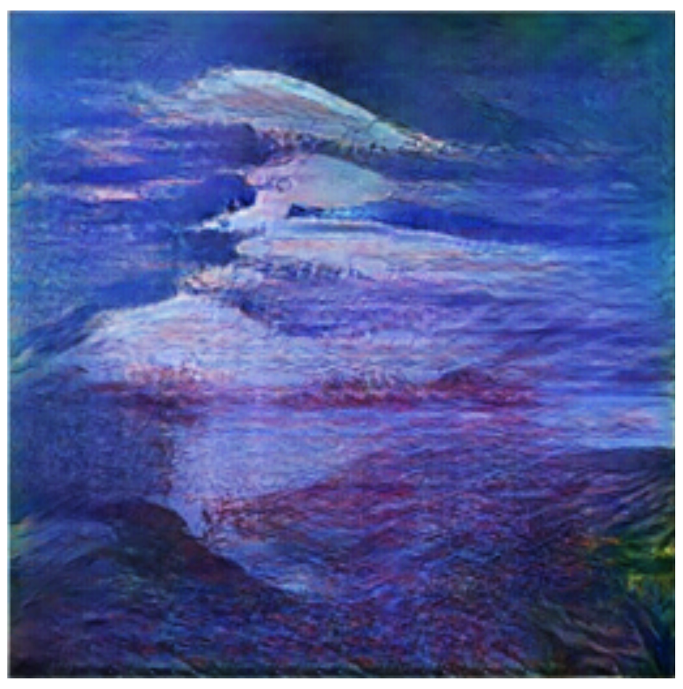
TOWARDS A PRINCIPLED EVALUATION OF MACHINE-GENERATED ART

Lia Coleman[1], Panos Achlioptas[2], Mohamed Elhoseiny[1,2]

1. KAUST 2. Stanford University

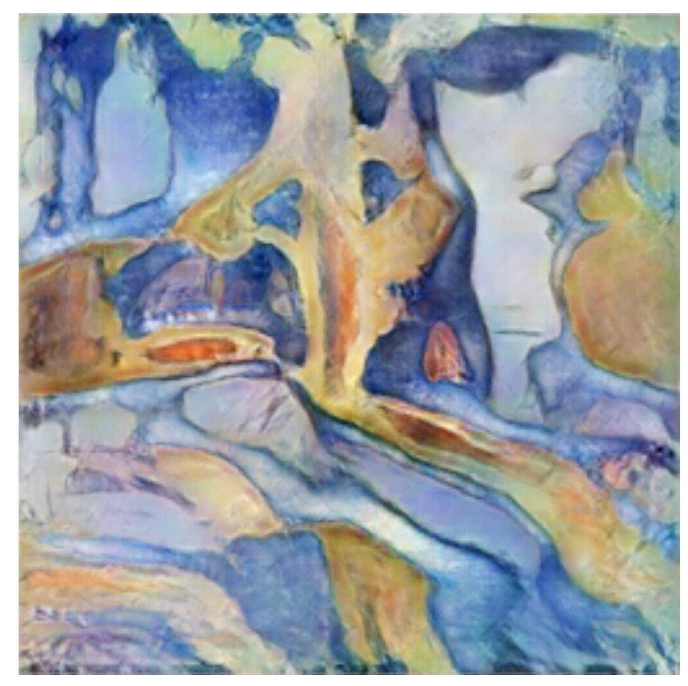
lia.coleman@kaust.edu.sa; optas@cs.stanford.edu; mohamed.elhoseiny@kaust.edu.sa

Our AI Generated Art. Exhibited in the NeurIPS 2019 Creativity Gallery.



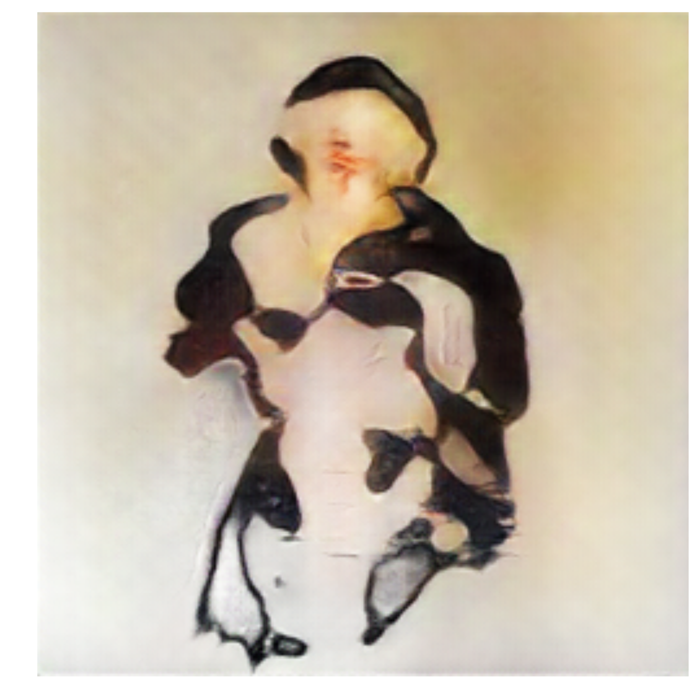
Reverie

Summer Haze



Tectonic

The Angel



His Last Visit

MOTIVATION

Do artists like AI generated "paintings" and can they tell them apart from human-made ones?

