Art and Machine Learning, 2018 Spring CMU
Assignment 3: Dream or Nightmare, Artist or Assistant
Name: Jacqueline Fashimpaur, Angela Liang, Anirudh Mani

Deep Pix2Stock Salad

CONCEPT

The concept of our work is an exploration of the notion of fair use of external data sources as an art student navigating the complex legalities embedded in the landscape of copyright royalties. Our initial idea for Assignment 3 was to explore social trends by using images tagged with a particular Instagram hashtag as our data set for training. However, we quickly encountered the issue of how to resolve potential copyright conflicts, particularly considering the potential desire to share our work with the public through personal portfolios. Our focus then shifted to obtaining a meaningful data set that was both easily obtainable and legal to use to create our art. This problem thus sparked the final concept for our project.

Our data set consists of royalty-free stock images of "salad" from Shutterstock. The juxtaposition of using perhaps the most laughably obvious avoidant source of image data rather than the widely popular trend-tracking images of Instagram reveals the deeply telling irony art student often face when creating their own art from the work of others. Originality in the digital era certainly brings the issue of ownership to question. This concept is especially relevant to our class, Art and Machine Learning, as many of our projects require us to seek external data sources in order to create novel artwork. We thus decided to use Shutterstock images of "salad" as our data set because these images serve as the canonically representative example of almost absurdly generic photos that are free to use only because of the garishly unignorable watermark defacing the images. The generation of a new set of images further deepens the irony because more stock images of salad is surely unnecessary and only serves to augment the already excessive supply of these uninspired images. However, the possibly formulaic nature of these images makes this data set particularly interesting in that the replication and reproduction of believable stock images of salad should be achievable through training with artificial neural networks. Though the concept underlying our work pushes the mind to think about the delicate balance between ethicality and legality when considering originality and ownership, the art is ultimately meant to be viewed with humor and appeal to the public by approaching this significant and weighty concept with whimsy in a light-hearted manner.

PROCESS

We decided to approach the issue of salad generation with two different methods. The first, explored by Jacqui, uses DCGAN to generate new salad images which are then textured and sharpened by Pix2Pix. The second, explored by Anirudh, uses Pix2Pix to generate images of salads from their outlines (or from the outlines of non-salad items). Both methods' processes are described separately below.

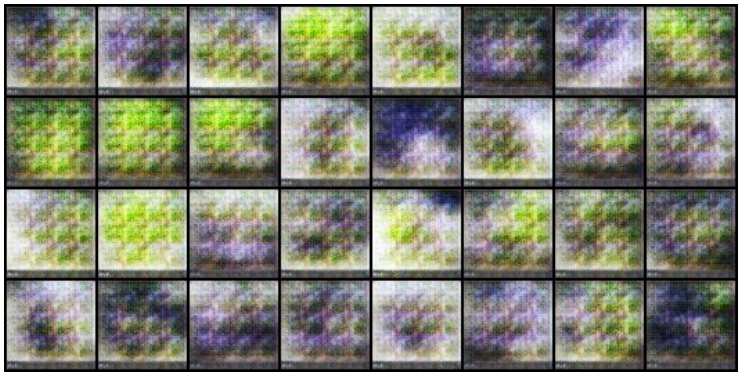
Method 1 - DCGAN:

I was a little enthusiastic to start, and decided to see what would happen if I trained a GAN on the small dataset Angela had manually collected (about 500 photos). After a little struggling with the formatting, I

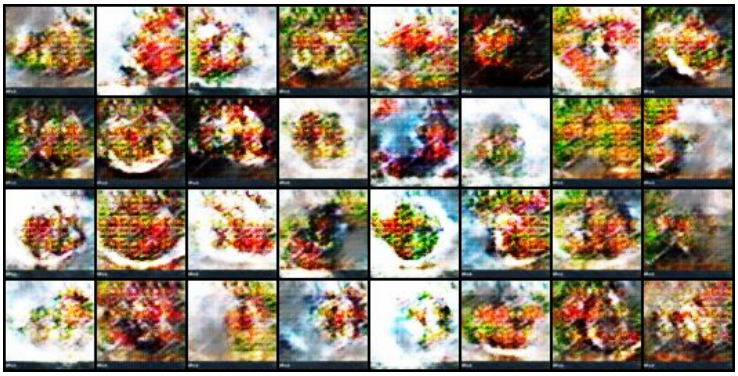
loaded this data into an AWS instance and trained a DCGAN model (the example notebook in the AMI, with minor adjustments) on it for 135 epochs. The results never got much better than those shown below, and after about 110 epochs started to overfit and look quite bad, in my opinion.

