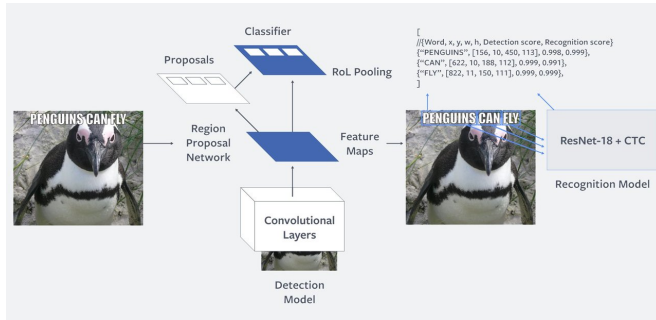


Rosetta system unleashed by Facebook can say whoa to offensive memes

13 September 2018, by Nancy Owano



Two-step model architecture: The first step performs word detection based on Faster R-CNN. The second step performs word recognition using a fully convolutional model with CTC loss. The two models are trained independently. Credit: Facebook

When a meme is beyond goofy and crosses the line to outright offensive, is anyone minding the store?

Say hello To Rosetta, which is a machine learning system that has been engineered to say whoa. Facebook has built and deployed this machine learning system. "It extracts [text](#) from more than a billion public Facebook and Instagram images and video frames (in a wide variety of languages), daily and in real time, and inputs it into a text recognition model that has been trained on classifiers to understand the context of the text and the image together."

Recognition of hate speech via automatic technology is never easy and it gets harder with the times. Rosetta can ease the load of trying to make sure it doesn't slip out undetected. Rosetta is a system that can determine the context of the text and image *together*.

What does that mean? Understanding words, understanding images...but now on to

understanding text *in* images?

Posting to "Facebook Code" site, Viswanath Sivakumar, Albert Gordo, and Manohar Paluri, describe the challenges that beckoned a solution like Rosetta. After all, creatives step beyond traditional articles that are text-centric.

They said a "significant number of the photos shared on Facebook and Instagram contain text in various forms. It might be overlaid on an image in a meme, or inlaid in a photo of a storefront, street sign, or restaurant menu. Taking into account the sheer volume of photos shared each day on Facebook and Instagram, the number of languages supported on our global platform, and the variations of the text, the problem of [understanding](#) text in images is quite different from those solved by traditional optical character recognition (OCR) systems, which recognize the characters but don't understand the context of the associated image."

OK, AI, can we talk memes? Our conversations have multiple condiments. With Facebook, images with text get posted every day—including memes. Rosetta is designed (1) to give screen readers a way to read what's written on them (2) to make sure they don't contain hate speech or violate the website's content policy,

Fast Company pointed out that the system has mostly been applied to still imagery, but Rosetta is just getting its feet wet; this is going to move deeper in. "Facebook plans to increasingly employ Rosetta to extract the meaning of text from video across all its applications," even though the technology is not ready to tackle all videos just yet.

Interestingly, *Fast Company's* Daniel Terdiman saw this as a weapon against memes as there has been a need for effective tools that services can rely on, to root out memes that can be harmful, in content that might otherwise fly under the radar. "We all love memes, and most of us have probably

helped spread them—passing on that cute photo with the ironic text to our many friends on Facebook, Twitter, and elsewhere. But sometimes memes can be harmful, spreading falsehoods about people or [organizations](#)." © 2018 Tech Xplore

Plain and simple, the Rosetta system can do a better job than was previously possible "in understanding harmful or false text used in memes that spread across Facebook and Instagram."

Mariella Moon in *Engadget* discussed how it works, and "it starts by detecting rectangular regions in images that potentially contain text. It then uses a [convolutional](#) neural network to recognize and transcribe what's written in that region, even non-English words or non-Latin alphabets," Moon said. To train the system, she added, Facebook used "a mixture of human- and machine-annotated public images."

What is Rosetta's status right now? Jacob Kastrenakes, *The Verge*: "Rosetta is said to be live now, extracting text from 1 billion images and video frames per day across both Facebook and Instagram."

What's next? Rosetta is not perfect; Facebook wants to get closer to perfection, though and has a to-do list. Moon said the company plans to keep on growing the number of languages it can understand and "to make it better at extracting text from video frames."

Does anyone sense there might be some who will send bad looks to Rosetta as it becomes more known? Maybe. Cohen Coberly in *TechSpot* wrote, "Rosetta will almost certainly be a controversial tool for certain members of the meme-loving public, but here's hoping the technology will prove smart enough to [distinguish](#) between silly-but-harmless content and truly offensive imagery."

Kastrenakes, *The Verge*: "Given the company's well-known moderation issues, a well-functioning system that can automatically [flag](#) potentially problematic images could be a real help."

More information: code.fb.com/ai-research/rosett...th-machine-learning/

APA citation: Rosetta system unleashed by Facebook can say whoa to offensive memes (2018, September 13) retrieved 1 July 2022 from <https://techxplore.com/news/2018-09-rosetta-unleashed-facebook-whoa-offensive.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.