How are their responses to the above questions different from those given by non-experts?

METHODS

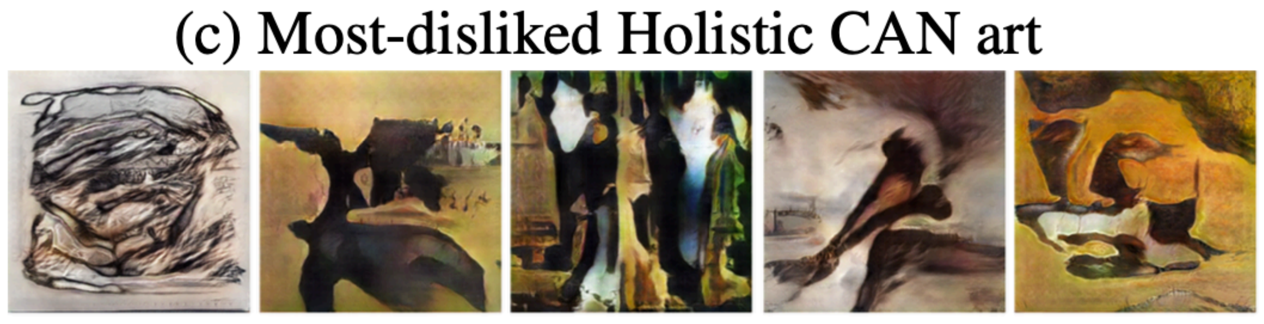
- We trained a CAN model [2,8] on the WikiArt dataset using multiclass cross entropy for our loss as in [8].
- Human evaluation study on 120 images: half synthetic generations; half human-art.
- Ask a group of 13 professional artists:
  Rate on a scale of 1-5 how much you like this work.
  Do you think this artwork was created by a human or generated by an AI?
- 5 responses per image from Amazon Mechanical Turk raters.
- In evaluation, we simulate real-world evaluation conditions: Split the data into two sets: the 'seen' set, and the 'unseen' set. Our evaluation procedure is:

1. From the 'seen' set, calculate the Cohen's Kappa (a measure of similarity) between each turker's likeability responses and the artist majority vote on likeability.
2. Choose a Cohen's Kappa threshold to exclude turkers from the 'unseen' set who deviate from artists the most.
3. (proof-of-viability) Verify that this Cohen's Kappa , which is derived from the 'seen' set, performs well on the 'unseen' set.

