Figure 1: Example sketches produced collaboratively by humans and artificial neural network model. Red strokes produced by humans; yellow strokes produced by model.

`collabdraw`: An Environment for Collaborative Sketching with an Artificial Agent

Judith E. Fan
UC San Diego
9500 Gilman Drive
La Jolla, CA 92093, USA
jefan@ucsd.edu

Monica Dinculescu
Google Brain
1600 Amphitheatre Parkway
Mountain View, CA 94043, USA
noms@google.com

David Ha
Google Brain
1600 Amphitheatre Parkway
Mountain View, CA 94043, USA
hadavid@google.com

C&C '19, June 23-26, 2019, San Diego, CA, USA

ACM ISBN 978-1-4503-5917-7/19/06.
https://doi.org/10.1145/3325480.3326578

Abstract
Sketching is one of the most accessible techniques for communicating our ideas quickly and for collaborating in real time. Here we present a web-based environment for collaborative sketching of everyday visual concepts. We explore the integration of an artificial agent, instantiated as a recurrent neural network, who is both cooperative and responsive to actions performed by its human collaborator. To evaluate the quality of the sketches produced in this environment, we conducted an experimental user study and found that sketches produced collaboratively were just as recognizable as those produced by humans on their own. Further control analyses suggest that the semantic properties of these sketches were indeed the product of collaboration, rather than attributable to the contributions of the human or the artificial agent alone. Taken together, our findings attest to the potential of systems enabling real-time collaboration between humans and machines to create novel and meaningful content.

Author Keywords
drawing; cooperation; artificial intelligence

ACM Classification Keywords
H.5.m [Information interfaces and presentation]: Miscellaneous

Introduction

The ability to express our ideas in visual form provides the foundation for important creative activities, including art, design, and science. A critical advantage of having such visual formats is that they are easily shared and naturally support collaboration. Sketching is one of the most accessible techniques for communicating our ideas quickly and collaborating in real time to produce meaningful content. What are the computational ingredients that enable successful coordination between agents to construct meaningful sketches together?

To investigate this question, we developed a web-based environment for collaborative sketching of everyday visual concepts. We integrated an artificial agent into this environment, instantiated as a recurrent neural network model [5, 10], to collaborate with each person who entered the environment. We explore the hypothesis that endowing this agent with: *first,* a well-defined goal it shares with its human collaborator, and *second*, the ability to adapt to this person's contributions, would support the collaborative production of recognizable sketches. To test this hypothesis, we conducted an experimental user study and found that the resulting collaborative sketches contained similar semantic information to those produced by humans on their own. Further control analyses suggest that the semantic information in these sketches were indeed the product of collaboration, rather than attributable to the contributions of the human or the artificial agent alone.

By providing formal quantitative evaluation of the semantic content in collaboratively produced sketches, our work complements recent investigations of human-computer collaborative sketching that have focused on user experience [9, 2, 1]. Taken together, these advances attest to the potential of systems enabling real-time collaboration between humans and machines to create novel and meaningful content. Moreover, they may inform the design of drawing interfaces for younger children [8], for whom the additional scaffolding provided by an artificial agent may enhance their ability to express their intentions more fully when drawing.

collabdraw: a collaborative sketching environment

Building this web-based collaborative sketching environment posed two key technical challenges: *first*, development of a cooperative artificial agent capable of sketching familiar visual concepts, and *second*, real-time deployment in a web browser. To solve the first challenge, we leverage a recently developed recurrent neural network model (`sketch-rnn`, [5]), that was previously trained on a large number of human-generated sketches of various visual concepts [6], and shown to have learned a latent distribution that can be sampled to generate novel sketches of those concepts. Importantly for our study, `sketch-rnn` can propose coherent continuations of partial sketches of a particular target concept, providing a natural interface for collaboration with a human who shares the goal of sketching that concept and has generated a partial sketch of it. To solve the second challenge, we used an implementation of `sketch-rnn` from Magenta.js [10], a high-level JavaScript API for doing inference with machine learning models in client-side applications, built with Tensorflow.js [13].

