How does a Machine Judge Photos? Comparing humans and algorithms

Wasim Ahmad
Syracuse University
wahmad@syr.edu

Abstract

Machine-vision technology has progressed to the point where it can do much more than just identify what's in a photo; it can tell what makes a photo good or bad. This study investigates how well the current technology used by a company recently acquired by computing giant Apple works by comparing software that uses this algorithmic approach to judge photos aesthetically to how professional photojournalists view the same photos. One-on-one interviews revealed that while humans varied in their responses to a photo, they often provided more than just surface-level commentary, adding extra elements related to context and their experience. Their preconceived biases also coloured their aesthetic evaluation.

Introduction

Machine perception has come a long way. Computers have advanced from recognizing simple text, to voice, and now a new frontier: images. But the pace of image-recognition technology has not kept up with the easier media of text and voice. With so many pixels and so much information to digest, the technology required for a computer to fully understand the context and content of a photo is still a long way off.

Still, the algorithms used today are becoming ubiquitous. Even services such as the lowly Flickr can now recognize basic items in photos, as can Google Photos. Apple's iPhone has become very good at recognizing faces and organizing them into albums.

Few have applied the technology to have it conjure up more meaning to an image than that. However, a French company, Regaind, has put its algorithms to use to try to better understand photos that are being run through its software. The company's software was good enough to catch the attention of Apple, which quietly purchased the company in September 2017. The service was shut down during negotiations. [1]

In 2016, Regaind created a public demonstration of its software, aimed at photographers who wanted a critique of their photographs. The program was called Keegan, the photo coach (https://keegan.regaind.io/). The underlying premise of this website was to use the algorithms created by this company for use in its business dealings to identify objects and categorize photos for another purpose: to critique a photo so the photographer could improve upon it. When a photographer uploaded a photo to the site, Keegan provided both written feedback and a numerical dataset, which ranked the photo according to several different metrics. An example is shown in Appendix A. The Keegan website was retired on Feb. 10, 2017, but Regaind still offered the technology in a more business-oriented format, without the qualitative, human-sounding feedback, to paying customers.

When the Keegan site was launched in 2016, it made waves in the photo industry. Whereas previously, photographers needed a knowledgeable human to obtain a critique of their photos in words, now a machine could provide the same service using algorithms. [2] The software could also output quantitative data about the photo, opening up a completely different avenue of study than that presented here.

This topic is of particular interest because of the potential seismic shift for a particular genre of photography: photojournalism. Photo editors

at a major event, such as the Olympics, can often receive upwards of 10 images per second from a working photographer, and going through them under ever-tightening deadlines is a difficult task. [3] If the technology existed to separate the good photos from the bad, the editors could work much faster. Of course, there is also a chance that the editors could be replaced.

The acquisition of Regaind by Apple increases the salience of this study. It is the most recent, and possibly the only study that examines the efficacy of software that may power image recognition technology on every iPhone and iPad on the market. The software was shut down mid-study, when Apple entered into negotiations with the company's founders, as hindsight revealed.

It is not an exaggeration to say that not only this software, but image recognition technology in general will shape the future of image editing across multiple industries, so understanding the logic and process behind such software is crucial. Human editors need to make decisions about photos for public consumption, and users need to curate their own personal libraries, so this study attempts to understand how this artificial intelligence works by examining the responses to the latest software in the field. It is through gaining this understanding that the implications of this technology on the media industry will be realized.

Literature Review

Previous research on this topic has mostly come from the realm of engineering. Some researchers have placed high consideration not only on how the aesthetic value of a photo affects machine perception, but also on how the technical aspects of a photo, such as compression and noise, affect a machine's evaluation of a photo. [4]

Other research has focused on what humans find memorable in photographs, and not surprisingly, photographs with human subjects tend to be more memorable than those without. Colour and "interestingness" were also factors affecting a photo's memorability. [5] If a machine tracks the same way, it could have far-reaching implications for the photo industry.

