kambhamettu
University of Delaware
chandrak@udel.edu

Kathleen McCoy
University of Delaware
mccoy@udel.edu

Abstract

Prior work identifies the challenges image quality issues can present when machine learning systems attempt to recognize image content in photographs taken by people with visual impairments. In this paper, we focus specifically on the impacts of image quality issues on assistive technologies such as visual question answering (VQA), where the presence of photometric quality issues may require additional time and effort on the part of people with visual impairments to retake images with unrecognizable content. We use existing state-of-the-art methods to enhance low light images with the objective of improving recognizability of image content in photos from people with visual impairments. We show that enhancing images significantly improves recognizability of image content and use this result to motivate future studies for improving the results of VQA systems.

1. Introduction

A variety of apps exist that take a photo and provide information to people with visual impairments using machine learning and human-in-the-loop systems [1][2][3][4][5]. One of these apps was the VizWiz system [1], which allowed people with visual impairments to submit a question for a crowd-worker to answer based on information in a corresponding photograph. Gurari et al. created the VizWiz dataset as the first VQA dataset composed entirely of questions and images submitted by people with visual impairments. In contrast to existing VQA datasets at the time, the VizWiz dataset was shown to contain significantly more unanswerable datapoints [6]. Further research by Bhattacharya et al. identified low-quality images as one of the reasons why a question may not have a consistent answer in the VizWiz dataset [7]. Most recently, Chiu et el. [8] published a set of eight labels identifying different types of low-quality images in the data and how instances of low-quality relate to unanswerable datapoints in VizWiz. The types of quality-issues identified are Blur (BLR), Bright (BRT), Dark (DRK), Framing (FRM), Obscured (OBS), Rotation (ROT), Other (OTH), and None (NON) [8].

Our work examines low-light image enhancement to improve photometric qualities after a person with visual impairments has taken a photo. To this end, we train the Zero-DCE network [9] to enhance low-light images and then use the trained network to enhance VizWiz images. Our proposed framework for enhancing images and then using them for assistive technologies shows promise for improving the recognizability of the image content and thereby, improving the quality of information people with visual impairments can obtain. Our main contributions are:

- Demonstrating that enhancing VizWiz images improves the ability to predict recognizability of image content.
- Proposing future work with people with visual impairments to assess how enhanced images can improve use of assistive technologies

2. Method

The images used to conduct our experiments are from the VizWiz Image Quality Issues dataset [8]. We use VizWiz data because it provides images collected from people with visual impairments along with crowd-sourced annotations for image quality issues and recognizability of image content. To enhance the images, we selected the Zero-DCE [9] neural network, which provides state of the art results for brightening dark images while reducing noise. We use Zero-DCE++ data for training, since the authors report that the original training data was lost [9]. Training was performed using a NVIDIA Quadro P1000 GPU and results of training were verified through replicating the test process from the Zero-DCE paper [9] on the DICM [10] and LIME [11] datasets.

Our objective is to create a system that automatically takes in a raw image, enhances the lighting if needed, and produces an answer to a question. Separating out dark images from the rest of the data is important, because enhancing an already bright image could cause it to become washed out and degrade the image quality. Therefore, it is necessary to create a procedure for automatically selecting which images are enhanced. We examined the histograms of enhanced images from the DICM [10] and LIME [11] datasets and VizWiz training images to determine a threshold value of 76 average intensity for images that would not become washed out when enhanced. Images below the selected threshold were enhanced with the Zero-DCE network before being used while images above the threshold were used without this additional step.

Our hypothesis that enhancing low-light images will make it easier to recognize image content for accessible applications suggests that when an image is enhanced, if the content was previously unrecognizable it will become recognizable and if the content was previously recognizable it will stay recognizable. To test this premise, we use the same pre-trained model as Chiu et al. [8] and ResNet-152 image features. A set of 1004 images from VizWiz with an average intensity below the threshold were run through the pre-trained model before and after being enhanced.

3. Results

Model Run	Recall	Accuracy	Precision
Before Enhancement	91.88	**89.74**	**94.26**
After Enhancement	**96.80**	89.34	89.75

Table 1: Results for predicting the recognizability of image content. Results are shown for a model tested on images before they are enhanced and the same model tested on images after enhancement.