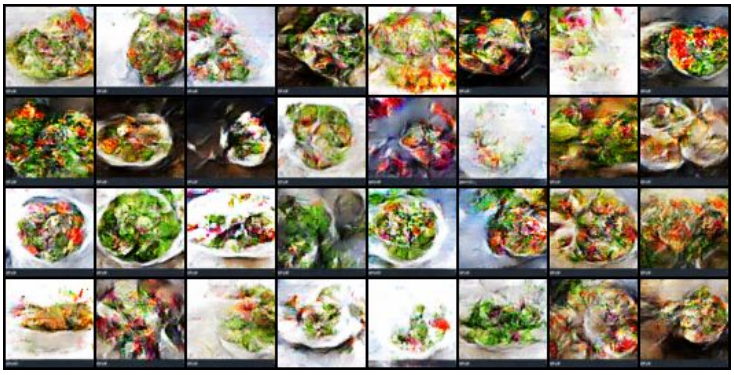
It became clear to me that to get reasonable results, I would need over one thousand images, so I wrote a short script to download all images on the first 30 pages of results for a search of the word “salad” on Shutterstock. This gave me about 3000 images, but many of them were sketches or other non-photos, or they featured a person more prominently than the salad, or they were otherwise not very good examples for our set. I briefly looked into using some kind of CV to filter these, but ended up doing it manually. This left me with about 2100 images, a decent dataset.

I fired up the notebook again, this time training it on the larger dataset for 100 epochs. These results were much better than before (supporting the idea “if it doesn’t work, just add more data”), and I was fascinated by how the generator evolved over time. I made a short video showing its progress, which can be seen here: https://www.youtube.com/watch?v=Q4mplQBRrfM. A few pictures of its progression are also below.



Epoch 6

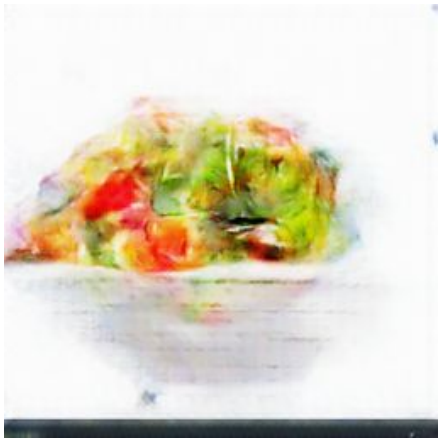

Epoch 27


Epoch 99

I liked these results pretty well, but the trouble was that they had too low resolution (only 64x64). I briefly experimented with making DCGAN generate larger images, but changing the network to do that turned out to be very non-trivial, and would require quadratically more data. I also considered simply scaling my images up, and interpolating/blurring them slightly so they wouldn't look as pixelated. However, these results did not impress me (see examples later in this section) and I thought I could do better.

So, I instead decided to try my hand with pix2pix. I wrote a script that would convert my high-res images of salads into low-res (64x64) versions, and used the combine_A_and_B python script that was already in the AMI to format the data as pairs of low-res and high-res images. I split this data into training and validation, and then trained the standard pix2pix notebook on this for only 20 epochs. The idea was to create a network that sharpened low-resolution images of salads, and added appropriate textures. Here is an example of that network in action (top row is input, second row is ground truth, bottom row is generator output):

After training this network, I gave it all of the images I had gotten from my DCGAN, and it polished them up. I think the extra effort of doing this pix2pix training was really worth it. Just compare its outputs to the upscaled and interpolated version of the same pictures!

Upscaled   Pix2Pix

Upscaled   Pix2Pix

The images made with pix2pix have much more well-defined edges, giving the impression that they are higher resolution, if still somewhat abstract. They also seems to be cropped slightly tighter, though I'm not sure whether this is pix2pix's doing, or a flaw in my formatting script. More results can be found in the "Result" section.

