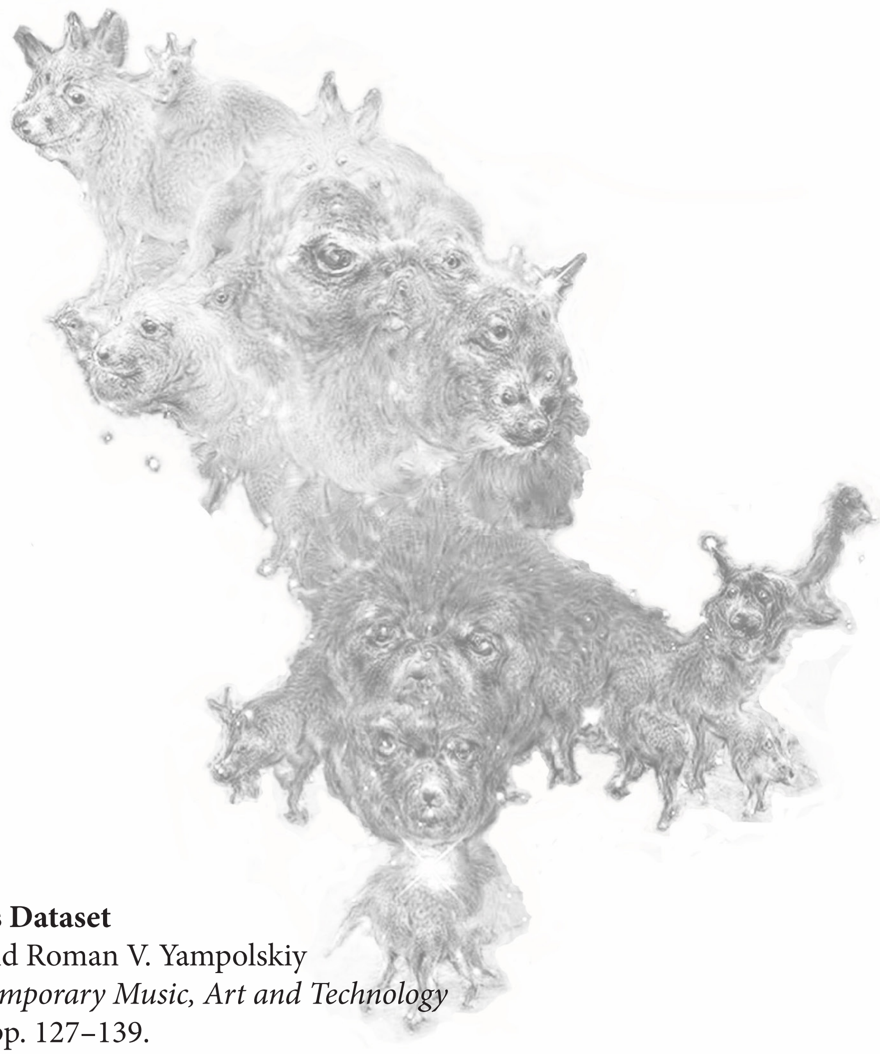
INSAM
JOURNAL OF CONTEMPORARY MUSIC, ART AND TECHNOLOGY



Optical Illusion Images Dataset

Robert Max Williams and Roman V. Yampolskiy

INSAM Journal of Contemporary Music, Art and Technology
No. 2, Vol. I, July 2019, pp. 127–139.

INSAM

UDC 612.821.89:004.8

Robert Max Williams*
University of Louisville
Louisville, Kentucky, United States

Roman V. Yampolskiy**
University of Louisville
Louisville, Kentucky, United States

OPTICAL ILLUSION IMAGES DATASET

Abstract: Human vision is capable of performing many tasks not optimized for during its long evolution. Reading text and identifying artificial objects such as road signs are both tasks that mammalian brains never encountered in the wild but are very easy for us to perform. However, humans have discovered many very specific tricks or illusions that cause us to misjudge the color, size, alignment, and movement of what we are looking at. A better understanding of these phenomenon could reveal insights into how human perception achieves these extraordinary feats. In this paper we present a dataset of 6,725 illusion images gathered from two websites, and a smaller dataset of 500 hand-picked images. We will discuss the process of collecting this data, models trained on the data, and the work that needs to be done to make this information of value to computer vision researchers.

