AI in Art: Simulating the Human Painting Process

Alexander Leiser and Tim Schlippe(✉)

IU International University of Applied Sciences, Bad Honnef, Germany
tim.schlippe@iu.org

Abstract. While AI is being used more and more to generate images, the generation usually does not resemble a human painting process. However, for applications in the field of art, it is useful to simulate the human painting process—e.g. in relation to location, order, shape, color and contours of the areas being painted in each step. Such applications are for example when a robot paints a picture or a program teaches humans to paint. Consequently, in this paper we evaluate and compare different approaches to simulate the human painting process. Additionally, we present our solution for this task which is based on a combination of filters and semantic segmentation. In our survey, this approach was rated as better and more realistic than the most realistic approach for this task so far which is a reinforcement learning approach: In all surveyed categories—location, order, shape, color and contours of the areas being painted in each step—always a significant majority of the participants prefers our approach to simulate the human painting process. When we displayed two time-lapse videos with the painting process of Edvard Munch's The Scream in parallel, even 79% found our generated process more realistic than the reinforcement learning-based process.

Keywords: AI art · AI in Art · Painting · Semantic Segmentation · Semantic Labeling · Machine Learning

1 Introduction

Painting is one of the most basic and oldest forms of art. Early findings of simple rock paintings are dating back almost 40,000 years [1]. While the general outcome of this art form is usually depicting a person or an object on a medium like paper or canvas, painters have developed different painting styles, techniques and methods over the past centuries to achieve this goal [2]. Since painting itself is a very visual form of art, the appeal of paintings can, at least in part, be conveyed through modern technologies. This also allows for simulation of the painting process as well as automated interaction with a brush, the main tool in this art form. Thus painting found its way into modern technologies such as AI and robotics. There are several examples of applications in this field:

M. Wölfel et al. (Eds.): ArtsIT 2021, LNICST 422, pp. 295–308, 2022.
https://doi.org/10.1007/978-3-030-95531-1_20

A team of scientists and engineers from IBM Japan, the University of Tokyo, and Yamaha Motors equipped an industrial robot with a camera and a paintbrush to explore the realms of creativity in machines and AI [3]. Other teams such as *AI_NORN*[1] and *cloudpainter*[2] also experiment with AI art, more specifically painting, to explore the outcomes when machines, which are capable of handling a paintbrush, are combined with modern AI technologies. The company Nvidia has a dedicated AI Art Gallery on their homepage[3] to show and support art projects which are generated or supported by AI. However, most of the works do not simulate a realistic human-like painting process or their simulation still has shortcomings. Consequently, the focus of this paper is on the simulation of the human painting process.

In the next section, we describe the painting process in more detail. In Sect. 3, we present the latest approaches of other artists and researchers. Section 4 characterizes our approach to simulate the human painting process. Section 5 describes our survey and the feedback on the approaches. We conclude our work in Sect. 6 and suggest further steps.

Fig. 1. Human painting process with a layering strategy [4].

1 https://ainorn.art
2 https://www.cloudpainter.com
3 https://www.nvidia.com/en-us/deep-learning-ai/ai-art-gallery

2 The Human Painting Process

The painting process of humans usually follows a specific pattern from coarse to fine details. While there are many different techniques in painting which largely depend on the paint and medium used [5,6], many artists use the common layering strategy to paint objects with a background [7]. This is a widely used process especially for realistic paintings like portraits or still-life images. Our work will focus on the common layering strategy described in [7]. As shown in Fig. 1, in this strategy one layer is painted over another, until the final picture is created. The layering process starts with a uniform background and then gradually adds details in size decreasing regions until all the desired details are contained in the painting.

With the simulation of such a layering process all kinds of images could be created—from very abstract images to very realistic detailed images. The focus of this paper is on the creation of images which contain one or more objects and have a background in relation to the aspects location, order, shape, color and contours of the areas being painted in each step.

3 Related Work

Several studies and practical machine learning based approaches to replicate or generate images and paintings and to simulate the process of painting itself have been conducted in the past. While some methods capture the general picture and thus the outcome of the painting process very well, they often fail to reproduce the process itself in a human-like manner. Examples are [8–11].

[12] use a generic algorithm which resembles images well with the help of certain constraints. However, the painting process is not human-like since it uses semi-transparent polygons to approximate the original painting as close as possible. [6] uses a differential renderer in combination with a generative adversarial network to generate human-like brushstrokes. While the resulting brushstrokes and pictures have a human-made appeal, the process of generating them is not in the way most humans would paint a picture since the intermediate images show random brush strokes which do not seem to help achieve the end result of the picture. An example are arbitrary purple and yellow brushstrokes in the middle of the image, where in the final image is a snow-covered mountain that contains white and blue tones.