In the communications realm, there has not been an analysis of the direct battle between humans and machines in this area, though there are studies on human vs. human competitions: e.g., photos from professional photographers vs. those from citizen photojournalists. [6] There has even been a study of professionals vs. professionals, looking at which newspaper staff are more professional and whether this professionalism produced better photography. [7] This study throws a machine into the mix, Regaind's Keegan, comparing its qualitative responses to photographs to insights from professional photojournalists obtained through in-depth interviews. The goal of the research is to determine how far along image recognition technology is, and to study whether in its present state, its perception can rival that of humans in journalistic fieldwork. The aim of this exercise is to see if software can achieve even a basic level of competency in identifying aesthetic qualities of a photo compared to photojournalists.

With that in mind, the following research questions are examined in this study:

RQ1: How close to a human response does a computer algorithm get when looking at the aesthetic qualities of a photograph?

RQ2: What contributes to the difference between a computer's interpretation of a photograph and a professional journalist's?

This approach holds appeal for both the engineering world and the communications world, putting to practical use this image-recognition technology and comparing it to human capability. Comparing human and machine results offers researchers an opportunity to further improve upon image-recognition technology until there is parity, at least from an aesthetic perspective. This will move image recognition to the next frontier of deciding which photographs are important in context. This is a skill that for the foreseeable future will require the hand of skilled human editors no matter how good the machines get.

Method

In this study, five photographs were run through Keegan, and its qualitative evaluations were

recorded. The photos were shot by the researcher or an associate and were not famous enough to have been published elsewhere. Although there are many famous photos that easily come to mind when considering photojournalism (many readers may have a ready image in mind, such as Nick Ut's Vietnam War-era "Napalm Girl" photo or Richard Drew's "Falling Man" photo from the 9/11 terror attacks on New York), there's a risk that the participants in the study would bring their own preconceived notions of these photos to their interpretations of the aesthetic qualities. To avoid this, the photos used were taken by the researcher so that viewers would not have any history with them. This is similar to an approach used in a previous study to prevent prior memories of photographs from interfering with the study. [8]

Ten current and former professional photojournalists and photo editors were chosen through purposive sampling for one-on-one, semi-structured in-depth interviews about the same five photos. They were asked first for their overall impression, and then asked to comment on items that Keegan frequently brought up, including composition and framing, background, exposure and lighting, colour, moment, blur, and a numerical rating. The participants all had a minimum of five years of experience, ranged in age from 27 to 64, and comprised five males and five females. For the in-person interviews, printed photos were used, and for phone interviews, e-mailed photos were used. Their interviews were recorded, transcribed and then inputted into NVivo for analysis.

First cycle coding was done as magnitude coding. [9] The criteria outlined in the questions (composition, background, colour, exposure, moment and blur) were coded as positive or negative. Keegan was also included in this process. Coding the human responses revealed additional themes that were unexpected, and pattern coding was used to group these thoughts together to reveal more information.

Results

The human responses differed in several ways from those of the software. This was not unexpected. However, what was unexpected was that in some cases, there were advantages to using the machine responses over human responses, the biggest being consistency.

Evaluating context

Context was one of the most dominant themes to come up. The participants consistently asked how a photo was going to be used. By contrast, a machine such as Keegan has no outward appearance of caring about context, but that doesn't mean that context isn't coded in. There's just no way to tell if Keegan was programmed by landscape photographers or photojournalists.

For instance, Jason, a photojournalist-turned-studio photographer, had this to say about an extensively altered portrait photo of a young child dressed as Thor, a comic book superhero: "I like what they did in this photo. I've seen some of this work down in Texas; a couple of guys used *Photoshop*, and it had a really nice effect. It's cute. It made me laugh. They definitely caught the moment."

