

First Impressions

A Probabilistic Generative Model: Is That Even a Good Idea?

Jane Sieving • Machine Learning • Fall 2019 • Olin College

Purpose

- Generative Networks are generally very slow, as they require educated guesswork and take a long time to converge on something recognizable
- From what I've seen, probabilistic models can take less training to produce something that is human-recognizable, because they rely heavily on the input data
- I've seen Markov chains used on text and music
 - could I find a meaningful way to apply one to image generation?

Results: Faces



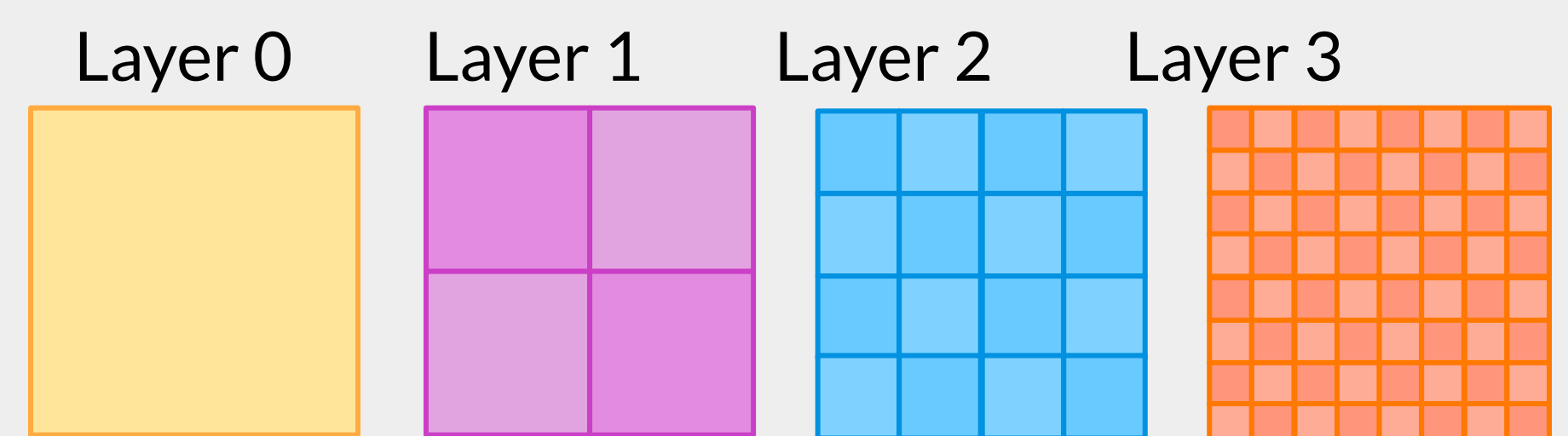
Methods

- Create probability data as an “impression” of a collection of images
- Layers of conditional data, organized by increasing resolution: each previous layer of data has half the resolution.
- Within each layer, each pixel has conditional probability tables for several relevant pixels in the *previous layer*: the bigger one it's part of, and the 4 touching that one. (In the color version, the bigger pixel's color values are also considered.)
- Training: count pixel values, downsample image, create per-pixel table of downsampled value and original value occurring together. Repeat downsampling and creating smaller tables.
- Generation: start with a single average pixel value from the counted distribution. Upsample, probabilistically choosing the new pixel value for each quadrant. Continue to upsample, factoring in neighboring “previous layer” pixels as they become available.
- Likelihood arrays based on “relevant” previous pixels are multiplied and normalized. If all hit 0, random values are chosen.

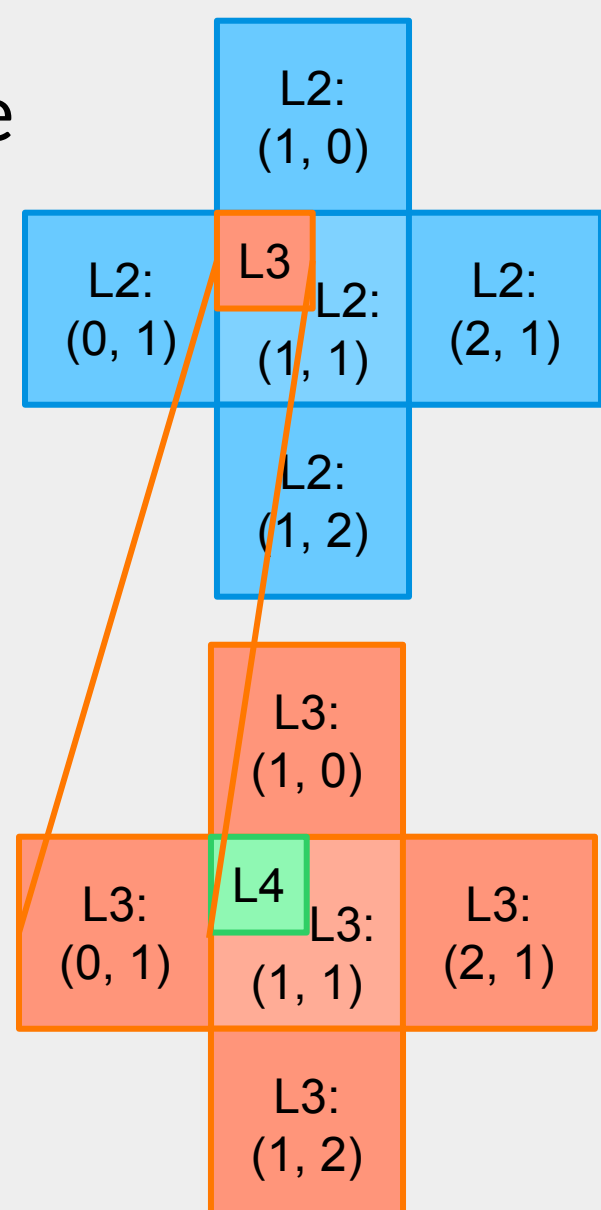
Questions

- Will this model work quickly to produce recognizable output?
- Will the output be coherent? What aspects of the input data will be maintained? (Shape, texture, color, position?)