To the best of our knowledge, there is no database which contains enough images of intermediate steps of the human painting process to train generative adversarial networks for more human-like simulations. Google's "Quick, Draw!" [13] database contains 50 million drawings across 345 different categories to train neural networks for the human doodling process. However, the images are mostly in the form of simple doodles. Since the data set only contains the vectorial information of doodles with their start points, end points and tractorial information, the images have a limited resolution and are in only one color, the dimensionality is too low for the simulation of a full painting process. For example, [14] used the collected data of sheep doodles to generate 10,000 sheep published in the book "Dreaming of Electric Sheep". The lack of training images with intermediate steps of the painting process is one reason why we did not choose generative adversarial networks or reinforcement learning approaches but the leaner approach, described in Sect. 4.

[15] trained a convolutional neural network with 117 collected, 4-minute long time-lapse videos of real and digital paintings, to synthesize the time-lapse video of new paintings. While the algorithm outputs decent time-lapse videos, it is not suitable for the simulation of the human painting process: As visualized in Fig. 2, the transitions between the different painting stages are blurry and do not resemble the single painting process steps. Furthermore, in a step different colors simultaneously appear in different regions of the image.

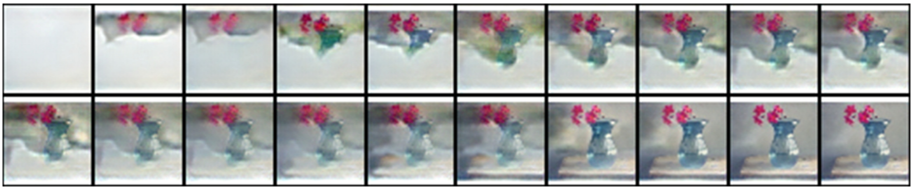


Fig. 2. Intermediate screenshots from synthesised time-lapse videos based on [15].

[16] and [17] apply reinforcement learning for the sequential decision-making. In reinforcement learning usually an agent is programmed to interact with an environment and improves its interactions with the given feedback coming from this environment [18]. Whereas [16] also focuses rather on the final result than on the actual process of generating the image, [17]'s approach tries to imitate the human painting process. [17] is close to the idea of [6] but captures the underlying picture sooner in a more human-like manor. The order of the brush strokes is not always human-like or intuitive. Nonetheless [17] serves as a good baseline since the results resemble the human painting process the closest of all described methods. Consequently, this approach was also evaluated in our survey for comparison with our own method.

4 Simulating the Human Painting Process Using Filters and Semantic Segmentation

While other methods focus on the resulting image or use not-so-human-like shapes or colors in the step-by-step expansion of the image during the painting process, our goal was to develop an algorithm that represents the human painting process as realistically as possible. Particularly with regard to the location, order, shape, color and contours of the areas being painted in each step, the generated painting process should be coherent. Additionally, it was important to us that we did not need huge amounts of image or video material of the painting process for training our computer vision models as is often the case with deep learning algorithms. With these conditions in mind, we developed an algorithm which is modular and thus flexible, which is lean, can be set up quickly, and emulates the layering process of a painter. Our simulation of the human painting process consists of the following components and steps, which are also visualized in Figs. 3, 4 and 5:

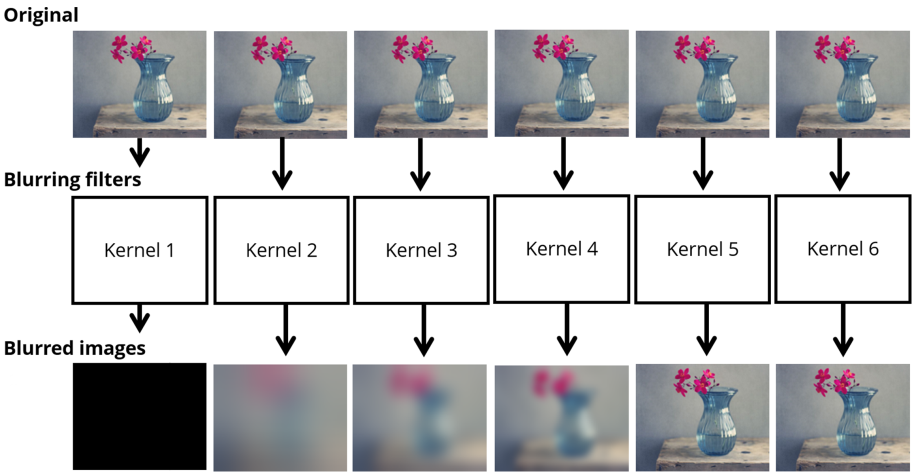


Fig. 3. Simulating the human painting process: Blurring filters.