Experimental user study

We conducted an experimental user study in order to evaluate how well our collaborative sketching environment supported the production of recognizable sketches. 90 participants recruited via Amazon Mechanical Turk

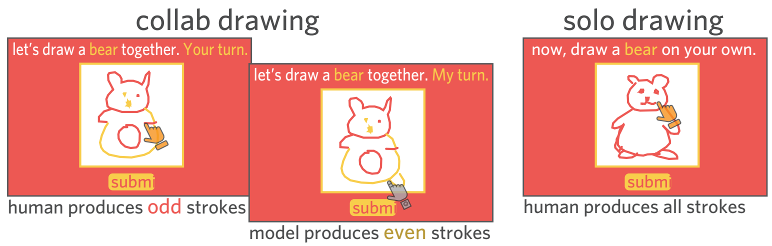

Figure 2: All participants drew each concept under both `collab` and `solo` conditions. On `collab` trials, the human and model alternated producing a single stroke until the human was satisfied with the drawing. On `solo` trials, the human drew the cued concept on their own.

(AMT) completed the study. In each session, participants drew eight different visual concepts (i.e., bear, cat, duck, lion, pig, rabbit, sheep, and swan) four times each, in a randomly interleaved sequence. In order to measure the impact of collaboration, controlling for variation between individuals in sketching behavior, participants sketched each object twice under collaborative (`collab`) and twice under non-collaborative (`solo`) settings.

On each `collab` sketching trial, the human participant and artificial agent (i.e., 'model') took turns producing one stroke each until the human participant was satisfied with the sketch (Fig. 2). The human participant always produced the first stroke, and thus all odd-numbered strokes in each collaborative sketch, whereas the model produced all of the even-numbered strokes. On each `solo` sketching trial, the human participant drew the object entirely on their own, providing a measure of the baseline recognizability of human sketches of each object in our dataset.

Analysis of sketch behavior

The primary goal of the user study was to test two critical aspects of the `collabdraw` environment: *first*, that the environment supports real-time collaboration between a human and artificial agent to produce meaningful sketches, and *second*, that the semantic information conveyed in these collaborative sketches was indeed the product of collaboration, rather than attributable to the contributions of the human or the artificial agent alone.

Extracting semantic information from sketches

Measuring the semantic properties of a sketch that determines its recognizability requires a principled approach for encoding their high-level visual properties. Here we capitalize on recent work validating the use of deep convolutional neural network (CNN) models to encode such properties in sketches [4, 14, 11]. Specifically, we used VGG-19 [12] trained to categorize objects in photos from the Imagenet database [3] to extract high-level feature-vector representations of each sketch. Prior to feature extraction, we rendered all sketches as 224px x 224px RGB images containing only black strokes of fixed width (5px), and that were cropped to include only the sketch plus 10px padding. Each 4096-dimensional feature vector reflected VGG-19 activations to sketches in the second fully-connected layer of the network (i.e., `fc6`). To extract the degree to which each sketch expressed the target concept, we applied an 8-way logistic classifier with L2 regularization to predict the identity of the sketched concept. The classifier was trained on a category-balanced set of 40K sketches of the same 8 concepts in the Quickdraw dataset.

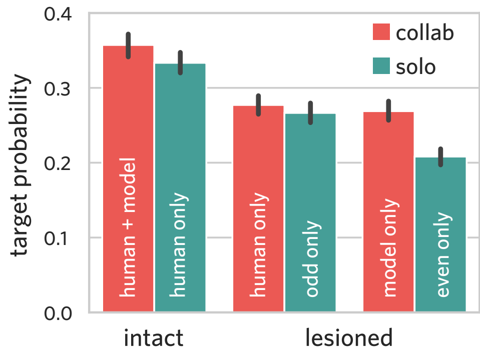

Figure 3: Average probability assigned to target concept by classifier for intact `collab` and `solo` sketches (left). Comparison with lesioned `collab` sketches in which only the human-generated (middle) or model-generated (right) strokes remain, as well as lesioned `solo` sketches in which only odd-numbered (middle) or even-numbered (right) strokes remain. Error bars reflect 95% confidence intervals.