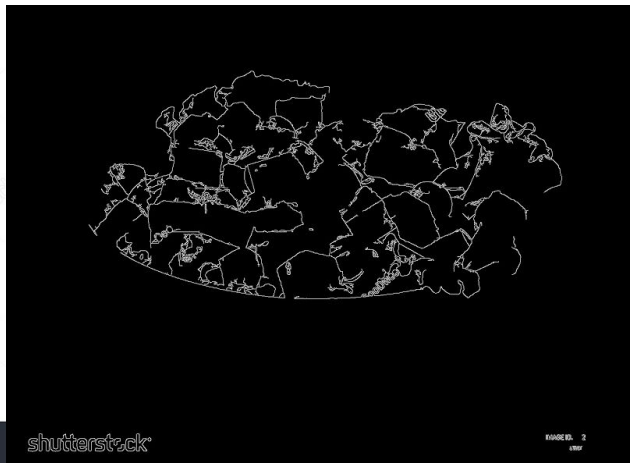
Method 2 - Pix2Pix:

The original idea was to take the salad stock images and convert them into their clustered versions, which would be posterized. However, the clusters obtained weren't good enough to show good posterized pictures as shown by the other dataset examples of pix2pix. The idea changed to using outlines of the salad in the images, and use them instead. Inspired by the pix2pix example which generates a cat from the example, I hoped to draw outlines of your favorite salad ingredients, and obtain some kind of a Shutterstock pictures of what we drew.

The first step was to detect outlines in the 2134 images I had from Shutterstock. Edge detection was performed on these images using OpenCV. I used the Canny edge detection algorithm and played around with the threshold values to obtain an optimal parameter values I found to give me good results. The objective was two-fold here,

1. Get edges of mainly the salad, and not too many as we wanted the GAN to be more robust on filling in the details
2. Get connected lines as much as possible, denoting continuous shapes

Another process which was tried was conversion of these RGB images to the HSV space, and then a mask was applied to get the greens of the salad image within a certain range, further edge detection on the masked image was performed. However, this approach didn't seem to work very well as well as a lot of these images had a lot more colors which weren't being represented well.