Enhancing images using the Zero-DCE model generates a five percent improvement in recall when predicting recognizability of image content (Table 1). We focus specifically on the results for recall, since this reflects how many images were labeled as recognizable but predicted to be unrecognizable. The increase in this category therefore indicates that enhancing images improves the prediction of recognizable images, validating our hypothesis. There is also a corresponding decrease of five percent in the precision of the model predictions from data points that were originally labeled as unrecognizable but predicted to be recognizable after enhancement. Since the model predictions are assessed using the original ground truth labels, one possible explanation for this decrease is that the image content has been changed so drastically by enhancement that the original crowdsourced label no longer accurately reflects the image content.

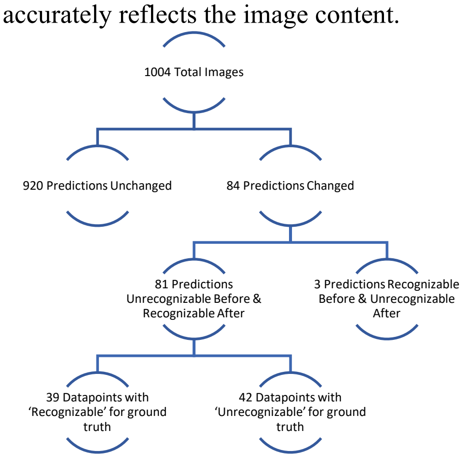

Figure 1: A visualization of how model predictions are distributed and the changes in predictions that occurred from before to after image enhancement.

To obtain a more detailed analysis of how frequently changes to ground truth labels may be occurring, we subdivide the datapoints into categories based on whether the prediction changed, what type of change occurred, and what the ground truth label is (Figure 1). We focus our examination on the 42 images where crowd-workers considered the content of the original image to be unrecognizable to see if the enhancement process has actually changed the recognizability of the image content. Examination showed 13 images where content appeared recognizable after enhancing, 7 images unrecognizable possibly due to noise from enhancement, 18 brightened images where content remained unrecognizable due to the presence of other quality issues, and 4 other images.

We also looked at the three images whose content was initially predicted as recognizable but was predicted unrecognizable after enhancement to determine the roll the enhancement process may have played in this change. In examining these three images, we determined that only one of the enhanced images suffered significant degradation of the image quality. The lowered quality was likely due to the image containing a lit graphics display where text was brightened during enhancement to a degree that caused letters to blur together.

4. Conclusions and Future Work

Overall, our results validate the hypothesis that enhancing images improves recognizability of their content. Our investigation also suggests that careful consideration must be made to ensure enhancement of images is actually beneficial and not detrimental. Using a threshold value for dark images shows promise for limiting the instances of over-enhancing images to a degree that the content degrades, since only one such case was identified in our test data. Based on our examination, we also propose that careful consideration should be made when enhancing data to ensure that data is re-examined after enhancement to maintain existing privacy protections.

Here we present the theory for benefits from enhancing image content in photos from people with visual impairments. Future work with the community of people with visual impairments will be needed to better understand whether incorporating such technology into accessible applications would simplify the photography process and improve results. Other future work in this area could include larger-scale studies to generate ground truth labels for enhanced images and integration of image enhancement with automatic detection of the types of quality issues present in an image.

References

[1] Jeffrey P. Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C. Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samual White, and Tom Yeh, "VizWiz: Nearly real-time answers to visual questions", in *Proceedings of the 23rd Annual ACM Symposium on User Interface Software and Technology,* New York, NY, USA,

2010, UIST '10, p/ 333-342, Association for Computing Machinery.
[2] Aira. Aira. Available at https://aira.io/, Nov 2021.
[3] CloudSight. Taptapsee. Available at https://taptaseeapp.com/, Apr 2021.
[4] Be My Eyes. Be my eyes. Available at https://www.bemyeyes.com/, Apr 2021.
[5] Microsoft. Seeing ai. Available at https://www.microsoft.com/en-us/ai/seeing-ai, Nov 2021.
[6] Danna Gurari, Qing Li, Abigale J. Stangl, Anhong Guo, Chi Lin, Kristen Grauman, Jiebo Luo, and Jeffrey P. Bigham, "VizWiz grand challenge: Answering visual questions from blind people," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
[7] Nilavra Bhattacharya, Qing Li, and Danna Gurari, "Why does a visual question have different answers?", in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV),* October 2019.
[8] 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.
[9] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* June 2020.
[10] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. In *IEEE Transactions on Image Processing,* 22(12):5372-5384, 2013.
[11] Bo-Wen Zeng and Kin Tak U. Low-light image enhancement algorithm based on lime with pre-processing and post-processing. In *2020 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR),* pages 96-101, 2020.