DoodleFormer: Creative Sketch Drawing with Transformers
Supplementary Material

Ankan Kumar Bhunia[1], Salman Khan[1,2], Hisham Cholakkal[1], Rao Muhammad Anwer[1,3], Fahad Shahbaz Khan[1,4], Jorma Laaksonen[3], Michael Felsberg[4]

[1] Mohamed bin Zayed University of AI, UAE [2] Australian National University, Australia
[3] Aalto University, Finland [4] Linköping University, Sweden
ankan.bhunia@mbzuai.ac.ae

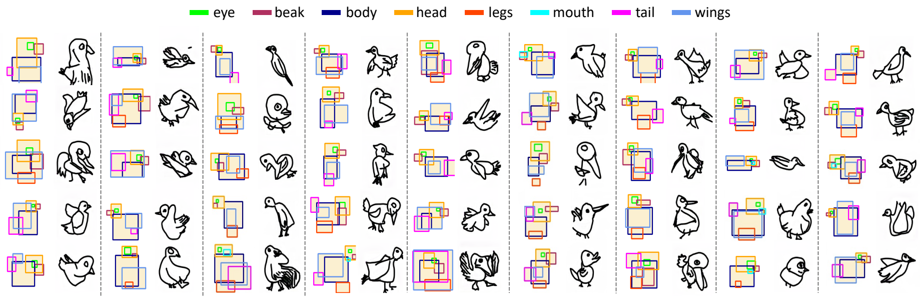


Fig. 1: Visualization of the *coarse-to-fine* creative sketch drawing process. The model first draws the holistic coarse structure of the sketch and then fills the fine-details to generate the final sketch. Both the coarse structure and the final sketch output are shown side-by-side in the figure. By first drawing the holistic coarse structure of the sketch aids to appropriately decide the location and the size of each sketch body part to be drawn.

In this supplementary material, we present additional qualitative results and additional user study details. In Sec. 1, we present the visualizations depicting the coarse-to-fine sketch generation process . Sec. 2 presents the additional details for the calculation of the evaluation metrics. Sec. 3 shows additional quantitative results. Sec. 4 provides additional details of user study experiments.

1 Coarse-to-fine Sketch Generation

Fig. 1 presents example visualizations depicting the intermediate results of our *coarse-to-fine* creative sketch generation. The proposed two-stage DoodleFormer framework decomposes the creative sketch generation problem to first capture the holistic coarse structure of the sketch and then injecting fine-details to generate the final sketch. Both the coarse structure from the first-stage PL-Net, and the final sketch output from the second-stage PS-Net, are shown side-by-side in Fig. 1. By first drawing the holistic coarse structure of the sketch aids to appropriately determine the location and the size of each sketch body part to be drawn.