Does collaboration yield recognizable sketches?
To evaluate the degree to which our system supported successful collaboration, for each `collab` sketch we extracted the pattern of probabilities assigned to each of the 8 concepts by the classifier, and compared this pattern to that assigned to sketches in the baseline `solo` condition. We found that the highest probability value was consistently assigned to the target concept in both conditions (`collab`: 0.357, 95% CI: [0.340, 0.374]; `solo`: 0.333, 95% CI: [0.316, 0.350]), validating our general approach to measuring semantic information in these sketches. If anything, there was modest evidence that `collab` sketches were slightly more recognizable than `solo` sketches, as revealed by a generalized linear mixed-effects model predicting classifier accuracy from condition, that included random intercepts for participant and concept ($b = 0.142, z = 1.78, p = 0.075$).

Beyond classifier accuracy, we further found a high degree of correspondence in the *pattern* of probabilities assigned to each concept between the `collab` and `solo` conditions. We quantified this correspondence using Jensen-Shannon divergence (JSD), which captures how different two probability distributions are, and lies in the range $[0, 1]$. We compared the observed mean divergence across concepts (JSD$=\ 0.0024$) to a null distribution where the assignment of probabilities to concept was scrambled independently for each condition. We found that the observed JSD fell far below the null distribution for every concept, providing strong evidence of non-trivial and similar structure in each pair of distributions.

Although the CNN backbone was trained on photos rather than sketches, the final classifier layer was trained using a subset of the Quickdraw sketch dataset, the same source of training data used for the artificial agent (`sketch-rnn`). This raises the possibility that the CNN-based classifier was not measuring semantic information in these sketches, or at least not the same information humans would use to recognize a sketch, but other properties that are idiosyncratic to the Quickdraw dataset. To test this possibility, we recruited a new group of naive human participants (N=30) to provide labels for each sketch. Consistent with the results so far, we found that the pattern of correct identifications and confusion errors made by these participants was highly similar between the `collab` and `solo` conditions (Spearman rank correlation = 0.842), validating the use of the CNN-based classifier to quantify semantic information in our sketch dataset.

Moreover, we found that similar amounts of effort were invested in producing sketches in both conditions: neither the number of strokes (`collab`=12.2; `solo`=11.7; 95% CI for difference=[-5.63, 5.06]), nor the amount of time taken to finish each sketch reliably differed between conditions (`collab`=39.0s; `solo`=36.8s; 95% CI for difference=[-41.6s, 38.9s]). Taken together, these results suggest that human participants and the artificial agent suceeded in producing sketches that carried as much relevant semantic information about each concept as those sketches produced by the human participant alone.

Did both parties contribute to the result?
An alternative explanation for the findings so far is that the human participant was primarily responsible for contributing strokes that conveyed relevant semantic information about the target concept. On this account, if we were to consider only the human-generated strokes in `collab` sketches, then the probability assigned to the target concept should remain unaffected. Conversely, if the artificial agent were primarily responsible for contributing semantically relevant information in each sketch, then considering only the model-generated strokes should be provide as much information about the target as intact `collab` sketches.

To test both of these alternative explanations, we conducted the same classifier-based analysis as above on 'lesioned' versions of each `collab` sketch in which either all of the human (odd-numbered) strokes or all of the model (even-numbered) strokes were removed. We found that the probability assigned to the target decreased substantially in both types of lesioned sketches, and to a comparable degree (human strokes only: 0.277, 95% CI: [0.264, 0.291]; model strokes only: 0.269, 95% CI: [0.257, 0.281]), suggesting that both parties contributed to the overall meaning of each sketch (Fig. 3). To provide a baseline for comparison, we additionally repeated the analysis on the `solo` sketches, lesioning either all of the odd-numbered or even-numbered strokes. We found that these lesions produced a comparable decrement in the amount of semantic information in each sketch (odd strokes only: 0.267, 95% CI: [0.254, 0.280]; even strokes only: 0.208, 95% CI: [0.199, 0.218]); if anything, the sketches containing only the human-generated even strokes conveyed less information about the target than those containing only the model-generated even strokes. In sum, these results suggest that the semantic properties of the `collab` sketches were indeed the product of collaboration, rather than solely attributable to the contributions of the human or the artificial agent.