1. *Blurring filters*: With the goal of coloring large areas first, the image is first blurred with various filters. As demonstrated in Fig. 3, the goal is to dilute the edges and colors in the image to different degrees so that the segmentation algorithm in step 2 outputs a different number of segments based on the details to be detected in the image. For our experiments we used 5 Gaussian filters with different kernel sizes to generate 5 blurred images. The implementation was done with OpenCV [19].
2. *Semantic segmentation*: We apply a semantic segmentation algorithm to the images blurred with different degrees to obtain smaller and smaller areas to be painted. As visualized in Fig. 4, the retrieved segments are given the color which has the most occurrence in that segment in the original image (*coloring*). For our experiments, we applied the unsupervised convolutional neural network based semantic segmentation described in [20] which minimizes similarity loss and spatial continuity loss to each blurred image.
3. *Stepwise adding colored areas*: Our goal is to add painted areas step by step. To avoid reapplying colors already applied in the painting process in the same place, we remove the areas that have the same color as the image on the left as shown in Fig. 5. Individual images are then created from the different color areas, with each new image corresponding to a step in the painting process,

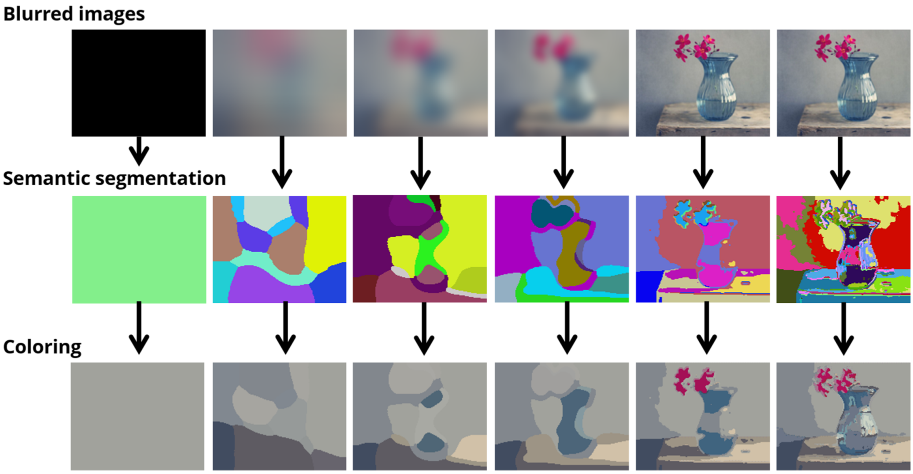


Fig. 4. Simulating the human painting process: Semantic segmentation.

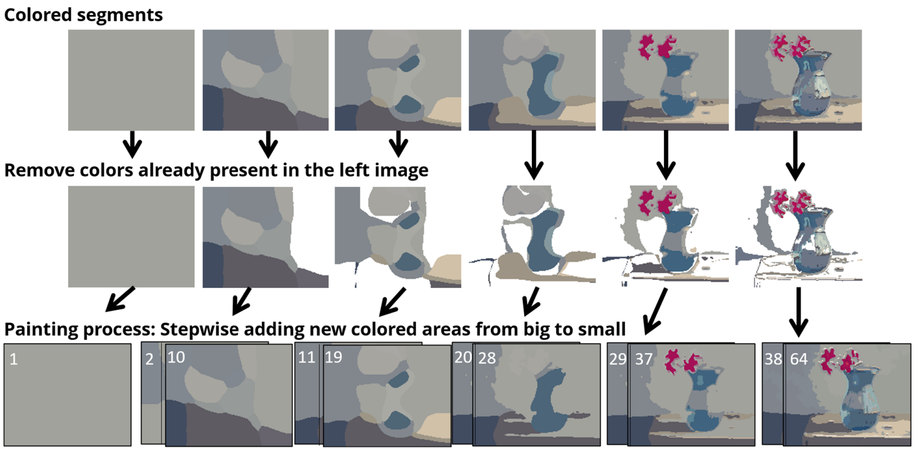
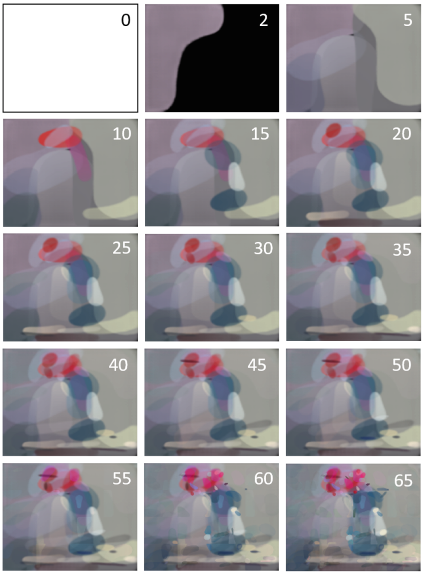


