DoodleFormer: Creative Sketch Drawing with Transformers

Ankan Kumar Bhunia[1] Salman Khan[1,2] Hisham Cholakkal[1] Rao Muhammad Anwer[1,4]
Fahad Shahbaz Khan[1,3] Jorma Laaksonen[4] Michael Felsberg[3]
[1]MBZUAI, UAE [2]Australian National University, Australia [3]Linköping University, Sweden [4]Aalto University, Finland

ECCV TEL AVIV 2022

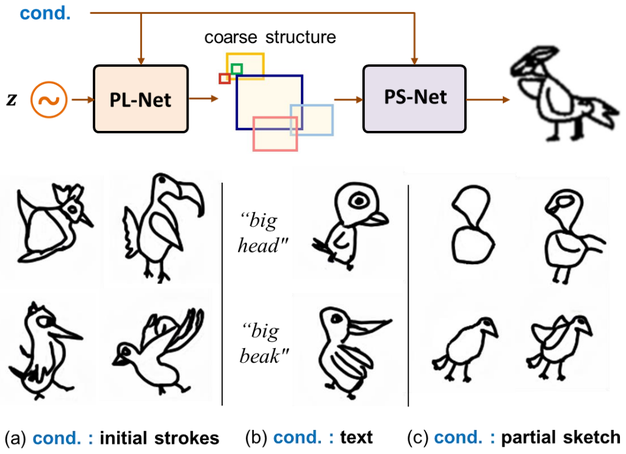
Motivation

Limitation of existing frameworks

- Poor Quality: The recent part-based method DoodlerGAN [1] doesn't employ an explicit mechanism to ensure that each body part is placed appropriately with respect to the rests. This leads to topological artifacts and connectivity issues.
- Lack of diversity: DoodlerGAN struggles to generate diverse sketch images, which is an especially desired property in creative sketch generation.

Why *Coarse-to-fine* Design?

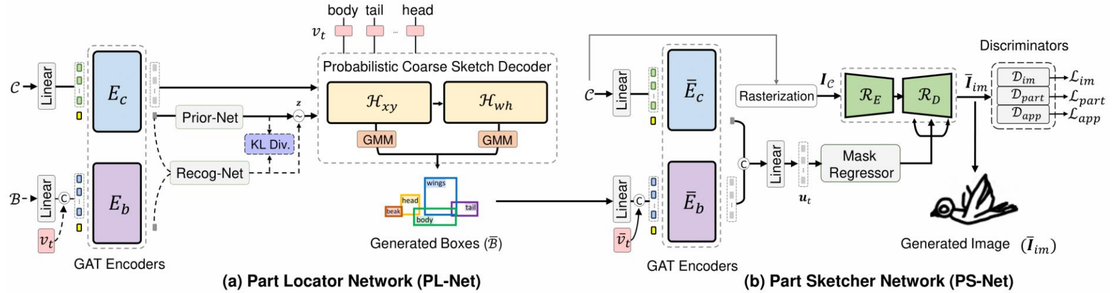
- Generally, a human artist *(i) first draws the holistic coarse structure* of the sketch and then *(ii) fills the fine-details* to generate the final sketch. By first drawing the holistic coarse structure of the sketch aids to appropriately decide the location and the size of each sketch body part to be drawn.



(a) **cond. : initial strokes** (b) **cond. : text** (c) **cond. : partial sketch**

- **1st stage**: PL-Net, takes the initial stroke points as the conditional input and learns to return the bounding boxes corresponding to each body part to be drawn;
- **2nd stage**: PS-Net, takes the predicted box locations as inputs and generates the final sketch image;