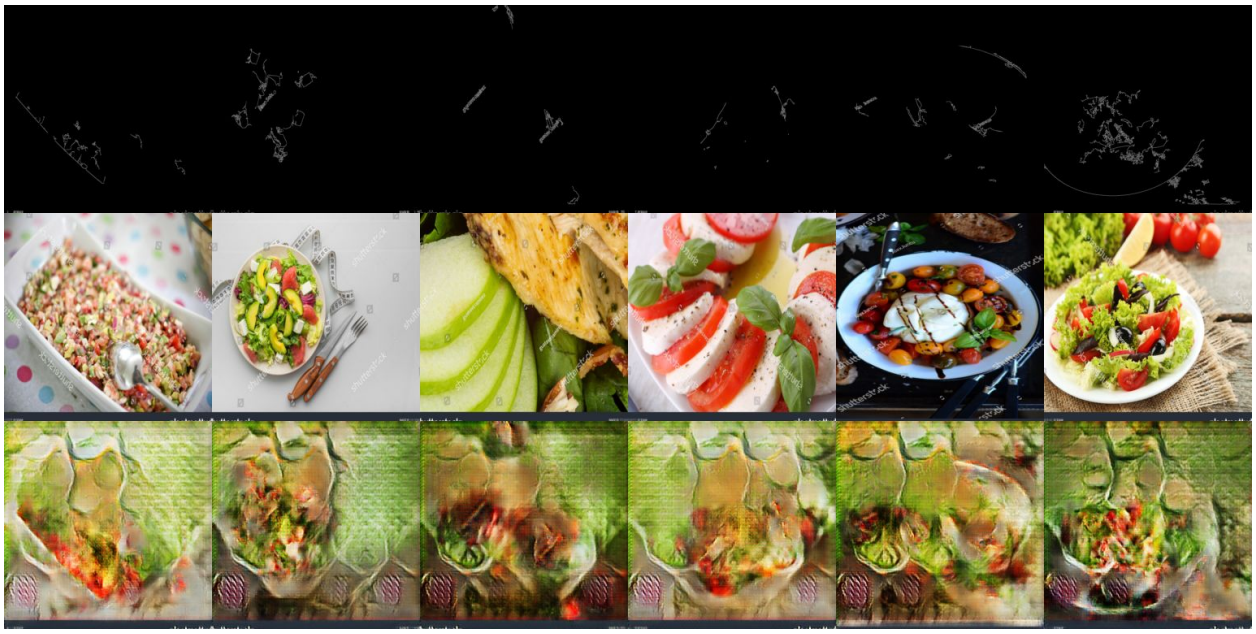


If we zoom in, we do see continuous lines, and this was our objective.

This data was then subsequently uploaded on the AWS Instance, where the provided pix2pix Keras implementation was trained on it. I tried the PyTorch version as I am more familiar with that, however, due to some dependency issues, it turned out to be way more time consuming than what I had initially planned for.

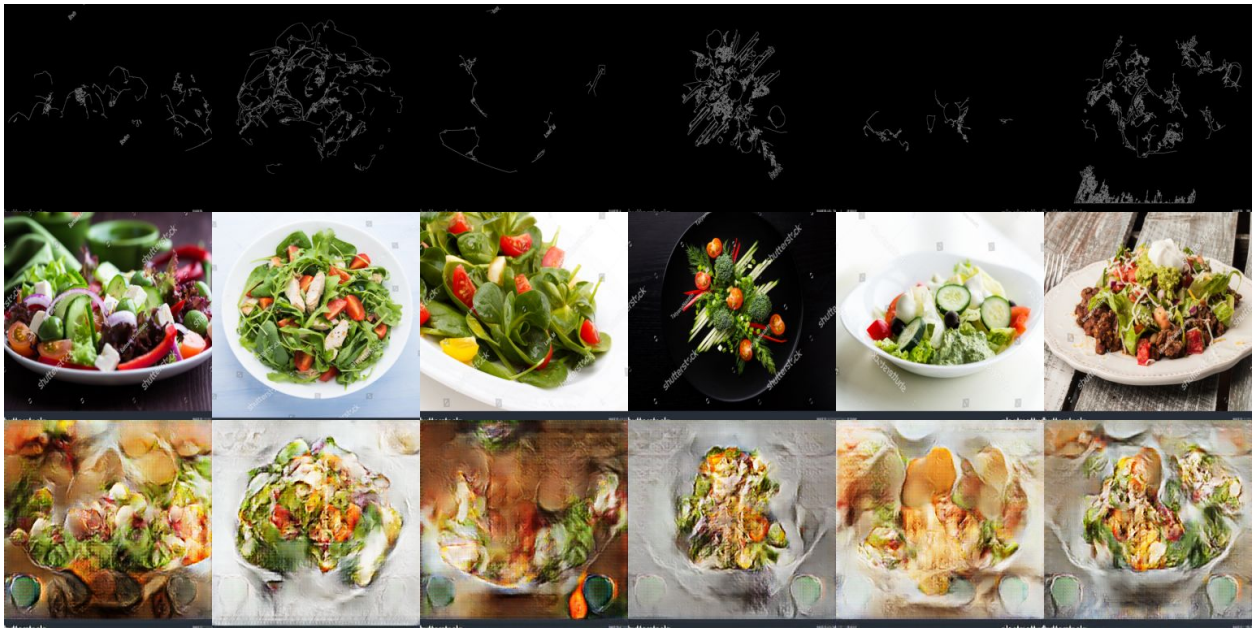
Here's the progress as per epoch,

Epoch 6:



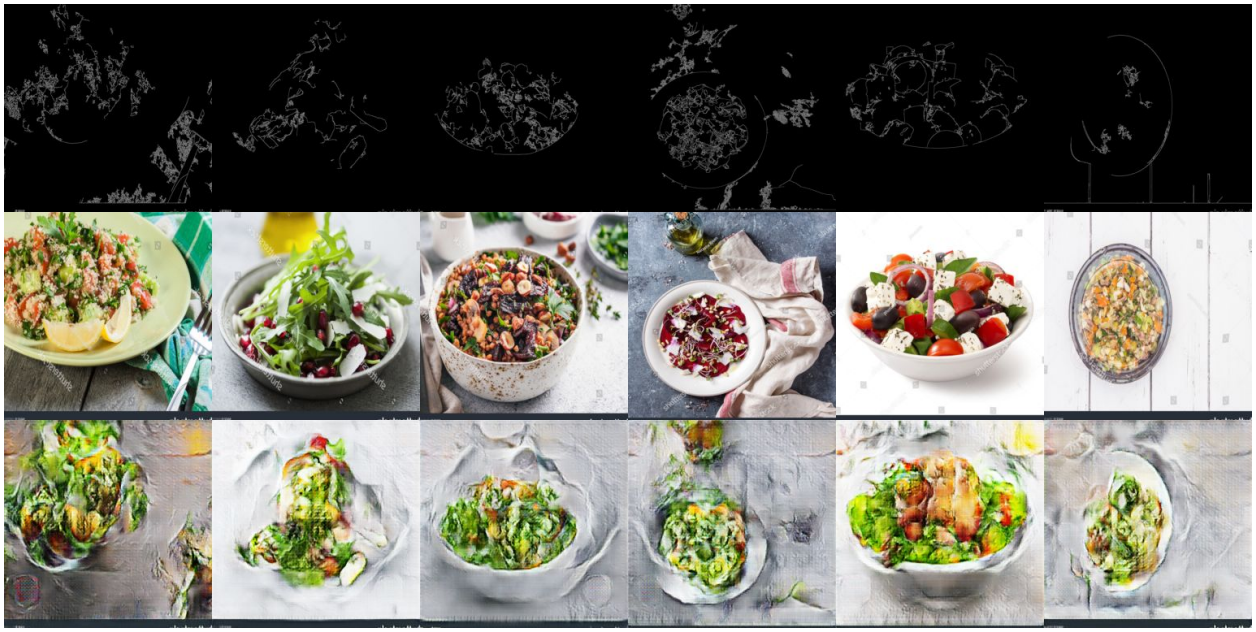
Here, as expected the fake image isn't as clear. However, an interesting progression to note is the background color of the image, of the table, and the utensil, etc. It gets better with every epoch, but the dataset contains a different variety of them, still the GAN trains to get more accurate with those.

Epoch 15:



Epoch 21:

Epoch 24:



Epoch 30:

I trained the GAN for 30 epochs, and after the training phase, the generated images were very close to the originals with respect to their background colors, however, the main content of the image itself - salad, was not as sharp. I believe it will only get better with more training, something I struggled to do with crashing kernels, and large training time.

RESULT

Method 1 - DCGAN:

Our results for this method are in the form of 512x512 “stock photos” of salads. Here are a few favorites:

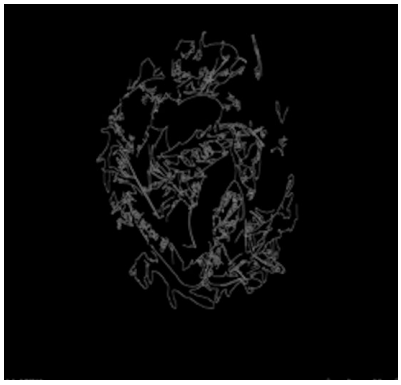
They can all be described as images of a single bowl of green salad sitting off-center on a plain background, a composition which is very common among Shutterstock photos of salad (or of anything, really). However, it’s worth noting that some of the generated images did not fit this description for one reason or another. Some pictures resembled fruit salad or potato salad more than traditional green salads (these alternative salad types were not excluded from the data set), while others looked like heads of lettuce without a bowl, or were impossible to identify as anything. Some examples of these

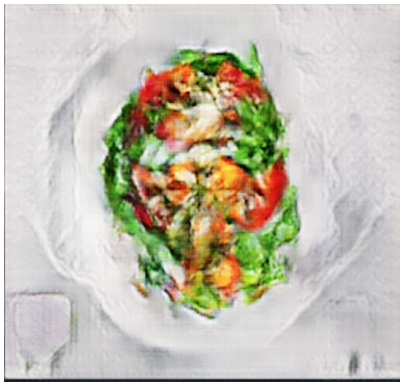
less-successful results are below. We are not really bothered by occasional “bad” outputs if we also get some very beautiful ones, which in this case we think we did.