Keywords: Computer Vision, Optical Illusions, Human Vision, Machine Learning, Neural Networks, Cognition

1. Motivation

Being able to understand and intentionally create illusions is currently only possible for humans. The ability to accurately recognize illusory patterns using a computer, and to generate novel illusion images, would represent a huge advancement in computer vision. Current systems are capable of predicting the effect of specific classes of illusions, such as color consistency illusions (Robinson, Hammon, and Sa, 2007) and length illusions (Garcia-Garibay and Lafuente, 2015; Bertulis and Bulatov, 2001). A reinforcement learning system learned to perceive

* Author's contact information: robertmaxwilliams@gmail.com
** Author's contact information: roman.yampolskiy@louisville.edu

color consistency illusions after training to predict color values where half of the image was covered in a tinted film, showing that perception of an illusion can emerge from the demands of seeing in a complicated world (Shibata and Kurizaki, 2012). It is also important to consider whether making a perceptual mistake similar to the mistakes of human perception constitutes having a visual experience similar to humans (Yampolskiy, 2017).

Recent work on generative adversarial networks (GANs) (Karras et al., 2017) has shown that high resolution images of faces can be created using a large dataset of 30,000 images. This quantity and quality of images is not available for optical illusions; as discussed below, naively applying their methods to this dataset does not have the same results.

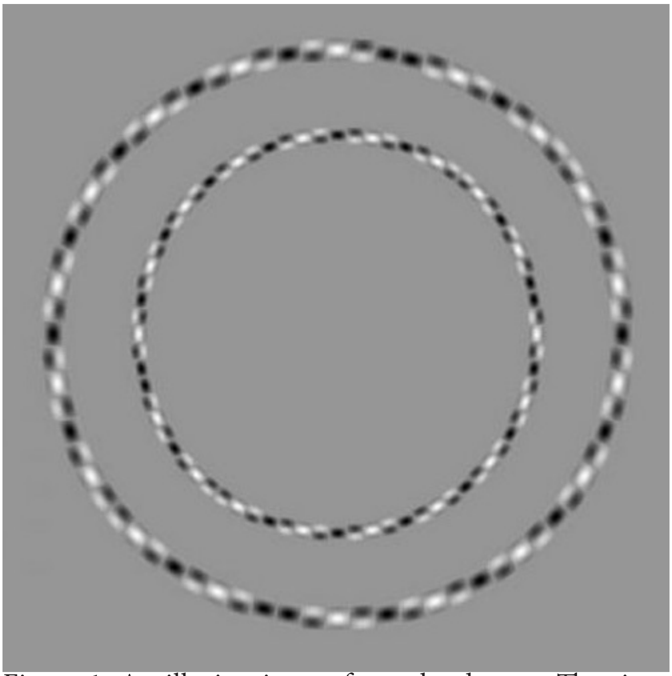


Figure 1: An illusion image from the dataset. The rings are circular and concentric, but the patterns and changes in contrast make them appear to be warped. © viperlib.york.ac.uk

The number of static optical illusion images available are in the low thousands, and the number of unique kinds of illusions is certainly very low, perhaps only a few dozen (for example, the Scintillating Grid illusion, Cafe Wall Illusion and other known categories). Creating a model capable of learning from such a small and limited dataset would represent a huge leap in generative models and our understanding of human vision.