Contrast that to how Leslie, a former photojournalist, acknowledges her bias about the same photo: "I will say [I rate this photo] a 5, because I just hate studio pictures … but that has nothing to do with it; it's a great, fun photo of your child or someone's child, so that's good, and I think that, you know, my bias comes from being a professional photojournalist. If I were a portrait photographer, I might give it a 10."

Keegan's programming seemed to be keyed in by portrait and studio photographers, because of all of the photos in the study, the child Thor was its favourite. He said this of it: "I'm interested, and I don't want to look away; congratulations! Composed quite well. Very dynamic. Overall, pretty good shot! 8.7/10; you deserve it, champ. Everything is so perfectly framed that you get the framing ribbon! Now you've got the idea. Feel free to send me as many photos as you want. I'll be glad to comment on them and give you my feedback. After 10 pictures, I will evaluate your level in terms of creativity and composition.

Misunderstanding images

Misunderstanding was a common theme. Even at the most basic level, the human participants

could figure out the intent of the photographer and recognize a distinctive feature, such as a silhouette, as a photographic choice rather than a mistake that required studio lighting to fix, as Keegan suggested. This tied in with experience. At times, Keegan failed to meet even the level of expertise of an entry-level photojournalist, although in some cases, neither provided the deep level of detail those with more experience provided. Keegan's advice for studio lighting was centred on a photo at a fair of the "Zipper" ride, where studio lighting wasn't needed or practical, and the silhouette was intentional. None of the photographers in the study made the same call as Keegan.

Contradictions and bias

The humans participants would sometimes contradict themselves about how they felt about an aspect of a photograph or they indicated bias knowing that a photo was shot with a cell phone. Professional photojournalists often frown upon cell phone photos. This was one area in which Keegan's objectivity was an advantage. Keegan did not seem to differentiate or care what device was used to shoot a photo, and its results were consistent, as opposed to the human participants, who often contradicted themselves in the same sentence. For instance, Jason, the photojournalist-turned-studio photographer, had this to say about the composition of the Dominican Day Parade photo: "I like its composition; I think it's a little loose." These two statements don't make sense in the same sentence without a contrast word. Keegan offered no such ambiguity.

The prejudice of professional photographers is a widely known industry issue. Photographers often frown upon using anything other than professional cameras and instantly dismiss what they consider snapshots with point-and-shoot cameras or phones. This was also true of the photojournalists in this study. Jessie, a photojournalist, had this to say about the photo of kids in a bounce-house: "This is definitely, like, a snapshot of 'Hey look, there's my son' or 'I gotta get a photo of this kid' type of photo." That attitude coloured the rest of her critique of the photo. When asked to rate the photo, she wanted to go lower than the scale allowed and give it a 0. By contrast, since Keegan was programmed by a company, it tended to be more tactful. For example, it had this to say about the same relatively poor photograph: "Nice timing, but a bit blurry. This pick is just … ok. Don't forget about the blur and background. A solid 5.7/10. Not bad, but I'm sure you can do better!"

The human bias was related to experience. In many cases, the photojournalists in this study were blunt with their critiques because they were battle-hardened by field experience. The more experience a participant had, the more detailed their critique, with photographers who also had photo-editing experience providing the most detailed responses. Keegan, by contrast, offered mostly surface-level and similar critiques, likely owing to its limited database of pre-programmed responses to photos.

Conclusions

The machines aren't there yet. But it's not easy to say why. Some research points to technical issues with photos. Resolution and compression, for instance, could put software at a disadvantage, but the same could be said for humans. [10] Print quality or monitor quality was brought up in some cases.

One limitation of the study was the photos themselves. Regaind shut down Keegan earlier than promised, so there was no opportunity to run more journalistic photos through it. The reason for this mysterious cut-off in communication became clear when Apple's acquisition of the company was reported in the media. [11] The photos chosen were a more general set used for exploratory purposes, but they ended up being the main photos used for the study. Regardless, the photos provided some insight into how the program perceives images.