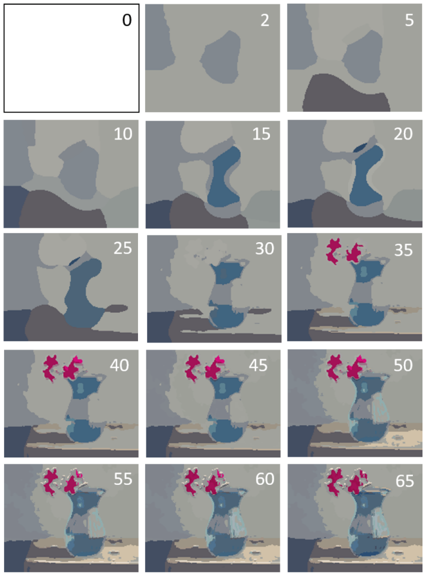
Fig. 5. Simulating the human painting process: Stepwise adding colored areas.

such as adding only one color to an area. The individual images are sorted in such a way that the images with the large color areas come first before the images with the smaller color areas.

Our approach requires no pre-training. It also eliminates the need for a pre-trained neural renderer to make the training process possible in a suitable amount of time as used by [20]. Due to the slim implementation, the algorithm can entirely be used locally or in cloud services such as Google Colab [21]. Single components like the *blurring* step and the *semantic segmentation* step can be replaced or extended. Images are generated in each intermediate step and can be extracted for further or other use. This makes our presented approach really versatile and accessible for future improvements. For example, in our experiments we used 5 Gaussian filters with different kernel sizes to generate 5 blurred images. But the number and the strength of the blurring effect could also be calculated based on the number of motives or the level of detail in the image. As illustrated in Fig. 6, our method paints in regions to slowly fill the canvas similar to the layering painting technique introduced in Sect. 2 instead of making semi-transparent brush strokes in seemingly random areas and combine them to the target image.



(a) [17]'s reinforcement learning approach.



(b) Our approach.

Fig. 6. [17]'s reinforcement learning approach vs. our approach.

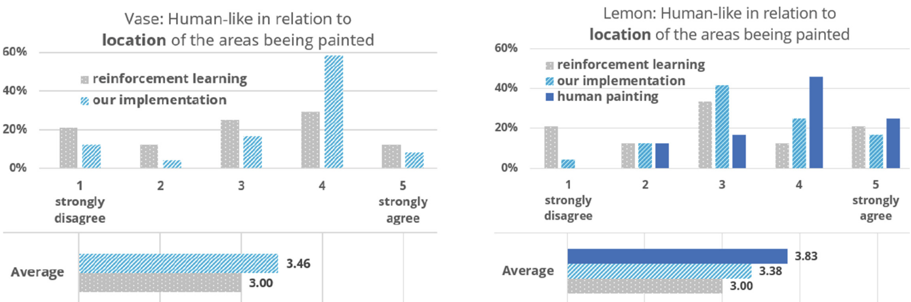
5 Experiments and Results

5.1 Experimental Setup

We evaluated our approach in comparison to the most realistic approach [17] (*reinforcement learning*) for simulating the human painting process in a survey. The study examined the location, order, shape, color and contours of the areas being painted in the intermediate steps. The progress of the painting processes was demonstrated as a time-lapse video and an image sequence for each of 3 pictures—a vase, a lemon, and Edvard Munch's The Scream. Of course, we did not tell the participant what painting processes they were shown. For the picture with the lemon, we also conducted a Wizard of Oz experiment: In addition to the two AI-based processes, we asked questions about the painting process that a human had actually performed, without telling the participants. For the pictures of the vase and the lemon, the participants were always shown only one simulation per page in the questionnaire, so that they could rate one approach without the influence of another approach. However, in order to also evaluate the direct comparison, for Edvard Munch's The Scream we displayed two time-lapse videos in parallel with the reinforcement learning approach and our approach. The participants evaluated most questions with a score. The score range follows the rules of a forced choice Likert scale, which ranges from (1) *strongly disagree* to (5) *strongly agree*. 24 people (14 female, 10 male) filled out our questionnaire. The participants of our user study were randomly selected volunteers between 19 and 71 years old who participated free of charge. The participants' painting routine varies from once a week to once a year or even never. Most people indicated that they are interested in art, but there are also some who are not interested in art. We appreciate these distributions as it was important to us to get feedback from different people.