2. Related Works

Research into biologically plausible models makes it possible to learn about visual phenomenon by conducting experiments on proxies for the real human vision system. Elsayed et al. found that by selecting the right models, adversarial examples for these models were also effective on time-limited humans (Elsayed et al. 2018). In their experiment, they created adversarial images for an ensemble of image classification neural networks designed to be similar to human vision. The adversarial images cause the machine learning classifier to classify them incorrectly by only making subtle changes to the image pixels, and they were testing whether these subtle changes would also cause humans to incorrectly classify the altered images. To make the neural networks similar to human vision, they preprocessed their input images to mimic some aspects of human vision, such as higher resolution in the center and lower resolution on the outside. Participants were shown an image in one of two classes, for example, an image of a snake or spider. The images were only shown for 63 milliseconds, meaning that there was not enough time to look at multiple places in the image or reason about its contents on a semantic level. Only the first few "layers" of human vision can work in that short a time span. Their result was that images with subtle changes that could fool an ensemble of neural networks also caused a significant decrease in accuracy for the time-limited humans. This means that current models learned using convolutional neural networks are internally similar to the simplest parts of human vision, and attacks on these neural networks transfer to the visual abilities of time-limited humans. The adversarial examples they created constitute a new class of optical illusions, which can fool the eye into making a mistake when first glancing at an image.

The Brain-Score metric measures internal and behavioral similarity between computer and primate image recognition (Schrimpf et al. 2018). As this metric is developed and models with higher scores are created, those models may be capable of experiencing additional kinds of optical illusions that are otherwise only experienced by primates.

To our knowledge, no dataset of this kind has been created before.

3. Data Collection

3.1. Image Sources

Twelve different websites that collect and display optical illusions (such as the one shown in Figure 1) were considered for inclusion in the dataset. Most proved to be too small or did not contain the right content. For instance, the site "Visual Phenomena & Optical Illusions" contains many interesting and visually powerful demonstrations of optical illusions, but very few still images that by themselves contain a visual effect ("Visual Phenomena & Optical Illusions" 2018). In the end, "Mighty Optical Illusions" (Mighty Optical Illusions, 2018) and "ViperLib" (Thompson and Stone 2018) proved to be the best sources of illusion images, both containing labeled, almost exclusively static images.

Mighty Optical Illusions is a blog-style website, with pages in chronological order labeled as different kinds of illusions and miscellaneous categories. These categories are used as training labels for the classification models. Most of the content on the site are static images, with only a few animations, meaning most of the data could be used.

ViperLib has image pages organized into exclusive categories, but many of them are animations which do not create an illusion when viewed as static images.

The "Illusions of the Year" contest also seemed to be a good source of images, but they only post the winning results publicly (Neural Correlate Society 2018). Emails to the website owner requesting all of the submissions were not answered.

3.2. Data Collection Results

We created a web scraper to go through each page of Mighty Optical Illusions and download the images on the page (source is available at Williams 2018). In total, 6,436 images were obtained, along with their metadata such as categories and page titles. ViperLib was scraped in a similar manner, obtaining 1,454 images also organized into categories and with page titles.

Each image from the Mighty Optical Illusions dataset has one or more tags describing it. Tags such as "anamorphosis" or "impossible objects" were associated with specific kinds of illusory effects, while other tags such as "murals" or "animals" describe the medium or contents of the images. To simplify the training of the classifier, a folder was created for each tag and all images using that tag were placed in its folder. This means that many images were duplicated across categories. In the multi-label classification experiment (Section 4.3), images are included in the datasets based on having or not having a particular tag, so no duplication occurs.

A subset of the data, referred to in Williams paper (2018) as "illusions-filtered," was selected manually as the highest quality illusion images. These images were selected based on having an immediate visual effect without needing any context, such as apparent motion illusions. This hand-picked subset of the data represents the classes of illusions that can be understood solely based on visual stimuli and seemed like the most likely candidates for illusions that a computer could experience, create, and discover. With the current state of machine learning, I expected that identifying and generating pattern-based illusions, such as motion illusions, would be a much easier task than understanding real world objects well enough to identify perspective illusions or Escher-like impossible objects.

3.3. Qualitative Analysis

To determine the feasibility of learning the dataset, we considered how meaningful the classes are and if the images were representative of their class enough to be learned.

