CNNs for Photorealistic Computer-Generated Imagery Detection

Gabe Mukobi
gmukobi@stanford.edu

Shawn Zhang
szhang22@stanford.edu

Stanford

Motivation

In this past decade, computer-generated imagery has improved drastically — so much so that the line between fantasy and reality is greatly blurring. Especially in this time of fake news, it is absolutely critical that we protect against disinformation and similar threats to authenticity.

- This project seeks to build a classifier that is able to detect whether an image is a real photographic image (PIM) or just a photorealistic computer-generated image (CG).
- Our inputs are RGB images of "real" and "fake" scenes, where we output both our classification as well as our confidence.

Dataset

- Data was gathered from the [1] Columbia Photographic Images and Photorealistic Computer Graphics Dataset.
  - 3600 images (2000 PIM + 1600 CG)
  - Sourced from personal collections, Google, and 3D-graphics websites

PIM

CG

Note: surveyed students showed an average classification accuracy of 90% on this dataset.

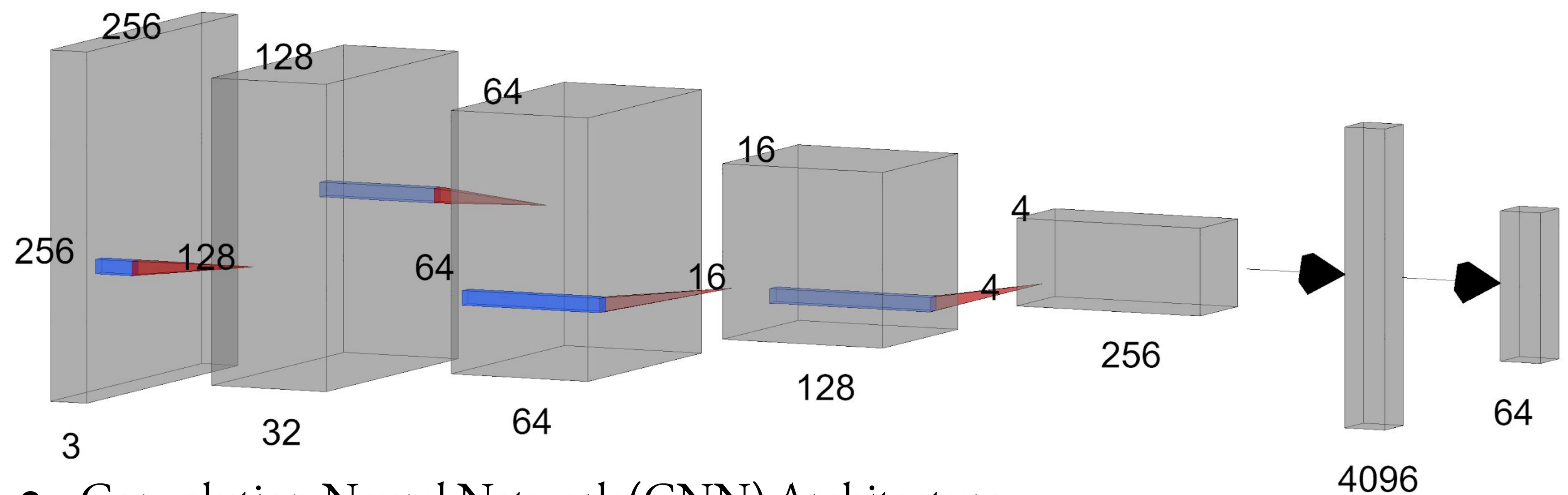
- To counteract overfitting, the data was then extended by the following datasets:
  - [2] Level Design Reference Database (1832 CG)
  - [3] RAISE Raw Images Dataset (2000 PIM)

PIM

CG

Implementation

Here we describe the convolutional neural network that yielded us the best results:

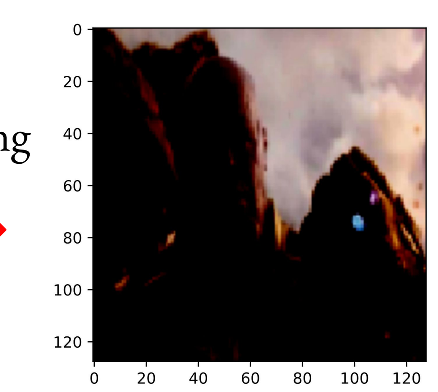


- Convolution Neural Network (CNN) Architecture:
  - 4 convolutional layers, each with batch normalization, ReLU, and max pooling
  - 2 fully connected layers with ReLU (first) and sigmoid (second/output)
- Note: We preprocessed our input images by resizing them to 256x256 and normalizing their color channels.
- Note: Our loss is calculated via Binary Cross Entropy (BCE) and optimized with the Adam algorithm.