(a) Top-liked Holistic CAN art

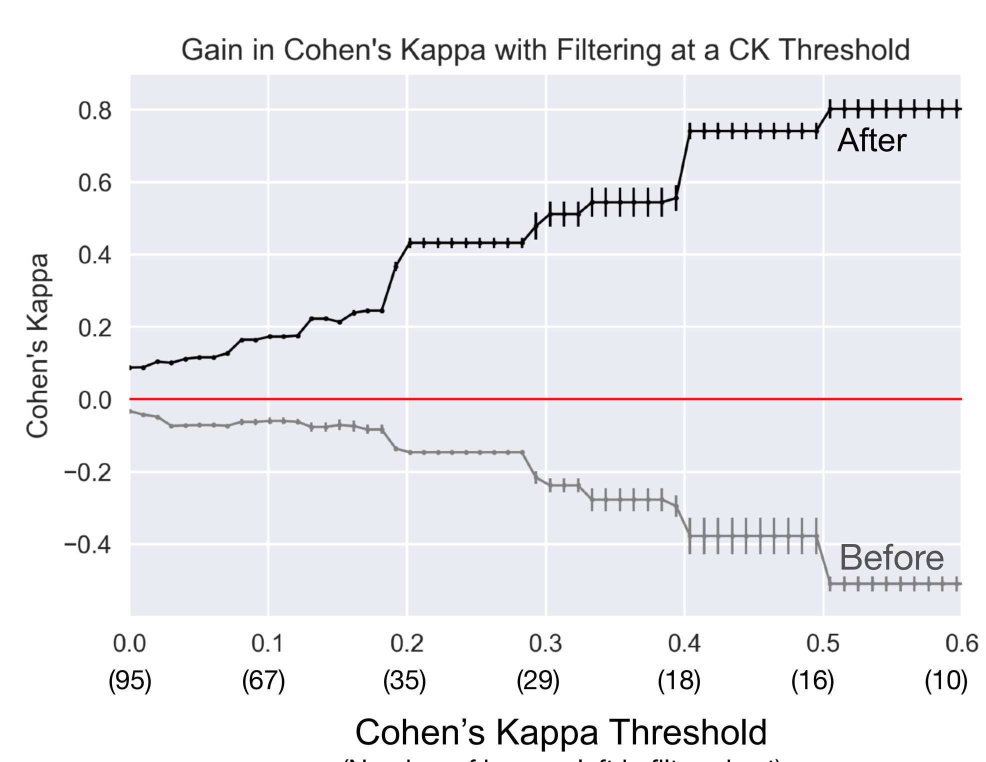
(b) Top-liked human art

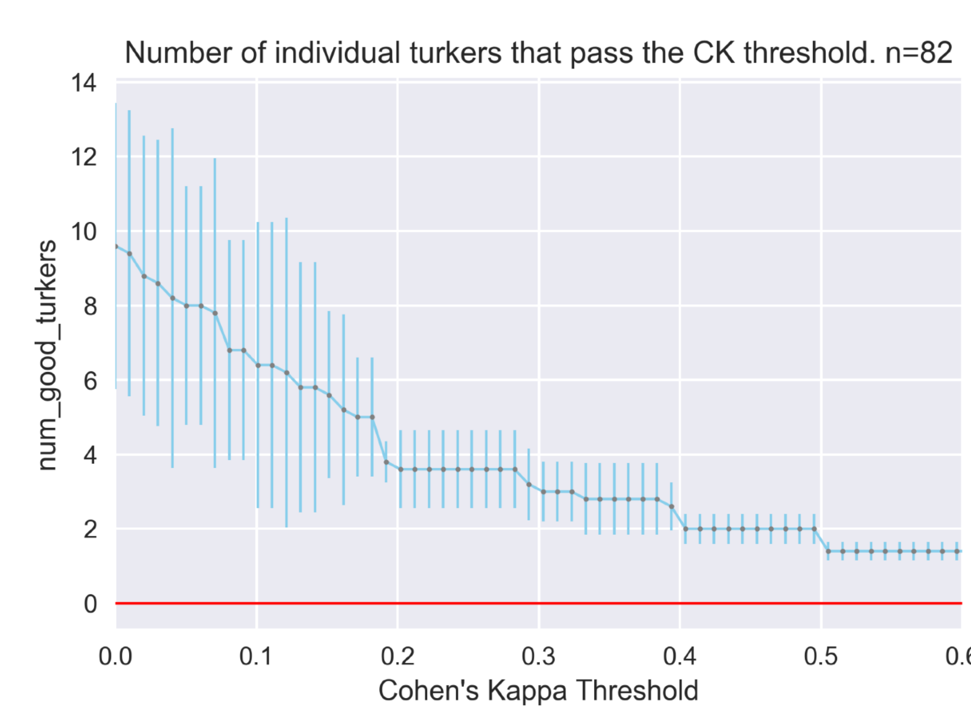
(c) Most-disliked Holistic CAN art



RESULTS

- Our best machine-generated image is on par with the 3rd best human-created artwork, both attaining 75% of artists' votes.
- We can use abundant Amazon MTurker labels to supplement the scarce labels of artists. Using Cohen's Kappa as a cutoff threshold for Turkers' labels, we get a large gain in Cohen's Kappa agreement with artists.
- Some aspects of likability in art are shared and learnable. A linear SVM operating on VGG16 image features achieves 78.2% ± 0.4% accuracy in likability prediction, while (educated) random guessing achieves 70.8%.





CONCLUSION

- Our analysis highlights the importance of considering artists' opinion when evaluating AI generated art.
- Using MTurk to help scale artist responses on likability for unseen images is a viable approach.
- Preliminary results that likability of artwork is learnable.

REFERENCES

1] S. DiPaola and L. Gabora. Incorporating characteristics of human creativity into an evolutionary art algorithm, 10(2):97–110, 2009.
[2] A. Elgammal, B. Liu, M. Elhoseiny, and M. Mazzone. CAN: Creative adversarial networks, generating" art" by learning about styles and deviating from style norms. ICCC, 2017.
[3] I. Goodfellow, Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In NIPS, 2014.
[4] D. Ha and D. Eck. A neural representation of sketch drawings. ICLR, 2018.
[5] P. Machado and A. Cardoso. Nevar–the assessment of an evolutionary art tool. In Proc. of the AISB00 Symposium on Creative & Cultural Aspects and Applications of AI & Cognitive Science, 2000.
[6] C. Martindale. The clockwork muse: The predictability of artistic change. Basic Books, 1990.
[7] A. Mordvintsev, C. Olah, and M. Tyka. Inceptionism: Going deeper into neural networks. Google Research Blog. 2015
[9] O. Sbai, M. Elhoseiny, A. Bordes, Y. LeCun, and C. Couprie. Design: Design inspiration from generative networks. In ECCV workshop, 2018.

Are you an ARTIST, DESIGNER or an ML RESEARCHER interested in art or design?

We value your input in measuring progress on AI generated art. If you are interested in helping us with your feedback, please fill out this form.

https://tinyurl.com/v5kapde