The Mighty Optical Illusions dataset was used for the classification experiments (in section 4.1) because it had more images and diversity. However, the labels seem somewhat arbitrary and are difficult for a human to understand from the images alone. Looking at Figure 2, it is not immediately obvious that images in each column belong together.

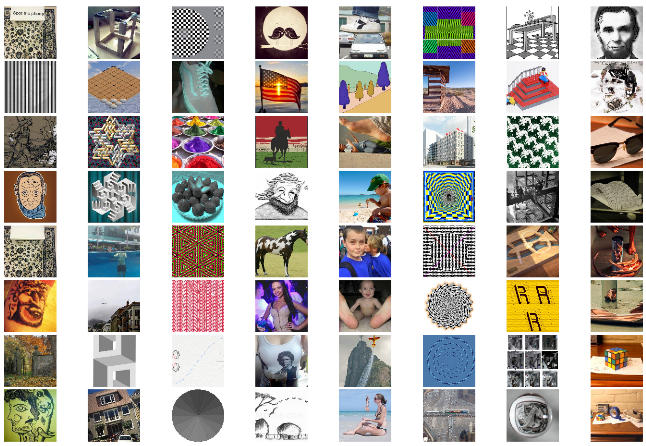


Figure 2: Images from the dataset. All images in the same column have a label in common. Labels are Spot The Object, Impossible Objects, Color Adapting, Multiple Meanings, Relative Sizes, Seemingly Bent, Escher Style, and Anamorphosis, from left to right.

The first class is "Spot the Object," where images contain something that is hard to find but is easy to see after it's been pointed out. Classifying whether or not an image contains a hidden object is a very difficult task, since some of these illusions can require minutes of searching to find the object. This means that to confirm that something does not have a hidden object, you would need to search for as long as the expected time needed to find a hidden object. The "Impossible Objects" category contained images or illustrations of perspective illusions and Escher-like geometries. Given a geometric scene, careful spatial reasoning is required to tell if there is impossible geometry. There are also images based in perspective illusions, with various combinations of clouds, refraction, reflective water, and slanted landscapes that create impossible seeming scenes. "Color Adapting" refers to the eye's ability to adapt to changes in lighting and illusions that are created by taking advantage of this ability, but this category includes a wider variety of images, and it seems that anything color related received this tag. The "Multiple Meanings" category contains images which have more than one appearance depending on how you look at them. Some are very subtle, so this category is difficult for the same reason as finding a hidden object. It overlaps heavily with "Impossible Objects," since many impossible objects appear as being two thing simultaneously which cannot exist at the same time.

"Relative Sizes" contains familiar objects in contexts that make them seem far larger or smaller than they really are. A mismatch between the apparent size of an object in an image and your commonsense knowledge about the actual size of objects is easy for humans to identify, but it seems like a task that would be very difficult for a neural network to learn without being specifically designed for this task. The category "Seemingly Bent" contains a large amount of illusions that are immediately apparent without additional context or knowledge, so I expected this category to be one of the easiest to identify with machine learning. "Escher Style" is the same as "Impossible Objects" but limited to impossible geometries and often the images are in the pencil-sketch style of Escher's artworks. "Anamorphosis" refers to images where viewing them with a specific perspective or lens changes their appearance. For instance, images of sculptures which produce an image when a cylindrical mirror is placed in the center and images that appear different when viewed up close or far away.

Within each class, many of the images do not contain illusions or are only meant as references. For instance, in many "Spot the Object" illusions, a second image with the hidden objects highlighted or circled is provided. How to make use of these images in a machine learning model is unclear. Many "Impossible Object" images also include images of how the object was constructed, none of which contain an actual illusion.

