Detecting Photorealistic Computer-Generated Images

CS 221 Project Proposal

Shawn Zhang and Gabriel Mukobi

szhang22@stanford.edu gmukobi@stanford.edu

https://github.com/mukobi/CS221-Project

Project scope and input-output behaviour of the system

Our project scope is to build a classifier that is able to detect whether an image is a real photograph or a computer-generated image (CGI). Our inputs are RGB images of "real" photographed scenes and "fake" computer-generated scenes. We output both a classification of whether the image is "real" or not as well as our confidence.

Evaluation metrics for success

With a standard binary classification problem, we use standard binary classification evaluation metrics such as accuracy, precision, recall, and AUC to evaluate our model.

Preliminary data with concrete examples of inputs and outputs

We are using the Columbia Photographic Images and Photorealistic Computer Graphics Dataset, a dataset for multimedia forensics research. For detailed information about the dataset, please refer to T.-T. Ng, S.-F. Chang, Y.-F. Hsu, M. Pepeljugoski, "Columbia Photographic Images and Photorealistic Computer Graphics Dataset," Technical Report 205-2004-5, Columbia University, Feb 2005.

This dataset includes 3600 images split into 4 different classes. The classes are:

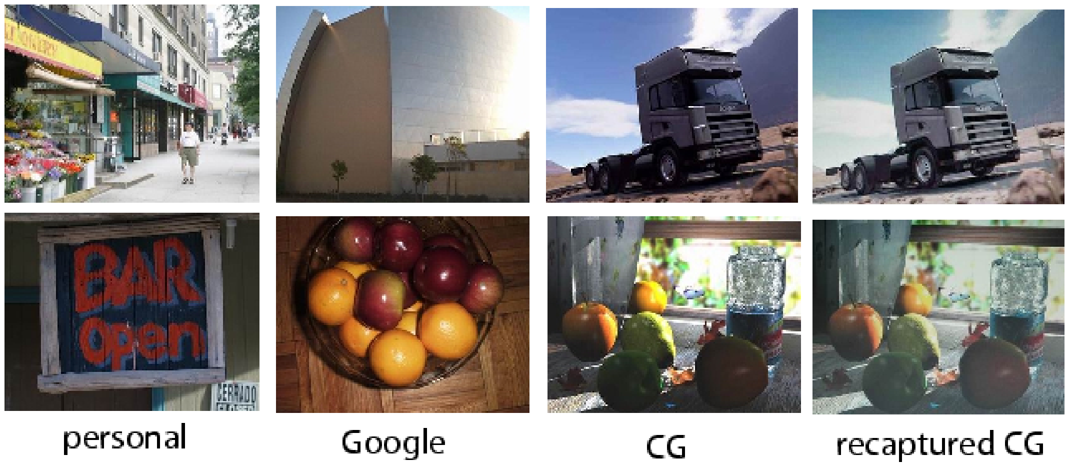
Personal: 800 photographic images (PIM) from the Columbia dataset authors' personal collections and 400 images from the personal collection of Philip Greenspun. These have reliable sources but limited diversity in content.

Google: 800 photographic images (PIM) from Google Images searches for the categories from the CG set. These have diverse image content but the ground truth may not be as reliable as the Personal class.

CG: 800 photorealistic computer-generated (PRCG or just CG) images taken mostly from 40 3D-graphics websites.

Recaptured CG: 800 of the same images from the CG class, but displayed on an LCD monitor and recaptured with a physical camera to restore the photographic effect.

Within these 4 classes, we primarily care about the binary distinction between "real" PIM images (Personal, Google) and "fake" CG images (CG, Recaptured CG). For any given input image, we wish to predict the corresponding output classification as either photographic images (PIM) or computer-generated (CG).



Example images from the dataset. Source: *Ng, Chang, Hsu, Pepeljugoski*.

Baseline, oracle, and the gap

For a baseline, we implement a linear classifier over all the source image data. For speed and compatibility with the linear classifier, we resize all images by scaling down to 32x32 pixels and flattening to a 1D vector. Such inputs were run through a single layer logistic regression. This achieves a classification accuracy of **69%**.

As an oracle, we use human subjects and had them each classify 21 PIM and 21 CG full-resolution images. Our oracle achieves an average classification accuracy of **90%**.

There is a significant gap between our baseline and oracle. The baseline performs fairly poorly, likely due to its inability to understand image data and very low-res input images. Our oracle performs quite well but not perfectly, perhaps indicating a need to collect more difficult-to-discern CG data.

Challenges and topics to address them

Our linear oracle performed quite poorly on image data. This could be improved significantly with the use of convolutional neural networks (CNN). Additionally, our dataset is from 2005, and computer graphics has evolved considerably since then. To gather more challenging data, we could create and test our own modern CG dataset.

Similar projects and related work

The authors of the Columbia image dataset used here took a shot at classifying images using the dataset. In 2005—notably before the widespread use of convolutional neural networks (we define this as 2012 with *Krizhevsky et al.*'s ImageNet)—the Columbia team was able to achieve a classification accuracy of **83.5%**.

Social implications and relevance

In a time of fake news, it's critical to protect against disinformation and similar threats to authenticity. Several others have built DeepFake detectors to combat faked facial animations. We believe more generalized synthesized image detectors like ours could be valuable for ensuring people know the authenticity of any kind of image.

Bibliography

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90. doi: 10.1145/3065386

Ng, T.-T., Chang, S.-F., Hsu, C., & Pepeljugoski, M. (2005). Columbia Photographic Images and Photorealistic Computer Graphics Dataset. *ADVENT Technical Report*, 205-2004-5.