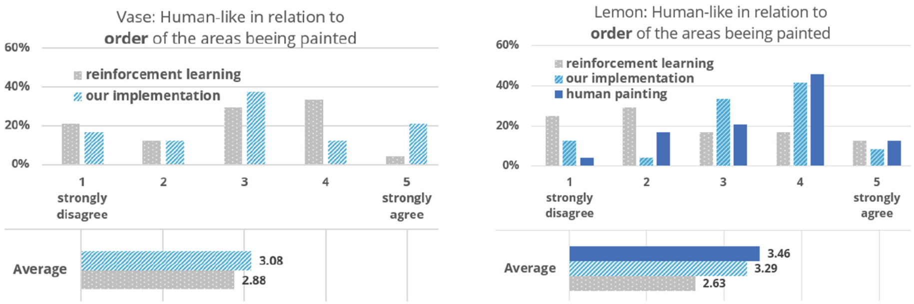
5.2 Evaluation in Relation to the Location of the Areas Being Painted

We asked the participants in our questionnaire how human-like they find the painting processes in relation to the location of the areas being painted. The goal was to find out whether the painting progress always happens in the right place. Figure 7 illustrates the feedback on the location for the vase and the lemon. While *reinforcement learning* was rated on average with 3.00 for *vase* and *lemon*, *our implementation* was rated better with 3.46 (*vase*) and 3.38 (*lemon*) on average. Thus, *our implementation* with regard to the location is rated 15% better for the vase and 13% for the lemon than *reinforcement learning*. The Wizard of Oz painting process *human painting* wins with an average of 3.83.



(a) Painting a vase. (b) Painting a lemon.

Fig. 7. Feedback on the location of the areas being painted.

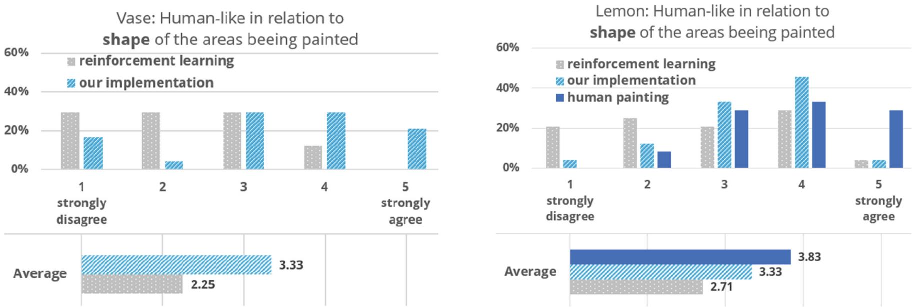


(a) Painting a vase. (b) Painting a lemon.

Fig. 8. Feedback on the order of the areas being painted.

5.3 Evaluation in Relation to the Order of the Areas Being Painted

Figure 8 illustrates our evaluation in relation to the order of the areas being painted. Overall, this evaluation is a bit worse than the evaluation of the location. But here, too, our method is rated on average between *reinforcement learning* and *human painting*: While *reinforcement learning* was rated with averages of 2.88 (*vase*) and 2.63 (*lemon*), *our implementation* was rated better with 3.08 (*vase*) and 3.29 (*lemon*) on average. This means that *our implementation* with regard to the order is rated 7% better for the vase and 25% for the lemon than *reinforcement learning*. *Human painting* outperforms the other approaches with an average of 3.46.

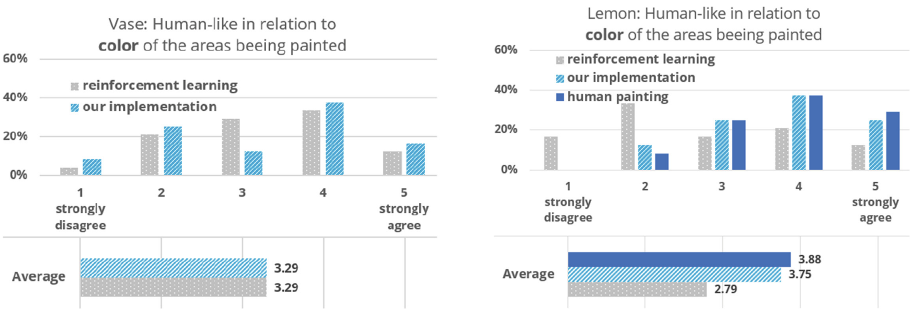


(a) Painting a vase. (b) Painting a lemon.