Many confounding factors make using this dataset in a traditional machine learning workflow difficult. This difficulty distracts from the key question: can machines perceive optical illusions as humans do? Expensive hand-sorting of the data could solve this problem, by isolating exactly the images of interest and putting them in consistent and meaningful classes of illusions. Illusions of restricted kinds could be automatically generated, such as variations on motion illusions based on known patterns. Overall, the dataset contains a large portion of images which clearly demonstrate illusions, but many non-illusions are present which makes learning difficult.

4. Machine Learning Results

Three different kinds of models were tested on subsets of the data. Two classifiers were trained to test how visually distinguishable the given classes are, and a generative model was trained to see if new instances of known illusions could be created by naively applying existing methods for image generation.

4.1. Single-label Classifier Results

A pretrained "bottleneck" model (TensorFlow 2018) was used to classify images from Mighty Optical Illusions. Only the last few layers had to be retrained, making use of transfer learning from a much larger dataset to learn how to classify images in general. In this case, the pre-trained model was "Inception v3", trained on over 14 million images in the ImageNet dataset. The pre-trained model converts the very high dimensional image data into a lower dimensional "feature space" vector. This means that the image, a vector of around two hundred thousand values, is compressed into a vector of around two thousand value which contain enough information to accurately classify an image.

Learning new image classes from this feature space representation requires significantly less data and computation time, meaning that these experiments can be run on a normal PC in a few minutes, instead of the days or weeks on a GPU that was required to train the original Inception v3 model.

Each image in the training data may belong to multiple classes, which was not accounted for in the model. In Section 4.3, a multi-label classifier was created for a different subset of the data. The results of training can be seen in Figure 3.

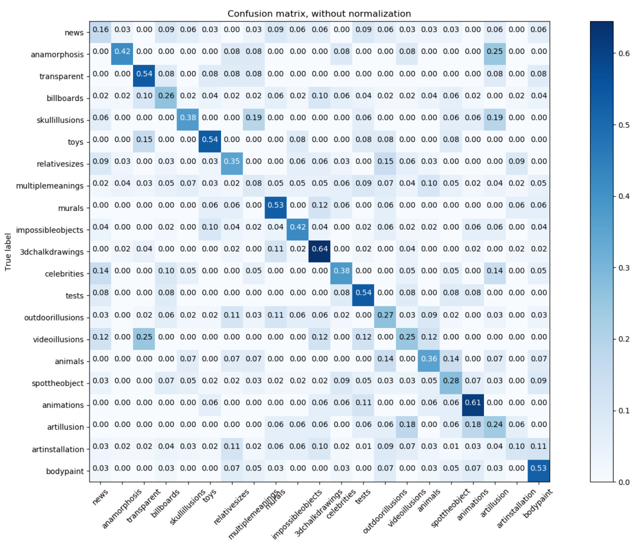


Figure 3: Confusion matrix for a classifier trained on the Mighty Optical Illusions data.

The model performed significantly better than random, meaning that the given classes are meaningful in a way that can be detected using a model trained on normal classes of images. Out of the 21 classes, most classes were predicted with 40-60% accuracy. The very poor accuracy on news, multiple meanings, and art illusions is explained by the lack of defining features for these classes. For example, the "art illusions" category overlaps evenly with most other categories.

An interpretation study could reveal more about how the neural network is able to distinguish these classes, such as the methods used in Zhang, Wu, and Zhu (2017) that show which areas of the image are important to classification and what the key features of each class are.

4.2. Generative Adversarial Network

A trial run using a generative adversarial network was attempted. Using HyperGAN (Martyn, 2017) on a hand-picked subset of the data with no hyperparameter optimization, nothing of value was created after 7 hours of training on an Nvidia Tesla K80. The training progression is shown in Figure 4.

When trained with homogeneous data (such as only using images of faces), GANs are able to create varied and convincing imagery. However, when applied to a varied, multimodal dataset, performance degrades and the generator only learns to generate a single type of image, a problem known as mode collapse (Barnett 2018).

The output produced by the GAN subjectively resembles some sort of scene or objects. It has learned many underlying patterns in the dataset, such as high contrast edges, varied shading, and spatially confined objects. On a small-scale visual level, the generated images appear to be a plausible photographic scene. However, on a larger scale it fails to recreate anything resembling the images in the dataset.

