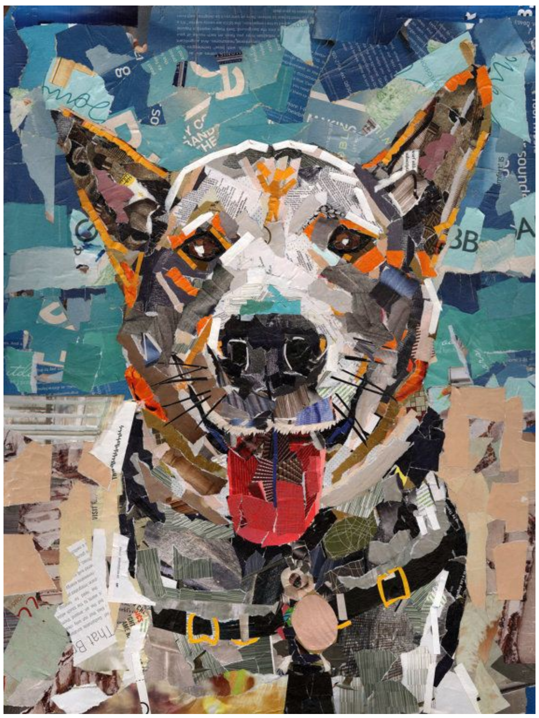
Capstone Proposal: Artistic collage-making

Isaac Wallis

Due to my deep interest in using machine learning for art, music, and other creative activities, I have decided to attempt to make, as my capstone, an algorithm that uses RL techniques to create photo-collages. In particular, I am interested in creating an algorithm that can take a number of images that are used as source material, cutting them up into various shapes, and placing them onto a canvas in such a way that they begin to look like a target image. A rough example of what I hope this algorithm will output is this image of a dog (assuming we use a photo of a dog as the target image)[1]:



I am not aware of previous algorithms that create this kind of photo-collage art, but ML is, of course, beginning to be used more and more frequently in the production of art. There are two previous approaches to algorithmic art-making which are particularly relevant to my work.

The first is image style transfer based on deep learning, where neural networks are used to extract the artistic style from one image, and then apply them over another image. My intuition tells me that this approach will not be very useful for my collage-making project. It is possible that, if the stylistic source images are carefully selected from existing works of collage art, that a rough approximation of a collage-art style could be transferred to some target image. But there are two obstacles, one practical and the other aesthetic. The practical obstacle is that deep learning algorithms require lots of data, and in this case that data would need to take the form of existing works of collage art in the appropriate style. I am not sure enough existing works can be found. The aesthetic obstacle is that paper-and-glue collages have the property that there is no exact duplication anywhere in the collage; each scrap of paper comes from a different location in the source material and therefore will have different text and imagery content visible. I don't think a CovNet-based style transfer algorithm would preserve this property; it is likely that the same source material would be used in multiple places throughout the collage.

The second relevant ML-based art generation technique that I have found is the work by Robert Johanssen[2] where an approximation of the Mona Lisa is generated by using Genetic Algorithms to place polygons on a canvas. I find this approach to be much more useful and similar to what I want to do. My proposed algorithm differs in that I want to use a combination of unsupervised learning techniques and reinforcement learning techniques, rather than a genetic algorithm with a similarity-based fitness measure.

1 Acquired from https://s-media-cache-ak0.pinimg.com/736x/32/8b/9d/328b9d83d240adca0ca10fea8b655297.jpg
2 https://rogerjohansson.blog/2008/12/07/genetic-programming-evolution-of-mona-lisa/

The Proposed Algorithm:

The inputs will consist of a set of *n* images to use as source material, plus a target image. The task of the algorithm is to cut shapes out of the source material, and place these shapes on a canvas in such a way that an approximate rendition of the target image, in collage form, is generated. In my proposed solution, the task of collage-making can be broken down into two steps. The first step is image segmentation, which must be applied both to the target image and to all the source images. This is essentially an unsupervised learning problem, accomplished through a clustering algorithm such as a modified k-means. Given an image, this algorithm must divide the image into *k* segments, where each segment is a contiguous patch that does not have too much internal color variance.

Once segmentation is accomplished, the set of all patches acquired from the source material shall be used as a palette which can be used to “fill in” the patches of the target image. This can be framed as a reinforcement learning problem. The agent takes actions, where the actions consist of choosing a target segment, then filling that target segment with a contiguous patch of pixels from a source segment. The reward for taking an action is based in part upon a similarity metric computing the similarity between the canvas and the target image. Therefore the learner is trying to find a state (where the state is just a mapping of source material onto target segments) that maximizes similarity between collage and target.

The similarity measure used to set the reward for actions will also serve as the main evaluation metric for the final collage. I will need to try a few different evaluation metrics to find one that works the best for my purposes. The simplest I can think of might be to compare the color histograms across each target segment. The most advanced might be some modification of Levenshtein distance into 2-D space across pixels, which, if applied as a reward function, might do a better job of placing source material that is texturally appropriate as well as appropriate in terms of color. I see that CovNets have been used to recognize visual texture in the past[3], so I may experiment with incorporating them into my similarity metric in some way as well.

Once I have an appropriate similarity metric, I can construct benchmark images to compare to the target image. The first, bad, benchmark will be an image of gaussian noise in RGB, where the distribution of pixels matches the histogram of the target image. Its similarity with the target image should be quite low and very easy to beat. The next benchmark image will be the target image after image segmentation is applied, where each segment is painted with the mean segment color. Beating this benchmark will require good selection of source material segments. Lastly, if I have time (as this would be a very advanced step for me), it might be worthwhile to try to construct a CovNet for style transfer, attempting to transfer a collage style onto the target image. If I manage to do this, it would be interesting to use the output of this style transfer algorithm as a third benchmark.

Since this project is artistic in nature, evaluating the collage algorithm itself—as opposed to evaluating the resulting collages in terms of the target image, which has already been discussed—is a bit of an impossible challenge. Art is aesthetic and subjective, and it is hard to say whether collage A is “better” than collage B in quantifiable terms. In general, if the algorithm generates collages that evoke the target images, in reasonable time-frames, then the algorithm is useful as an artistic tool.

---

3 https://www.robots.ox.ac.uk/~vgg/publications/2015/Cimpoi15a/cimpoi15a.pdf

Project Design Steps:

1. Implement segmentation algorithm, with a way to display each segment.
2. Experiment with similarity measures that can be applied per-segment and across the entire image. Select one.
3. Write a reinforcement learner that will map the source segments onto the target segments according to rewards derived from the similarity measure. This learner will need to take into account the following things when placing segments:
    c) Source segments should not be rescaled. If there is no source segment which is large enough to fill the target segment, the target segment should be divided into smaller segments.
    d) Source segments can be rotated, and the place in the source segment that provides the pixels for the target segment can be moved around(in other words, the source segment can be translated too).
    e) Source pixels that are placed on the collage already cannot be placed in duplicate elsewhere on the collage. This limitation makes the system more Markovian, in the sense that the order in which target segments has impacts further down the road.
4. Using grid search, optimize the parameters of this reinforcement learner in order to minimize training time while maximizing similarity.
5. As an optional step: Since the final collage may need a little tweaking according to the user's aesthetics, it would be good to give the user the ability to manually change the segment mappings after the fact.