As mentioned in the “Process” section, the pix2pix texturing really helps these images come to life. As with most GAN-generated pictures, they look a little distorted, but the texturing at least gives them some of the hard edges one would expect from a stock photo, and in the best examples it also makes the diagonal “shutterstock” watermarks a lot more visible. This is important because it brings back a lot of the themes of this project like copyright and ownership: even these images--which the network generated by itself--are saddled with a distracting watermark, reminding us of their origin and preventing them from being truly marketable.

Method 2 - Pix2Pix:

Here are some results after full training:





Our plan was to generate outlined images of random objects, and turn them into salad. However, this was relatively a new idea, and due to my kernel crashing, with the model unsaved, I couldn't generate those. We thought, it'd be a fun idea to see anything turn into salad.

It was very interesting to note how the GAN learnt to closely approximate the background colors, be it of the plate, or the table or any cutlery. The main content of these images, the salad itself needs to be sharper. We believe that can be achieved with more training, as the outlined dataset is good enough. This is something we plan to do even after this project.

REFLECTION

The project evolved extensively from its initial conception to the final result. The final result was chosen as a representative exemplar that both achieves the original goal at the outset of the project of generating a believable stock image of salad with the additional goal of producing a piece of artwork that is aesthetically appealing and provokes creative interest. Though the final result is certainly not an exact replica of a typical stock image of salad that may be found on Shutterstock, we believe it is a satisfying result for this project because of the relatively high degree of imitation that was achieved, particularly in replicating some of the most distinctive and recognizable elements of stock images (gray bar and watermark).

In regards to the DCGAN, we thought it was funny how early in the training the gray Shutterstock bar at the bottom of each image appeared. Before it knew it was making salads, it knew it was making stock photos! We are happy with the results, but we would have liked to try "uncropping" the photos as well. DCGAN can only generate square images right now, but most stock photos are not perfect squares. It would have been nice to generate rectangular images that more closely resembled the structure of a real stock photo, but we think that the style and content of the results are both quite nice, and we would definitely consider this method successful. Outline detection method worked well as well,

however its training is really time consuming, and we believe we can train it some more to get better results. It generates the salad images pretty from the outlines, and the outlines data works well too.

Additional ideas that arose throughout the process include an expansion of these techniques to subjects beyond salad. As mentioned earlier, a possible extension of this project is to try to generate a salad from the outline of any object. Another possible idea would be to generate stock images of other subjects with new training data. Women eating salads is another canonical stock image subject that adds complexity due to variation as well as fine-grained facial detail. Perhaps additional training data or a refinement of the methods attempted through this project would be able to produce high-quality images of a similar nature when applied to broader domains involving intricacies.

The project was a great learning experience for all of us as we were able to develop the seedling idea into a fully established concept with weighty meaning carrying implications beyond the classroom that are both highly relevant and significant. Unexpected difficulties along the way pushed the progress of the project forward and served as impetus for further innovation. We were thus able to explore a variety of machine learning methods to produce artwork that we believe is both visually and mentally stimulating. Our hope remains loyal to our original concept in bringing light to the topic of ownership and copyright when attempting to produce original and creative art work through our artistic interpretation of the stock image of salad.

COLLABORATION

The collaboration of our group was successful in that each member contributed his/her strengths to the project. The idea for the project was a product of combining the individual interests of each group member. Angela's focus was primarily on developing the concept and meaning of the art as well as much of the preprocessing and development work, while Jacqui and Anirudh focused more on method and implementation in producing the art. Jacqui primarily explored image generation through DCGAN and sharpening with pix2pix, while Anirudh worked on edge detection with OpenCV and image generation with pix2pix. Throughout the process, our group discussed developing ideas as the project progressed, and any differing perspectives were able to be resolved through these discussions. Overall, the work seemed to be distributed relatively evenly amongst the group such that each member was able to bring his/her strengths to creating the resulting art.