The GAN could be pretrained on a larger dataset to overcome the issue of having such a small dataset. Dataset expansion techniques, such as rotating, cropping, and scaling images, could also be applied to increase the amount of data available for training. The GAN was run with default parameters, which are likely far from ideal for this dataset. Tweaking the parameters to better suit the specifics of this dataset may prevent mode collapse and increase the quality of training.

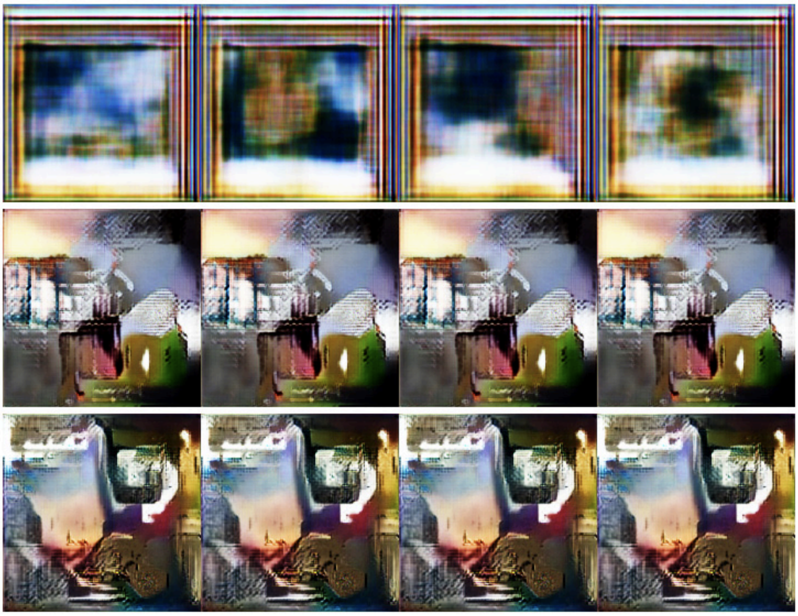


Figure 4: Failure of GAN to generate imagery similar to the dataset. Top to bottom is the progression from start to finish. Images in the same row are from the same training step but with different random input vectors (Goodfellow et al., 2014). The lack of variety is abnormal and may lead to insights into how to correct the problem.

4.3. Multi-label Classifier

In multi-label classification, each image can be in more than one class, and the classifier outputs true or false for each label. Most of the images in the Mighty Optical Illusions dataset have more than one label, so this technique is appropriate for the dataset.

4.3.1. Model

For this experiment, ResNet50 with pre-trained weights was used (Simonyan et al. 2015). The final classification layers were removed and replaced with a densely connected layer with ReLU activations and a prediction layer with sigmoid activation on a single output. This is the same bottleneck training technique used in Section 4.1. An instance of this model was created for each of the target classes and trained separately. To obtain a vector representing all of the labels, the output of all of the models is concatenated together.

4.3.2. Data

Labels that occur less than 70 times in the dataset were removed, as well as labels that do not indicate the content of the illusions, such as “Animals” or “Murals.” Only 8 of the 42 labels met these specifications: Spot The Object, Impossible Objects, Color Adapting, Multiple Meanings, Relative Sizes, Seemingly Bent, Escher Style, and Anamorphosis. In an initial testing run, images with none of these labels were left in the dataset to provide negative samples. This leaves a large majority of the images with no label, meaning that a model that always predicts 0’s for each class will be largely accurate, and the model failed to learn to classify on any of the labels significantly better than random. To better evaluate the model, the data was made into even splits for each label: 50% images which have the label the model is being trained on, and 50% with any other label. For example, the split for “Color Adapting” would consist of 50% images that have the “Color Adapting” label (and possible other labels) and the other 50% would consist of images randomly sampled from the rest of the dataset, contained any label except “Color Adapting”. This is repeated for every label. Another similar dataset was made, but with a third category of images with no labels to provide negative examples. In that dataset, the data for each label was split with 50% having the label the model is being trained on, 25% having any other labels, and 25% with no labels at all.

