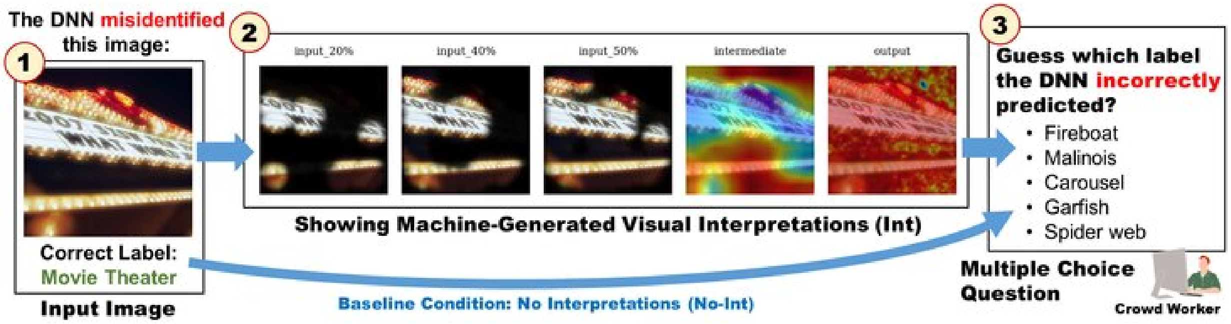
TechXplore

Users don't understand computer explanations for image labeling errors

October 27 2020, by Jessica Hallman


The DNN misidentified this image:
1
Correct Label: Movie Theater
Input Image
2
input_20%
input_40%
input_50%
intermediate
output
Showing Machine-Generated Visual Interpretations (Int)
Baseline Condition: No Interpretations (No-Int)
3
Guess which label the DNN incorrectly predicted?
• Fireboat
• Malinois
• Carousel
• Garfish
• Spider web
Multiple Choice Question
Crowd Worker

An example of the workflow of "Guessing the Incorrectly Predicted Label" task in the research. In the project, each worker was presented with an image and told that the deep neural network incorrectly predicted its label (Step 1). Some workers were also presented with visual interpretations in the form of a saliency map (Step 2). Each worker was then asked to guess the incorrectly predicted label -- "Carousel" in this example -- from five options, four of them being distractors (Step 3). Credit: Pennsylvania State University

When images are uploaded to online platforms, they are often tagged with automatically generated labels that indicate what is shown, such as a dog, tree or car. While these labeling systems are often accurate, sometimes the computer makes a mistake, for example, recognizing a cat as a dog. Providing explanations to users to interpret these mistakes can be helpful, or sometimes even necessary. However, researchers at Penn State's College of Information Sciences and Technology found that explaining why a computer makes certain mistakes is surprisingly difficult.

In their experiment, the researchers set out to find if users could better understand image classification errors when having access to a saliency map. A saliency map is a machine-generated heat map that highlights the regions in images that the computer pays more attention to when deciding the image's label, for example, using the cat's face to recognize a cat. While saliency maps were designed to convey the behavior of classification algorithms to users, the researchers wanted to explore whether they could help explain mistakes the algorithm makes.

The researchers showed images and their correct labels to human participants and asked them to select from a multiple-choice question the incorrectly predicted label that the computer had generated. Half of the participants were also shown five saliency maps, each generated by a different algorithm, for each image.

Unexpectedly, the researchers found that displaying the saliency maps decreased, rather than increased, the average guessing accuracy by roughly 10%.

"The takeaway message (for web or application developers) is that when you try to show a saliency map, or any machine-generated interpretation, to users, be careful," said Ting-Hao (Kenneth) Huang, assistant professor of information sciences and technology and principal investigator on the project. "It doesn't always help. Actually, it might even hurt user experience or hurt users' ability to reason about your system's errors."

However, Huang explained that computer-generated output is important for users, especially when they need to use this information to make decisions about important things like their health or real estate transactions.

"Say you upload photos to a website to try to sell your house, and the website has some kind of automatic image labeling system," said Huang. "In that case, you might care a lot if a certain image label is correct or not."

While this work contributes to a potential direction for future research, the researchers look forward to even more human-centric artificial intelligence interpretations being developed in the future.

"Although an increasing number of interpretation methods are proposed, we see a big need to consider more about human understanding and feedback on these explanations to make AI interpretation really useful in practice," said Hua Shen, doctoral student of informatics and co-author of the team's paper.

Huang and Shen will present their work at the virtual AAAI Conference on Human Computation and Crowdsourcing (HCOMP) this week.

Provided by Pennsylvania State University

Citation: Users don't understand computer explanations for image labeling errors (2020, October 27) retrieved 10 April 2024 from https://techxplore.com/news/2020-10-users-dont-explanations-image-errors.html