Cassie Haas

Professor Ian Bryce Jones

Loop as Form

5/6/2024

Visualizing Machine Repetition

When using a machine to identify images, the machine first has to be trained as to what these images are supposed to look like. There are general conventions to how this is done, but most are fit to the situation at hand. For example, convolutional neural networks are common – they examine small areas in an image at a time, combining their understanding of subspaces into one cohesive whole. When it's necessary to use less computing power, regression models and decision trees are often common. As computers don’t have eyes, images must be radically altered into a format that can be mathematically understood. Take, for example, the MNIST dataset[1].

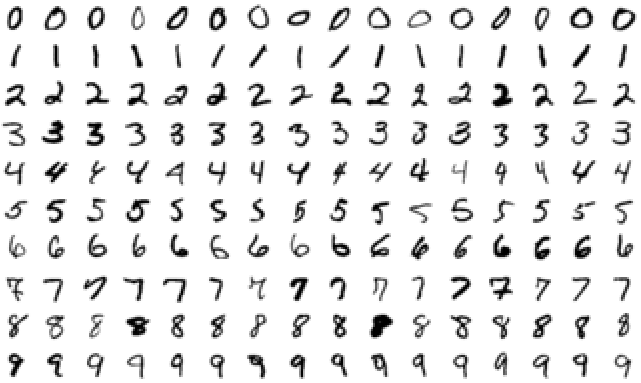

Figure 1: Sample of entries in MNIST dataset

The MNIST dataset is a collection of handwritten digits gathered in an effort to better process, primarily, mail. There are 70,000 images total, and each is black and white without background, shrunk down to 28x28 pixels each. For most models to process these images(excluding cases like the convolutional neural network) it is necessary to change their

[1] “MNIST Database,” Wikipedia, April 30, 2024.

form such that each pixel can be seen as a variable with a given value; in this case, the value is the brightness or darkness of the pixel. Each image is laid out into one vector 784 values long. This is equivalent to taking each row and putting it directly next to the row above it, until there are 28 groupings, each 28 values long, in one row.

This process, however, is not visually accessible to the user. There is a stark divide between how humans interpret visual input on the momentary level to how computers do. At its core, computers *rely* on repetition to understand what they see. They cannot look at a face once and then recognize it later without much more information. So the question lies here: how might we visualize the distortion of images that must happen in order to be "seen" by a machine?

For the purposes of exploring this idea through the lens of the MNIST dataset, 784 images were taken of an individual's bust in one location. They moved, spoke, and fidgeted while photos were being taken, but did not shift drastically from their location.