4.3.3. Results

The model trained on the original, biased data only learned to predict false for every label and fit some of the training data. The models trained on balanced data, however, were able to generalize to the held-out validation set with some success. The validation accuracies for the two models trained on balanced data are shown in Figure 5. Both models failed to learn some classes, and scored very well on some labels, such as Color Adapting. Inspecting images in each class shows that there are many surface characteristics, such as the high contrast shapes in Color Adapting, that make it easy to recognize the label without being able to identify the presence of a color adapting illusion. This dataset allows the model to “cheat” and identify illusions by the way they are presented instead of imitating the human visual system and identifying them as illusions. A means to deactivate illusions without changing surface characteristics would enable a more rigorous test of the model. For instance, in “Skye’s Oblique Grating,” if the odd check marks are not rotated 90 degrees from the even check marks, the illusion disappears. A model that memorized the appearance of the Skye’s Oblique Grating might miscategorize the deactivated version as being an illusion.

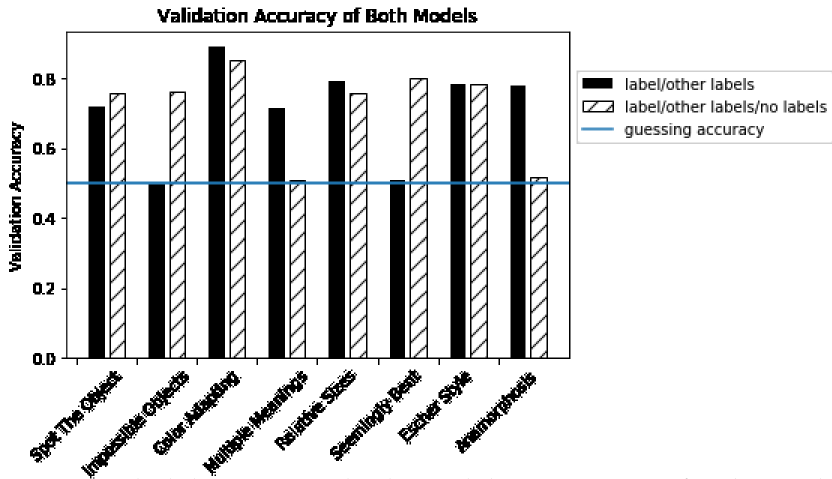


Figure 5: Black bars are single class validation accuracy for the model trained on data split between having that label and not having that label but having a different label. Striped bars are accuracy for the model trained on data split between having that label, having any other label, and having no label. The horizontal line shows the baseline accuracy for a model that always guesses the most likely class, 50%.

5. Conclusion and Future Work

The only optical illusions known to humans have either been created by evolution (for instance, eye patterns in butterfly wings) or by human artists. Both artistic designers of illusion images and the glacial process of evolution have access to active vision systems to verify their work against. An illusion artist can make an attempt at creating an illusion, observe its effect on their own eyes, and add or remove elements to try to create a more powerful illusion. In an evolutionary process, every agent has a physical appearance and a vision system, allowing for patterns to be verified in their environment constantly. A GAN trained on existing illusions would have none of these advantages, and it seems unlikely that it could learn to trick human vision without being able to understand the principles behind the illusions. Because of these limitations, it seems that a dataset of illusion images might not be sufficient to create new illusions and a deeper understanding of human vision would need to be obtained by the network somehow. This could be done by having a human giving feedback as the network learned, or by learning an accurate proxy for human vision and trying to deceive the proxy as in Elsayed et al. (2018).

Appendix A.

Downloading the Dataset Images are currently hosted on the machine learning cloud platform "Floydhub."