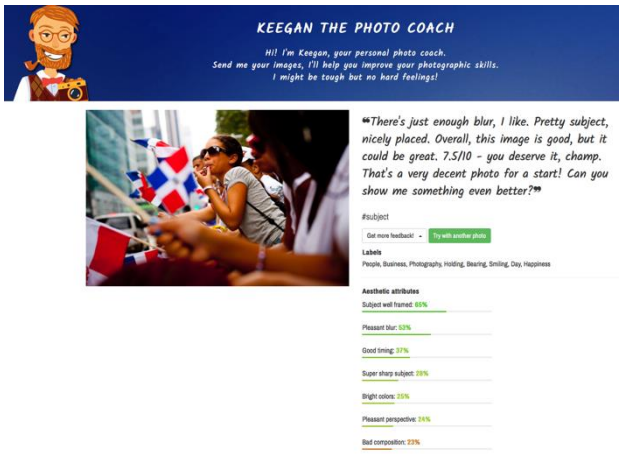
Since this study began, new software and products have come out that utilize machine vision. Amazon, for instance, released a device that takes a photo of users and offers fashion advice. While the technology has significant implications for journalism, there's a wide range of consumer-based applications to be explored, an area ripe for future study.

References

1. Romain Dillet, “Apple quietly acquired computer vision startup Regaind,” *TechCrunch*, September 29, 2017. https://techcrunch.com /2017/09/29/apple-quietly- acquires-computer-vision-startup-regaind/.
2. Michael Zhang, “Keegan is an Online A.I. Photo Coach Who Critiques Your Photos,” *Petapixel*, October 8, 2016. https://petapixel. com/2016/10/08/keegan-online-photo- coach-critiques-photos/.
3. Jack Crager, “How Getty's Olympics Photos are Shot, Edited, and Sent into the World in Just Two Minutes,” *Popular Photography*, August 3, 2016. http://www.popphoto.com /how-olympic-images-reach-your-eyes-in-two-minutes-flat.
4. Xiaoou Tang, Wei Luo, Xiogang Wang, “Content-based Photo Quality Assessment,” *IEEE Transactions on Multimedia*, 15, no. 8 (2013):1930-1943
5. Phillip Isola, Jianxiong Xiao, Devi Parikh, Antonio Torralba, and Aude Oliva, “What Makes a Photograph Memorable?,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36, no. 7 (2014): 1469-1482.
6. Tara Buehner Mortensen and Ana Keshelashvili, “If Everyone with a Camera Can Do This, Then What? Professional Photo-journalists' Sense of Professional Threat in the Face of Citizen Photojournalism,” *Visual Communication Quarterly*, 20, no.3 (2013): 144-158.
7. Thomas Coldwell, “Professionalization and performance among newspaper photog-raphers,” *International Communication Gazette*, 20, no. 2 (1974) :73-81.
8. Phillip Isola, Jianxiong Xiao, Devi Parikh, Antonio Torralba, and Aude Oliva; *What Makes a Photograph Memorable?*
9. Johnny Saldaña, *The coding manual for qualitative researchers*, 3rd Edition (Los Angeles: SAGE 2016).
10. Xiaoou Tang, Wei Luo, Xiogang Wang, “Content-based Photo Quality Assessment.”
11. Romain Dillet, “Apple quietly acquired computer vision startup Regained.” *TechCrunch*, September 29th 2017. https:// techcrunch.com/2017/09/29/apple-quietly-acquires-computer-vision-startup-regaind/.

Appendix A

Following is a sample of the output obtained from running a photo through Keegan the photo coach. As you can see, it offers a few sentences of critique for each photo inputted by the user, followed by a detailed analysis of several attributes of the photo.



Appendix B

These are the five photos used in the study, in the same order presented to the participants. The three phone interview participants viewed these on their computer screens at the highest resolution available for each photo, depending on the camera used. The seven in-person interview participants viewed them as 8.5 x 11” print-outs on Canon Lustre photo paper, printed on a pigment ink-based printer, the Canon Pro-10.