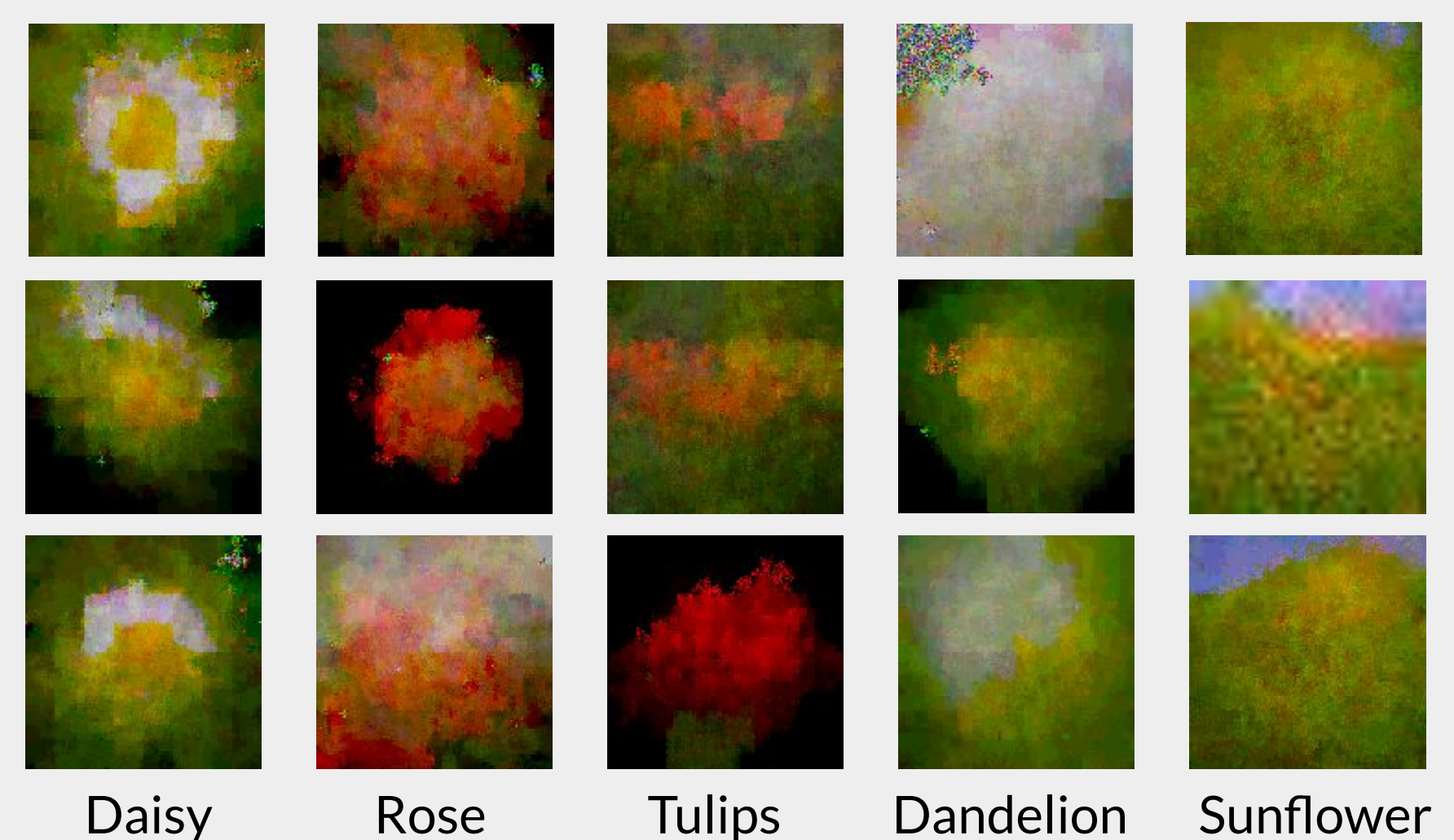
An Impression contains layers of data where the higher layer pixels are averaged to create lower layers of pixels. Each coordinate in each layer stores the possible pixel values, dependent on several nearby pixels.



To predict **every pixel**, the model looks at the **5 nearest pixels** in the previous layer. Once **all pixels for a layer** have been generated, it becomes the information that the **next layer** will be conditioned on.



Results: Flowers (128x128)



Analysis

- Can produce results which *give the impression of* the source images in reasonable time
- Color images are easier to recognize
- Images with consistent subject position are best
- Preserves color well, position when consistent, shape when consistent, texture not really
- Not useful - would not fool a machine - mediocre wall art at best