Fig. 9. Feedback on the shape of the areas being painted.

5.4 Evaluation in Relation to the Shape of the Areas Being Painted

In the category *shape* we are significantly better than *reinforcment learning*: As shown in Fig. 9, the question if the shape of the areas being painted is human-like was rated only with an average score of 2.25 (*vase*) and 2.71 (*lemon*). *Our* *implementation* was rated better with 3.33 on average for *vase* and *lemon*. Comparing the scores shows the significance of the shape: *Our implementation* with regard to the shape is rated 48% better for the vase and 23% for the lemon than *reinforcement learning*. *Human painting* again performs best, in this category with an average of 3.83.



(a) Painting a vase. (b) Painting a lemon.

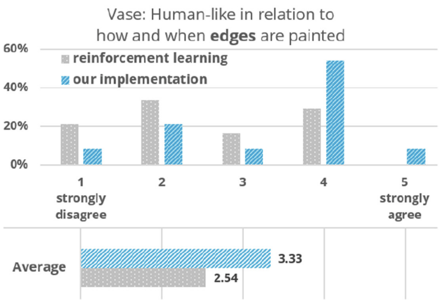
Fig. 10. Feedback on the color of the areas being painted.

5.5 Evaluation in Relation to the Color of the Areas Being Painted

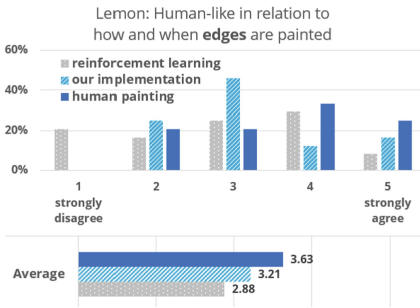
Then we asked the participants in our questionnaire how human-like they find the painting processes in relation to the color of the areas being painted. The results are demonstrated in Fig. 10: While for *vase* both approaches *reinforcement learning* and *our implementation* were rated 3.29 on average, *reinforcement learning* performed with 2.79 and *our implementation* with 3.75 on average for *lemon*. Thus, *our implementation* with regard to the color is rated equal for the vase but for the lemon 34% better than *reinforcement learning*. *Human painting* outperforms the other approaches again, this time with an average of 3.88.

5.6 Evaluation in Relation to How and When Edges Are Painted

The final aspect which we evaluated was how human-like the painting process is in relation to how and when edges are painted. As illustrated in Fig. 11, the trends are as in the other aspects: For the *reinforcement learning* the question was rated with an average score of 2.54 (*vase*) and 2.88 (*lemon*). *Our implementation* was rated significantly better with 3.33 (*vase*) and 3.21 (*lemon*) on average. This means that *our implementation* with regard to the edges is rated 31% better for the vase and 12% for the lemon than *reinforcement learning*. *Human painting* again performs best, in this category with an average of 3.63.

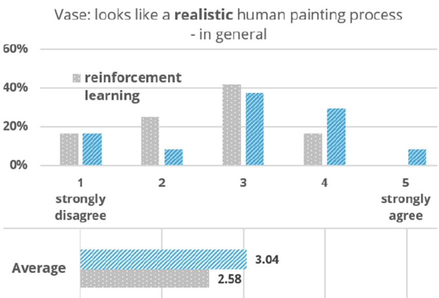


(a) Painting a vase.

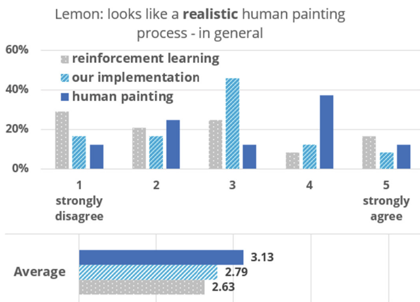


(b) Painting a lemon.

Fig. 11. Feedback on how and when edges are painted.



(a) Painting a vase.



(b) Painting a lemon.

Fig. 12. Feedback on how realistic the human process is in general.

5.7 Evaluation of the General Realistic Look

In the previous questions we asked about the individual aspects in order to analyze the strengths and weaknesses of the procedures. But we also wanted to use the questionnaire to find out how human-like the participants find the procedures in general. As visualized in Fig. 12, for the *reinforcement learning* the question was rated with an average score of 2.58 (*vase*) and 2.63 (*lemon*), whereas *our implementation* was rated with 3.04 (*vase*) and 2.79 (*lemon*) on average. This shows that the general impression of *our implementation* is by 18% and 6% better. *Human painting* achieves an average of 3.13.

