

AI vs. Real Interior Design Classification

Rahul Jilakara

20058017

GitHub link: <https://github.com/jilakararahul/deep-learning-CA1/tree/master>

Summary

The study in the paper "AI Knows Aesthetics: AI-Generated Interior Design Identification Using Deep Learning Algorithms" is repeated and expanded upon in this project. The main goal is to use a deep learning model to figure out if interior design pictures are real or made by AI. The original study used a private dataset, but this replication used tools that were open to the public to make a balanced dataset and then used a transfer learning method with a ResNet50 model that had already been trained. The classifier was very accurate and good at generalization, showing that it is possible to tell the difference between AI-generated interior design and pictures taken in real life.

Introduction

The rise of visual material made by AI, especially with tools like DALL·E, Midjourney, and Stable Diffusion, makes it harder to tell if something is real. In areas like interior design, clients, professionals, and researchers need to be able to tell the difference between renderings made by AI and real designs. Because of this need, this project is making a binary classifier that can say whether a picture is "AI-generated" or "real." This work copies part of the paper "AI Knows Aesthetics" and shows the same high-performance results using data from public sources.

Methodology

Dataset Preparation

The dataset was made using images that were available to the public so that the conditions of the original study could be repeated. For real interior design pictures, a collection from Kaggle was used. It had 19 groups, such as Scandinavian, Asian, and Modern styles. Randomly choosing 50 images from each group made up a representative subset. This made a real image dataset that was balanced and varied. These were put together in a shared directory structure so that they could be processed further. Images made by AI were gathered from a number of trustworthy sources, such as Lexica.art, PromptHero, and Hugging Face Spaces. These pictures were originally in a bunch of different forms, like .webp, .jpeg, and .png, and they were all different sizes. All images were preprocessed using the same pipeline so that the raw data for training would be the same. All picture formats had to be changed to .jpg, the images had to be resized to 360x360 pixels, and the files had to be renamed one after the other (for example, real_1.jpg, ai_1.jpg). The finished dataset had two folders: one with 500 real images and the other with 500 images made by AI.

Model Building

The model was made with the TensorFlow and Keras frameworks, and transfer learning was used to make training quick and accurate. The model was built around a ResNet50 design that had already been trained on ImageNet. This is ResNet50, a 50-layer convolutional neural network that solves the vanishing gradient problem with residual links. It starts with a convolutional layer and then has a group of identity and convolutional

blocks that are organized into stages. It is very good for deep feature extraction because each block uses shortcut connections to let gradients flow straight through the network.

The layers of ResNet50 that had already been trained were frozen during training to keep the representations they had learned, and a unique classification head was added. This head had a GlobalAveragePooling2D layer to flatten the space between the dimensions, then a fully connected dense layer with 256 units and ReLU activation. To keep things from fitting too well, a dropout layer with a rate of 0.3 was used. In the end, a thick layer with a single neuron and a sigmoid activation function was added to give a chance for a binary classification. The Adam optimizer and the binary cross-entropy loss function were used to put together the model, and accuracy was used as the main evaluation measure. It was trained for five epochs, with 32 people in each batch and 20% of the time spent validating.

Results

Training and Validation Accuracy/Loss

```
Epoch 1/5
48/48 ————— 451s 9s/step - accuracy: 0.9526 - loss: 0.1221 - val_accuracy: 0.9739 - val_loss: 0.0614
Epoch 2/5
48/48 ————— 414s 9s/step - accuracy: 0.9818 - loss: 0.0626 - val_accuracy: 0.9530 - val_loss: 0.0997
Epoch 3/5
48/48 ————— 423s 9s/step - accuracy: 0.9738 - loss: 0.0681 - val_accuracy: 0.9869 - val_loss: 0.0246
Epoch 4/5
48/48 ————— 411s 9s/step - accuracy: 0.9892 - loss: 0.0324 - val_accuracy: 0.9869 - val_loss: 0.0286
Epoch 5/5
48/48 ————— 410s 9s/step - accuracy: 0.9943 - loss: 0.0236 - val_accuracy: 0.9896 - val_loss: 0.0287
```

Figure: Training Accuracy Graph

Prediction Evaluation

A random batch of predictions from the validation set demonstrated that the model correctly categorized nearly every image:

```
1/1 20s 20s/step
Ground truth: [0 1 0 0 0 0 0 0 1 1 1 1 1 0 1 0 0 0 0 1 0 1 1 0 1 0 1 0 0 1]
Predictions:  [0 1 0 0 0 0 0 0 1 1 1 1 1 0 1 0 1 0 0 0 1 0 1 1 0 1 0 1 0 0 1]
```

This steady match between predicted and real labels shows that the model can work well with data it hasn't seen before.

Confusion Matrix

The confusion matrix showed that the sorting worked well. Out of all the guesses, only four were wrong. In particular, three pictures made by AI were wrongly thought to be real, and one real image was thought to be made by AI. The matrix looks like this:

- AI correctly predicted as AI: 176
- AI predicted as real: 3

- Real correctly predicted as Real: 203
- Real predicted as AI: 1

Evaluation Metrics (Binary Classification)

Metric	Value
Accuracy	98.69%
Precision	98.54%
Recall	99.51%
F1 Score	99.02%

These metrics demonstrate a very effective model that consistently maintains high recall and precision, making it reliable for use in practical settings.

Conclusion

Using open datasets and a standard deep learning pipeline, this project successfully copies the main goal of the original study. Using a ResNet50-based transfer learning model and a balanced and well-preprocessed dataset, the system got over 98% accuracy on validation data with very few wrong predictions. The confusion matrix and metric tests show that the learned model is strong and trustworthy.

This replication not only backs up the original paper's results, but it also shows that similar results can be reached using open datasets and processes that can be used again and again. In the future, researchers might look into finetuning the ResNet50 layers even more, adding attention mechanisms or transformer-based models, using Grad-CAM to make visual explanations, or putting together a live web app for real-time AI vs. real design detection.