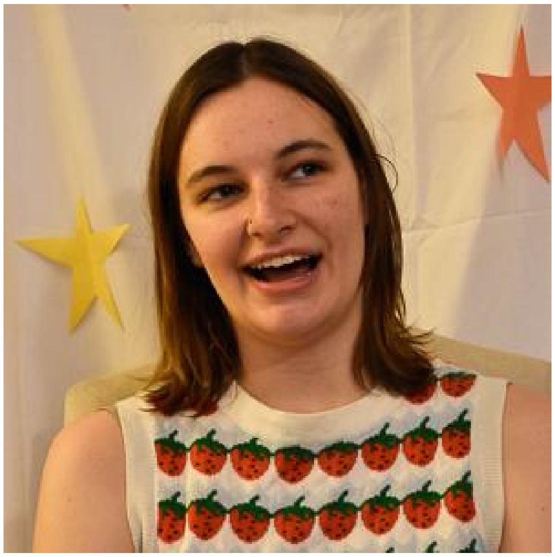

Figure 2: Sample image taken

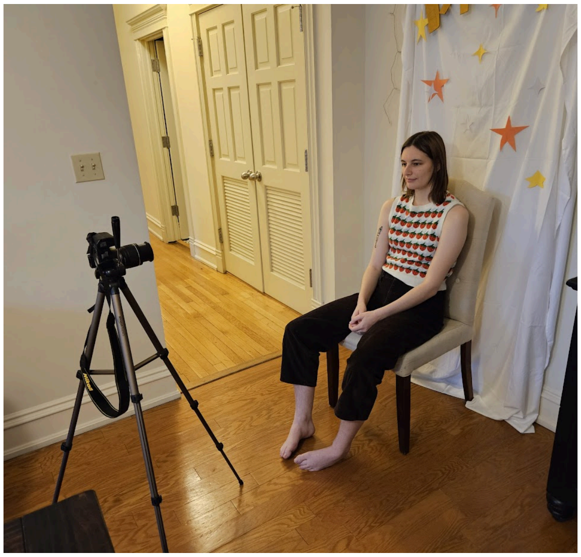

Figure 3: Image of setup with backdrop, subject, and camera

Each of these 784 images was then downscaled to 28x28 pixel size. At this scale, the image was still recognizable but had lost much of its detail. Following the process applied to the MNIST dataset, each image was stretched out into one long vector(though the MNIST deals in monochrome images, color was kept here to retain intelligibility). The vectors were then transposed and lined next to one another. This created a new image with size 784x784 which displays a representation of the transformation that must happen to images in order to be processed.

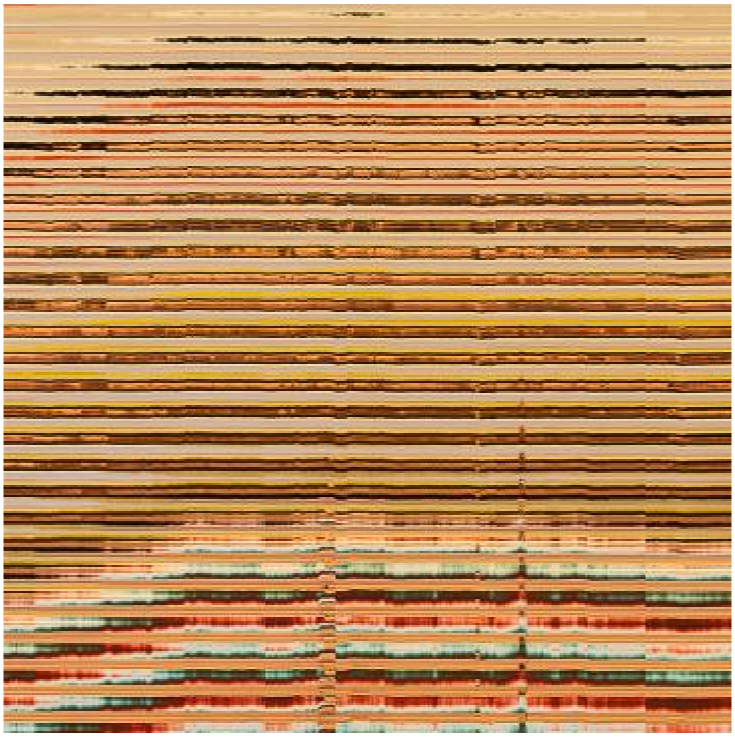

Figure 4: Image at end of its manipulation

The final image poses questions unto itself. Is this still a portrait of the subject? It is created from portraits of them, and attempts to recreate them. Through repetition, however, their form is dissipated. Where their body, clothes, the walls around them end and fold into each other is no longer clear. Looking at the image, it's difficult to not at least *try* to piece together the original subject. Through repetition the photo's subject has been almost completely obscured. Within the image, however, lies the record of a process. Taking the photos took near 800 photos over the course of 20 minutes. The images that make up this composite display us joking, laughing, talking about our families – all while I kept the camera going. Through this, labor is recorded as well.

Further exploration of this topic might see different representations of the form. How might the image be displayed if each pixel were the average of its values across the individual photos? Is it necessary to take photos in one position to still have a cohesive whole? Issues of size of the final image are a concern as well. For each increase in pixel size of the compressed images, the amount of photos necessary to create a composite must be increased drastically as well. Just changing from 28x28 size images to 30x30 would require an extra 116 images with a barely perceptible change in form. A longer project might see the process broken down more - 784 images are taken of one person in one location, and the process is repeated for multiple locations. Those would then be combined much like this into a collection of representations. Regardless of how composites are created or photos are gathered, each stores records of the repetitive labor required to gather the images as well as a distorted form of the body itself.