In order to also evaluate the direct comparison, for Edvard Munch's The Scream we displayed two time-lapse videos in parallel with the *reinforcement learning* approach and *our implementation*. Figure 13 indicates that here, too, participants find our implementation significantly more human-like in each single aspect and also in general. Even 79% find *our implementation* more realistic than *reinforcement learning*.

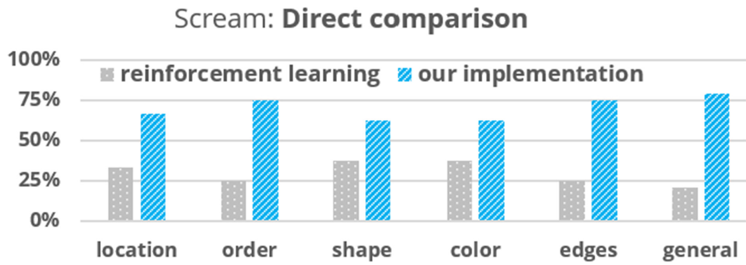


Fig. 13. Direct comparison of location, order, shape, color, edges and in general for the paining process of Edvard Munch's The Scream.

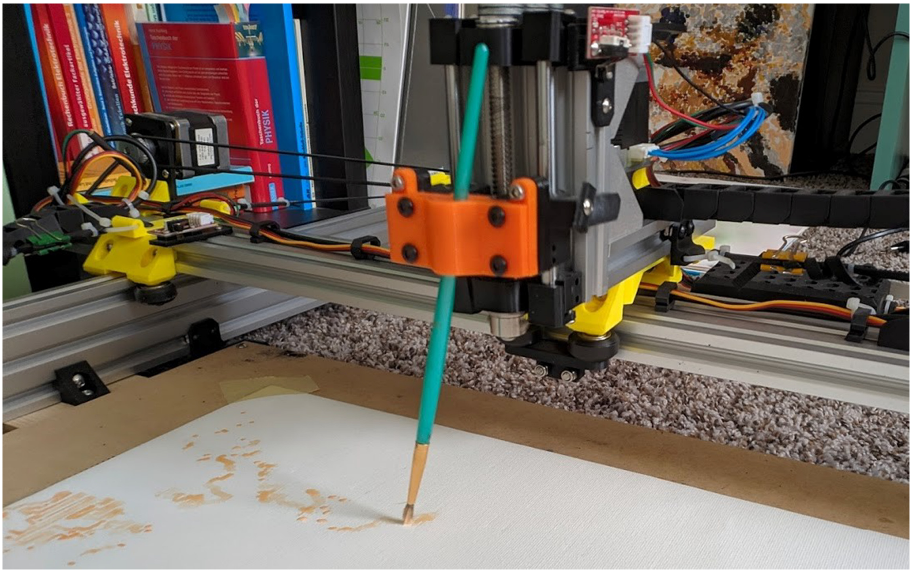


Fig. 14. Painting robot.

6 Conclusion and Future Work

In this paper we have evaluated and compared different approaches to simulate the human painting process. Additionally, we presented our solution for this task which is based on a combination of filters and semantic segmentation. In our survey, this approach was rated as better and more realistic than the most realistic approach for this task so far which is a reinforcement learning approach. When we look at the individual aspects evaluated, we perform best on the shape of the areas being painted. Our advantage over other methods is that the use of filters and semantic segmentation does not approximate shapes, which would be necessary in other deep learning approaches to limit the parameter space [12,17].

Future work may include the combination of our implementation with a robot which possesses the equipment to paint or software which teaches how to paint. Figure 14 demonstrates our painting robot Paintbot[4]. The painting robot is similar to a CNC machine and built in a standard 3-axis design, focusing on a large and flat work area to paint canvases. The machine has 60 different acrylic paints directly accessible. It also has a sponge, water reservoir and a cloth for cleaning the brush between color changes. The idea is to use our algorithm to generate the intermediate images of the painting process which will serve as input of the G-code generator software to control the robot and paint the image in a human-like manor.

Our approach is suitable for paintings that clearly have coarse to fine levels (e.g., those with a clear separation of background and foreground). Future studies may explore how to treat other types of paintings, e.g., those with local fine structure and no global arrangement, or—the other way around—with global structure but no local details. To refine our algorithm we plan to add edge detection to draw outlines after painting the background as demonstrated in the second step of the human painting process in Fig. 1. Having achieved that the image is divided into separate painting areas/layers, we could further divide these areas into individual brush strokes. When combining our algorithm with our painting robot, this would be done implicitly by the painting robot, but not simulated beforehand in our algorithm. Whereas the goal of our study was to evaluate the feedback of people with different interest and experience in art, the perception of people with more experience compared to those with less experience can be investigated in future studies.