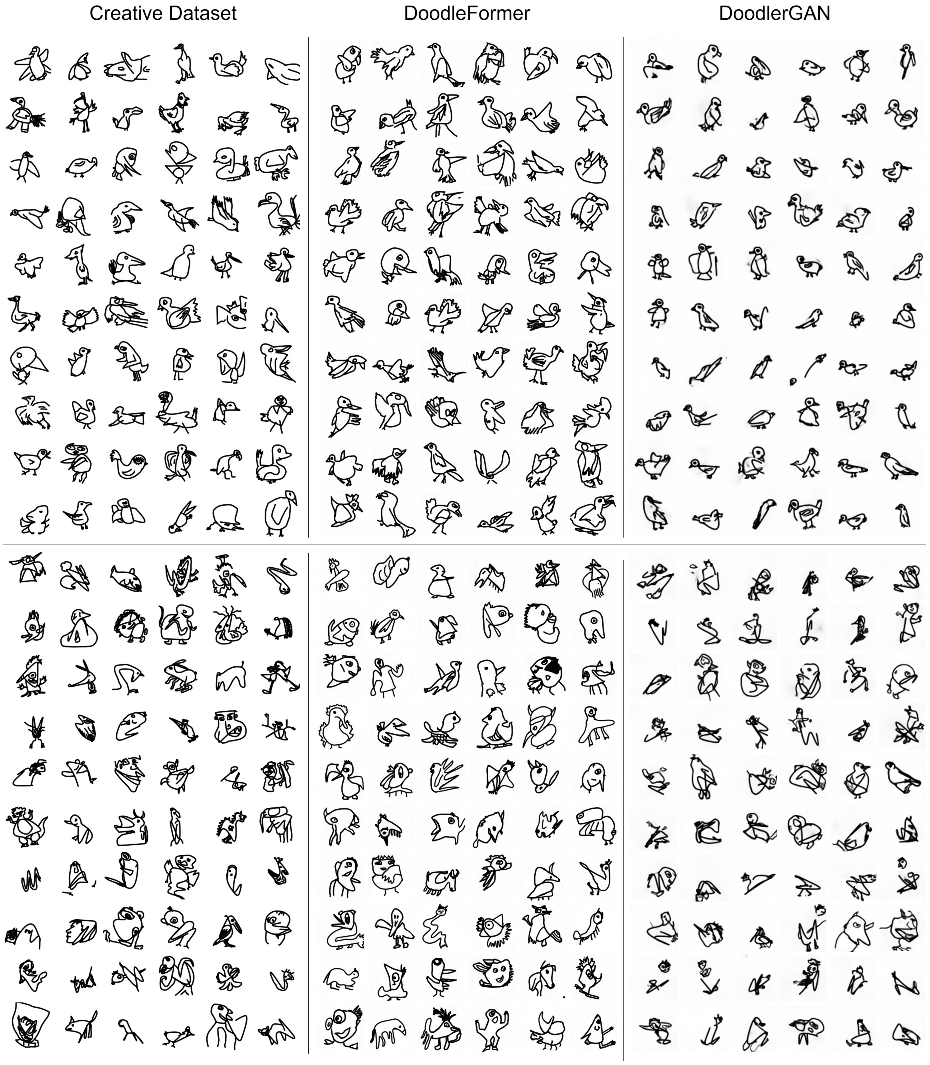
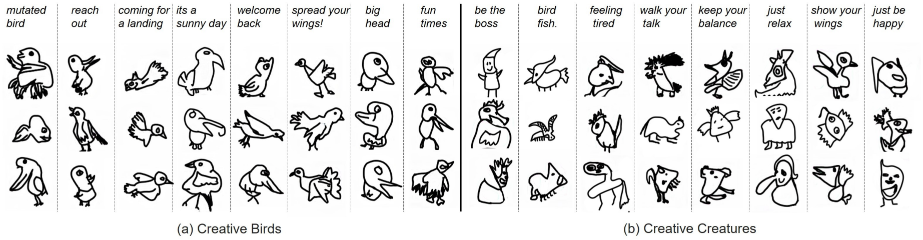


Fig. 2: Additional qualitative comparisons on Creative Birds and Creative Creatures datasets. In this figure, we compare the generated sketches using the proposed DoodleFormer (in the middle column) with DoodlerGAN [1] (in the right column). The human drawn creative sketch images from the datasets are shown in the left column. DoodlerGAN suffers from topological artifacts. Also, DoodlerGAN generated sketches have lesser diversity in terms of size, appearance and posture. The proposed DoodleFormer alleviates the issues of topological artifacts and disconnected body parts, generating creative sketches that are more realistic and diverse.

2 Additional Details of Evaluation Metrics

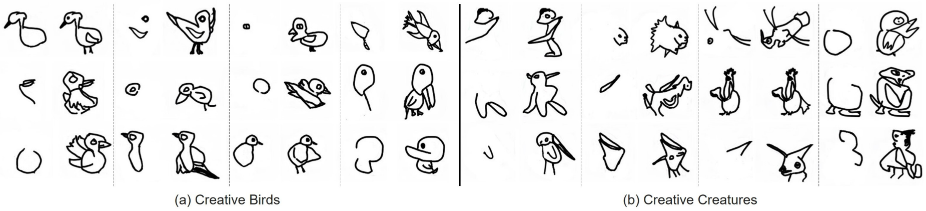
We quantitatively evaluate our proposed approach based on four metrics: Frèchet inception distance (FID), generation diversity (GD), characteristic score (CS), and semantic



(a) Creative Birds

(b) Creative Creatures

Fig. 3: Additional qualitative results for text to creative sketch generation of DoodleFormer on Creative Birds and Creative Creatures dataset. For each input text, we show three different samples generated using DoodleFormer. The results demonstrate that our approach is capable of generating diverse sketch images while ensuring that the generated sketches are well matched with the user provided text inputs.