Future Work

The current study affirms the potential of systems enabling real-time collaboration between humans and machines to create novel and meaningful content. Next, we plan to expand our dataset of collaborative sketches to include a larger number of sketches of a wider variety of concepts. More broadly, we aim to further develop our collaborative sketching environment for a range of other applications, including ones that promote creative improvisation during sketch production [7], and others that provide scaffolding for individuals who may experience difficulty sketching on their own, such as young children [8] and neurological patients [15]). In the long run, such technologies may even help users to discover new modes of creative visual expression.

Acknowledgements

Thanks to Bria Long and Mike Frank for helpful discussion.

REFERENCES

1. Nicholas Davis, Chih-PIn Hsiao, Kunwar Yashraj Singh, Lisa Li, and Brian Magerko. 2016b. Empirically studying participatory sense-making in abstract drawing with a co-creative cognitive agent. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*. ACM, 196–207.

2. Nicholas Mark Davis, Chih-Pin Hsiao, Kunwar Yashraj Singh, and Brian Magerko. 2016a. Co-creative drawing agent with object recognition. In *Twelfth artificial intelligence and interactive digital entertainment conference*.

3. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. (2009).

4. Judith E. Fan, Daniel L. K. Yamins, and Nicholas B. Turk-Browne. 2018. Common Object Representations for Visual Production and Recognition. *Cognitive Science* (2018).

5. David Ha and Douglas Eck. 2017. A neural representation of sketch drawings. *arXiv preprint arXiv:1704.03477* (2017).

6. Jonas Jongejan, Henry Rowley, Takashi Kawashima, Jongmin Kim, and Nick Fox-Gieg. 2017. Google QuickDraw. (2017). `https://quickdraw.withgoogle.com/`

7. Pegah Karimi, Nicholas Davis, Kazjon Grace, and Mary Lou Maher. 2018. Deep Learning for Identifying Potential Conceptual Shifts for Co-creative Drawing. *CoRR* abs/1801.00723 (2018). `http://arxiv.org/abs/1801.00723`

8. Bria Long, Judith E Fan, and Michael C Frank. 2018. Drawings as a window into developmental changes in object representations. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society*.

9. Changhoon Oh, Jungwoo Song, Jinhan Choi, Seonghyeon Kim, Sungwoo Lee, and Bongwon Suh. 2018. I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 649.

10. Adam Roberts, Curtis Hawthorne, and Ian Simon. 2018. Magenta. js: A javascript api for augmenting creativity with deep learning. (2018).

11. Patsorn Sangkloy, Nathan Burnell, Cusuh Ham, and James Hays. 2016. The sketchy database: learning to retrieve badly drawn bunnies. *ACM Transactions on Graphics (TOG)* 35, 4 (2016), 119.

12. Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

13. Daniel Smilkov, Nikhil Thorat, Yannick Assogba, Ann Yuan, Nick Kreeger, Ping Yu, Kangyi Zhang, Shanqing Cai, Eric Nielsen, David Soergel, and others. 2019. TensorFlow. js: Machine Learning for the Web and Beyond. *arXiv preprint arXiv:1901.05350* (2019).

14. Qian Yu, Yongxin Yang, Feng Liu, Yi-Zhe Song, Tao Xiang, and Timothy M Hospedales. 2017. Sketch-a-net: A deep neural network that beats humans. *International journal of computer vision* 122, 3 (2017), 411–425.

15. Daniel Zalewski. 2015. Life Lines. *The New Yorker* (2015).