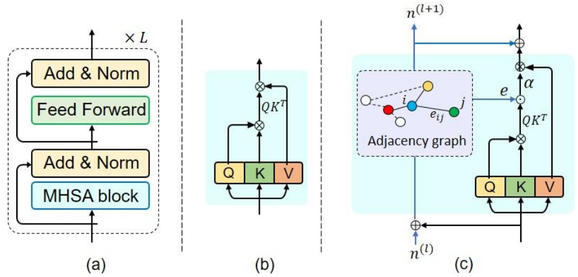
Methodology



Legend —→: train & test - - →: only train $\mathcal{C}$: initial stroke points $\mathcal{B}$: target boxes v_t: part embedding ⓒ: concatenation ⓢ: sampling ▯: *cls* token

We propose a novel two-stage *transformer-based encoder-decoder* framework, DoodleFormer, for creative sketch generation. DoodleFormer decomposes the creative sketch generation problem into the construction of holistic coarse sketch composition followed by injecting fine details to generate final sketch image.

- GAT Encoder blocks: Our framework comprises of *graph-aware transformer* (GAT) block-based encoders to capture structural relationship between different regions within a sketch.



- While the standard self-attention module is effective towards learning highly contextualized feature representation, it does not explicitly emphasize on the *local structural relation*. However, creative sketches are structured inputs with definite connectivity patterns between sketch parts. To model this structure, we propose to encode an adjacency based graph implemented with spectral graph convolution.
- GMM-based probabilistic coarse sketch Decoder: we further introduce probabilistic coarse sketch decoders that utilize *GMM modelling for box prediction*. This enables our DoodleFormer to achieve diverse, yet plausible coarse structure for sketch generation.
- Different from the conventional box prediction that directly maps the decoder output features as deterministic box parameters, our GMM-based box prediction is modeled with M normal distributions

Loss Objectives: The PL-Net loss is the weighted sum of the *reconstruction loss*, and the *KL divergence* loss.

The training of PS-Net follows the standard GAN formulation where the PS-Net generator is followed by additional discriminator networks to obtain *image-level, part-level, and appearance adversarial losses*.

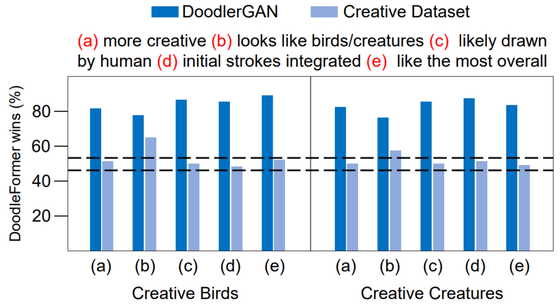
Experiments

Quantitative analysis of DoodleFormer

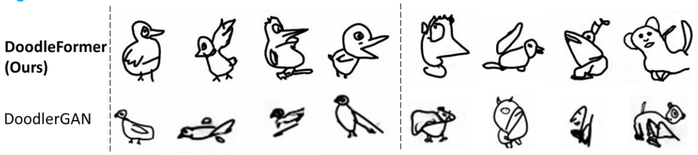
Methods	Creative Birds			Creative Creatures			
	FID(↓)	GD(↑)	CS(↑)	FID(↓)	GD(↑)	CS(↑)	SDS(↑)
Training Data	-	19.40	0.45	-	18.06	0.60	1.91
SketchRNN [12]	82.17	17.29	0.18	54.12	16.11	0.48	1.34
StyleGAN2 [17]	130.93	14.45	0.12	56.81	13.96	0.37	1.17
DoodlerGAN [10]	39.95	16.33	**0.69**	43.94	14.57	0.55	1.45
DoodleFormer (Ours)	**16.45**	**18.33**	0.55	**18.71**	**16.89**	**0.56**	**1.78**

User study analysis

Higher values indicate DoodleFormer is preferred more often over the compared approaches (DoodlerGAN [1] and human drawn datasets).



Qualitative results



For additional results for different applications (Text-to-sketch, Sketch completion an House plan generation) of Doodlerformer see our paper.

Conclusions

Qualitative, quantitative and human-based evaluations show that our DoodlerFormer produces diverse, yet realistic creative sketches.

Scan code for project page

1 Ge, Songwei, et al. "Creative sketch generation." arXiv preprint arXiv:2011.10039 (2020).