(a) Creative Birds

(b) Creative Creatures

Fig. 4: Additional qualitative results for text to creative sketch completion of DoodleFormer on Creative Birds and Creative Creatures dataset. The results depict that DoodleFormer accurately completes the missing regions of the given partial input sketch.

diversity score (SDS). Following [1], we use an inception model trained on the QuickDraw3.8M dataset [3] to calculate these metrics. We generate $10,000$ sketches based on a randomly sampled set of previously unseen initial strokes. Then, the generated sketches are resized into 64×64 images and passed through the trained inception model to calculate the above mentioned metrics.

3 Additional Qualitative Results

Fig. 2 shows additional qualitative results of our proposed DoodleFormer for creative sketch generation on both datasets (Creative Birds and Creative Creatures). For better visual comparison, we also show the creative sketch images from the datasets and the DoodlerGAN [1] generated sketch images in the same figure.

In Fig. 3, we present additional qualitative results for text to creative sketch generation on both datasets (Creative Birds and Creative Creatures). For each user provided input text, we show three samples generated using our proposed DoodleFormer. The results demonstrate the effectiveness of our approach towards generating diverse sketch images while ensuring that the generated sketches are well matched with the user provided text inputs. Fig. 4 shows qualitative results for creative sketch completion. The results show that DoodleFormer accurately completes the missing regions of these challenging incomplete sketches. Fig. 5 shows that GMM-based modelling can generate multiple bound- ing box layouts with more diverse sketch results (w.r.t appearance, size,

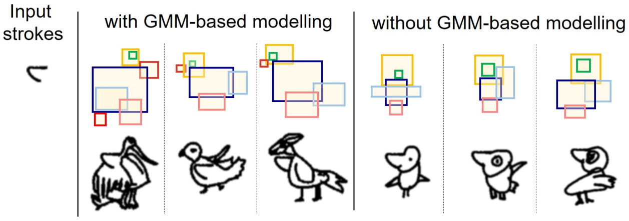


Fig. 5: Qualitative analysis of GMM-based modelling for common initial strokes. The introduction of the GMM-based modelling substantially improves the diversity of generated sketches.

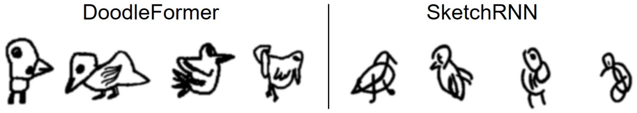


Fig. 6: A visual comparison of creative sketches generated by DoodleFormer and SketchRNN [2] on Creative Birds dataset.

orientation and posture). Fig. 6 presents a qualitative comparison of DoodleFormer with SketchRNN [2].

4 User Study Additional Details

Here, we present additional details of our user studies on 100 human participants to evaluate the creative abilities of our proposed DoodleFormer. We compare the DoodleFormer generated sketches with DoodlerGAN and human-drawn Creative dataset sketches. Specifically, we show participants pairs of sketches – one generated by DoodleFormer and the other by a competing approach. Each participant has to answer 5 questions for each of these pairs. The questions are: which one (a) is more creative? (b) looks more like a bird/creature? (c) is more likely to be drawn by a human? (d) in which case, the initial strokes are well integrated? (e) like the most overall. Our proposed DoodleFormer performs favorably against DoodlerGAN [1] for all five questions on both datasets. Further, the DoodleFormer generated sketch images were observed to be comparable with the sketches drawn by human in Creative Datasets for all the five questions.

Acknowledgements. This work has been funded by the Academy of Finland in project USSEE (345791) and supported by the Aalto Science-IT project.

References

1. Ge, S., Goswami, V., Zitnick, C.L., Parikh, D.: Creative sketch generation. In: ICLR (2021) 2, 3, 4
2. Ha, D., Eck, D.: A neural representation of sketch drawings. In: ICLR (2018) 4
3. Xu, P., Hospedales, T.M., Yin, Q., Song, Y.Z., Xiang, T., Wang, L.: Deep learning for freehand sketch: A survey and a toolbox. arXiv preprint arXiv:2001.02600 (2020) 3