- https://www.floydhub.com/robertmax/datasets/illusions-jpg
- This contains all images that were downloaded, using the same numbering scheme as the metadata on the linked github repository.
- https://www.floydhub.com/robertmax/datasets/illusions-filtered
- This folder contains images hand picked for having obvious visual effects without having to follow special instructions.

List of References

Barnett, Samuel A. 2018. "Convergence Problems with Generative Adversarial Networks (GANs)." CoRR abs/1806.11382.

Bertulis, Algis, and Aleksandr Bulatov. 2001. "Distortions of length perception in human vision." *Biomedicine* 1 (1): 3–23.

Elsayed, Gamaleldin F., Shreya Shankar, Brian Cheung, Nicolas Papernot, Alex Kurakin, Ian Goodfellow, and Jascha Sohl-Dickstein. 2018. "Adversarial Examples that Fool both Computer Vision and Time-Limited Humans." eprint: arXiv:1802.08195.

Garcia-Garibay, Otto B., and Victor de Lafuente. 2015. "The Muller-Lyer illusion as seen by an artificial neural network." 25745398[pmid], Front Comput Neurosci 9 (February): 21. issn: 1662-5188. doi:10.3389/fncom.2015.00021. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4333816/. 5

Goodfellow, Ian J., Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. *Generative Adversarial Networks*. eprint: arXiv:1406.2661.

Martyn, Mikkel et al. 2017. "HyperGAN." https://github.com/255BITS/HyperGAN.

Mighty Optical Illusions. 2018. "Mighty Optical Illusions." Accessed January 15, 2018. https://www.moillusions.com/.

Karras, Tero, Timo Aila, Samuli Laine, and Jaakko Lehtinen. 2017. "Progressive growing of gans for improved quality, stability, and variation." arXiv preprint arXiv:1710.10196.

Robinson, Alan E., Paul S. Hammon, and Virginia R. de Sa. 2007. "Explaining brightness illusions using spatial filtering and local response normalization." Vision Research 47 (12): 1631–1644. issn: 0042-6989. doi:https://doi.org/10.1016/j.visres.2007.02.017. http://www.sciencedirect.com/science/article/pii/S0042698907000648.

Schrimpf, Martin, Jonas Kubilius, Ha Hong, Najib J. Majaj, Rishi Rajalingham, Elias B. Issa, Kohitij Kar, et al. 2018. "Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like?" bioRxiv preprint.

Shibata, K., and S. Kurizaki. 2012. "Emergence of color constancy illusion through reinforcement learning with a neural network." In 2012 IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL), 1–6. November. doi:10.1109/DevLrn.2012.6400580.

Simonyan, Karen, and Zisserman, Andrew. 2015. "Very Deep Convolutional Networks For

Large-scale Image Recognition" In 2015 ICLR. abs/1409.1556. arXiv: 1409.1556. http://arxiv.org/abs/1409.1556.

Neural Correlate Society. 2018. "Illusions of the Year." Accessed January 15, 2018. http://illusionoftheyear.com.

TensorFlow. 2018. "How to Retrain an Image Classifier for New Categories." Accessed April 10, 2018. https://www.tensorflow.org/hub/tutorials/image_retraining.

Thompson, Peter, and Rob Stone. 2018. "ViperLib." Accessed January 15, 2018. http://viperlib.york.ac.uk. "Visual Phenomena & Optical Illusions." 2018. http://www.michaelbach.de/ot/.

Williams, Robert Max. 2018. "optical-illusion-dataset." Accessed September 21, 2018. https://github.com/robertmaxwilliams/optical-illusion-dataset.

Yampolskiy, Roman V. 2017. "Detecting Qualia in Natural and Artificial Agents." CoRR abs/1712.04020. arXiv: 1712.04020. http://arxiv.org/abs/1712.04020.

Zhang, Quanshi, Ying Nian Wu, and Song-Chun Zhu. 2017. "Interpretable Convolutional Neural Networks." eprint: arXiv:1710.00935. 6

Article received: February 25, 2019
Article accepted: May 15, 2019
Original scientific paper