References

1. Aubert, M., et al.: Pleistocene Cave Art from Sulawesi. Indonesia Nat. **514**, 223–227 (2014)
2. Driscoll, S.: Painting. Salem Press Encyclopedia (2019)
3. earthryse: An AI-based Robot that Creates Fine Art Paintings (2021). https://earthryse.prowly.com/130623-an-ai-based-robot-that-creates-fine-art-paintings. Accessed 24 May 2021
4. 3KICKS fine art studio: Sean Cheetham's Demo in Advanced Portraiture Class (4/25/11) (2011). http://3kicks.blogspot.com/2011/05/sean-cheethams-demo-in-advanced.html. Accessed 24 May 2021
5. Durani, B.: Acrylic Painting Techniques: A Series of Nature Themed Acrylic Paintings. Ph.D. thesis, Yeshiva College, Yeshiva University (2020)
6. Nakano, R.: Neural Painters: A Learned Differentiable Constraint for Generating Brushstroke Paintings. ArXiv abs/1904.08410 (2019)
7. Reyner, N.: How to paint with layers - in acrylic and oil (2017). https://nancyreyner.com/2017/12/25/what-is-layering-for-painting/. Accessed 26 May 2021

[4] https://vialps.com

8. Singh, J., Zheng, L.: Combining Semantic Guidance and Deep Reinforcement Learning for Generating Human Level Paintings. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 16387–16396, June 2021
9. Kotovenko, D., Wright, M., Heimbrecht, A., Ommer, B.: Rethinking Style Transfer: From Pixels to Parameterized Brushstrokes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 12196–12205, June 2021
10. Zou, Z., Shi, T., Qiu, S., Yuan, Y., Shi, Z.: Stylized Neural Painting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 15689–15698, June 2021
11. Liu, S., et al.: Paint Transformer: Feed Forward Neural Painting with Stroke Prediction. CoRR abs/2108.03798 (2021). https://arxiv.org/abs/2108.03798
12. Johansson, R.: Genetic Programming: Evolution of Mona Lisa (2008). https://rogerjohansson.blog/2008/12/07/genetic-programming-evolution-of-mona-lisa. Accessed 24 May 2021
13. Google Creative Lab: The Quick, Draw! Dataset. https://github.com/googlecreativelab/quickdraw-dataset (2017), accessed: 2021–05-24
14. Diaz-Aviles, E.: Dreaming of Electric Sheep (2018). https://medium.com/libreai/dreaming-of-electric-sheep-d1aca32545dc. Accessed 24 May 2021
15. Zhao, A., Balakrishnan, G., Lewis, K.M., Durand, F., Guttag, J., Dalca, A.V.: Painting Many Pasts: Synthesizing Time Lapse Videos of Paintings. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8432–8442 (2020)
16. Ganin, Y., Kulkarni, T., Babuschkin, I., Eslami, S.M.A., Vinyals, O.: Synthesizing Programs for Images using Reinforced Adversarial Learning. In: Dy, J.G., Krause, A. (eds.) Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10–15, 2018, Proceedings of Machine Learning Research, vol. 80, pp. 1652–1661. PMLR (2018)
17. Huang, Z., Zhou, S., Heng, W.: Learning to Paint with Model-based Deep Reinforcement Learning. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 8708–8717 (2019)
18. François-Lavet, V., Henderson, P., Islam, R., Bellemare, M.G., Pineau, J.: An Introduction to Deep Reinforcement Learning. Found. Trends Mach. Learn. **11**(3–4), 219–354 (2018)
19. Culjak, I., Abram, D., Pribanic, T., Dzapo, H., Cifrek, M.: A Brief Introduction to OpenCV. In: 2012 Proceedings of the 35th International Convention MIPRO, pp. 1725–1730 (2012)
20. Kim, W., Kanezaki, A., Tanaka, M.: Unsupervised Learning of Image Segmentation based on Differentiable Feature Clustering. IEEE Trans. Image Process. **29**, 8055–8068 (2020). https://doi.org/10.1109/TIP.2020.3011269
21. Pessoa, T., Medeiros, R., Nepomuceno, T., Bian, G.B., Albuquerque, V., Filho, P.P.: Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications, p. 1. IEEE Access (2018)