Example Pipeline:

Input Image

Preprocessing

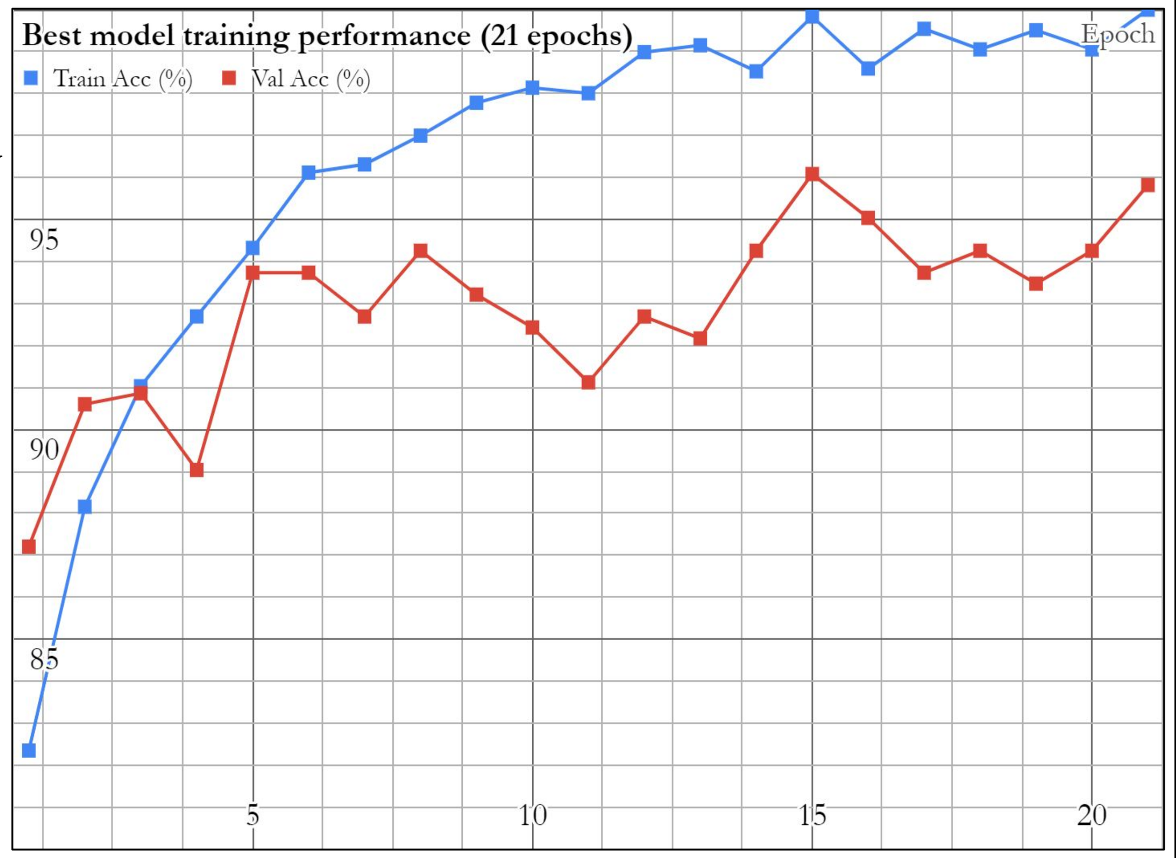


Feed Network

Prediction: CG
Confidence: 81.37%

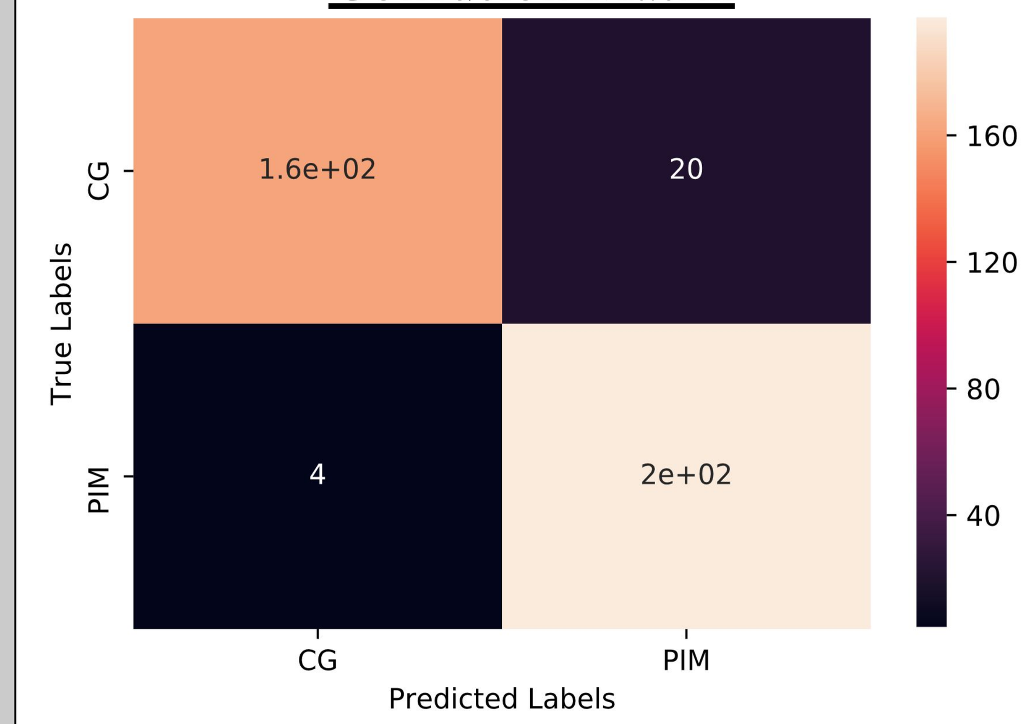
Results

- Best model performance:
  - **99.837%** train accuracy
  - **96.084%** validation accuracy
- On Regularization
  - Best model used only batch normalization
  - Dropout, L2 regularization, and simpler models explored but yielded poorer results
- Additional experimentation
  - Preprocessing (crop/resize)
  - Network/filtering size



Analysis

Confusion Matrix

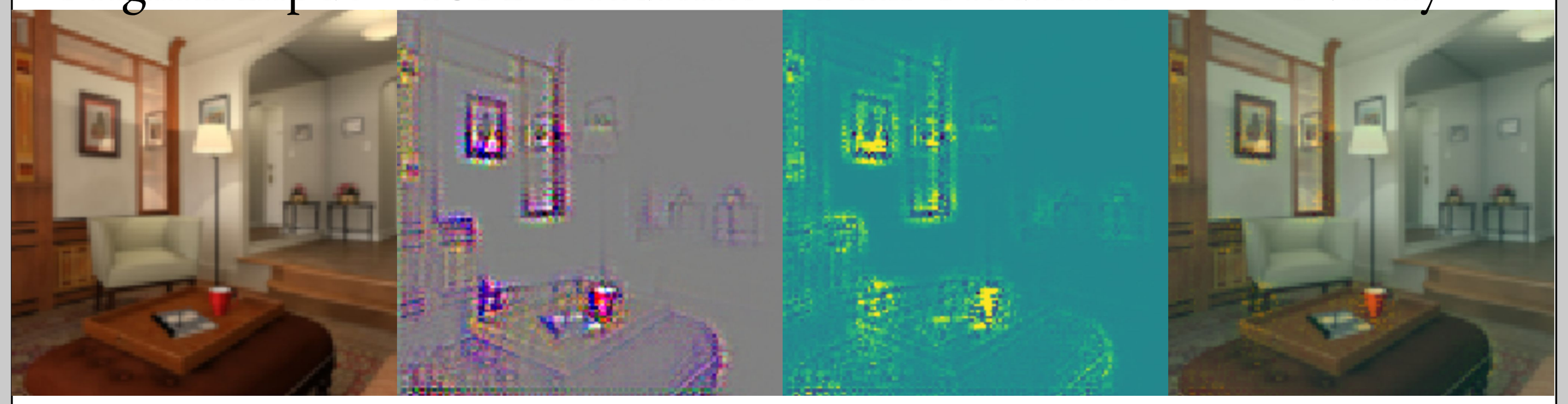


From our validation confusion matrix:

- The model is more likely to misclassify CG than PIM.
- This is an indication on CGI's advance in photorealism.

Saliency Map

Original Input RGB Gradients Max Gradients Overlay



Analyzing the saliency map above, we infer that:

- The most influential pixels seem to compose the image's edges
- The model could be distinguishing PIM vs. CG from edge aliasing (i.e. CG images likely have more pristine edges).

Future Work

On future endeavors for this project:

- Our data consists of CG images from 2005-2014. We wish to collect more recent CG images, challenging our model and demonstrating CGI's evolution over time.
- We may include CG/PIM hybrid images in efforts to perform segmentation and isolate the computer-generated portions.

References

[1] Ng, T.-T., Chang, S.-F., Hsu, C., & Pepeljugoski, M. (2005). *ADVENT Technical Report*, 205-2004-5.

[3] D.-T., Dang-Nguyen, C., Pasquini, V., Conotter, G., Boato. (2015). *ACM Multimedia Systems*.

[3] Level-design Reference Database. (2017). *http://level-design.org/referencedb*.