Tell, Draw, and Repeat: Generating and Modifying Images Based on Continual Linguistic Instruction

Appendix

A. Object detector and localizer network

All of the evaluation metrics for the Generative Neural Visual Artist (GeNeVA) task rely on the object detector and localizer network and hence, it needs to have high detection and localization performance. We report the performance of the trained object detector and localizer network on the test set images of both Collaborative Drawing (CoDraw) and Iterative CLEVR (i-CLEVR) datasets in Table 1.

Dataset	Precision↑	Recall↑	F1-Score↑	NRMSE↓
CoDraw	0.962	0.972	0.964	0.121
i-CLEVR	1.000	1.000	1.000	0.060

Table 1. Mean test set Precision, Recall, and F1-Score for the object detector and localizer network. Normalized Root Mean Squared Error (NRMSE) is the root mean square distance between the localizer's predicted and ground truth object centroids normalized by the image dimensions. ↑: higher is better, ↓: lower is better.

B. Relational Similarity metric: rsim

B.1. Additional details

For both CoDraw and i-CLEVR datasets, we determine front-behind and left-right relationships by comparing the coordinates of their centre predicted by the object detector and localizer network. We run the network on both ground truth and generated images to predict the centre coordinates (rather than using perspective coordinates provided by the renderer as these are only available for the ground truth for i-CLEVR).

B.2. Appropriateness for evaluation

The CoDraw and i-CLEVR datasets are constructed such that there is at most one object of each object class per image. Hence, we train the object detector to predict only binary presence of each object class and the localizer regresses only one set of centroid coordinates per class. This design breaks if multiple instances of an object class are generated or if the object detector frequently misclassifies objects. However, qualitatively assessing the generated images, over-generation is rare and the object detector accuracy is very high (cf. Table 1).

Since all objects in ground truth scenes occur at most once, generations with multiple instances per class are out-of-distribution. The model cannot learn to exploit this flaw, since rsim is not optimized during training. Thus, over-generation is not a failure mode we have observed. Additionally, rsim is position-sensitive: over-generation would not necessarily produce the correct relative positions of objects since the object localizer only localizes one instance per class. For datasets with multiple instances per class, the rsim metric should be modified such that the denominator is the union of ground-truth and predicted detections, which will penalize over-generation.

B.3. Shortcomings

Quantitative measures for attributes like "boy kicking" are currently a missing piece. We share this shortcoming with all text-to-image Generative Adversarial Network (GAN)-based methods and most of the conditional GAN literature. At the moment, conditional GANs are evaluated using Inception Score (IS) and Fréchet Inception Distance (FID), both of which do not account for attributes. An evaluation metric that accounts for attributes will be a valuable contribution for future research.

B.4. Comparison with existing metrics

The Scene Similarity Metric (SSM) used by Kim et al. [1] is well-suited for the setting of predicting object location and attributes. SSM is a weighted score across recall and considers objects that face the wrong direction, incorrect expressions, poses, clip art size, distance between object positions in ground truth and predicted image, and left-right and front-behind relationships. SSM achieves the highest score for exact reconstructions. In our case, we want to not just reward reconstructions but also plausible generations where left-right, front-behind relationships are correct. Our main focus here is to generate complete images instead of predicting object location and attributes. Several attributes, such as boy or girl poses / expressions, or object directions have lower detector accuracy and consequently would reduce metric reliability (cf. Section B.2).

B.5. Qualitative evaluation

We provide generated image examples with scores spread out between the minimum value (0) and maximum value (1) on the rsim metric in Figure 1. This is to provide readers with a more intuitive understanding of how the metric captures which spatial relationships match between the ground truth and the generated image.

C. Generation Examples

We present selected examples generated using our best model (D Subtract) on two datasets. Examples generated for CoDraw are presented in Figure 2 and examples generated for i-CLEVR are presented in Figure 3. We also present random examples from all the models present in the ablation study for a qualitative comparison on the CoDraw dataset. These are shown in Figure 4 (Baseline), Figure 5 (Mismatch), Figure 6 (G prior), Figure 7 (Aux), Figure 8 (D Concat), Figure 9 (D Subtract), and Figure 10 (Non-iterative).

D. Generalization to new background images

GeNeVA-GAN was trained using the empty background image as the initial image. We ran an experiment where we used a different image (intermediate ground truth image from the test set containing objects) as the initial image. We present generated examples from this experiment in Figure 11. The model is able to place the desired object at the correct location with the correct color and shape over the provided image. This shows that the model is capable of generalizing to a background it was not trained on and it can understand the existing objects from just the initial image without any instruction history for placing them.

E. i-CLEVR Dataset Generation

To generate the image for each step in the sequence, an object with random attributes is rendered to the scene using Blender [2]. We ensure that all objects have a unique combination of attributes. Each object can have one of 3 shapes (cube, sphere, cylinder) and one of 8 colors. In contrast to CLEVR, we have a fixed material and size for objects. For the first image in the sequence, the object placement is fixed to the image center. For all the following images, the objects are placed in a random position while maintaining visibility (not completely occluded) and at a minimum distance from other objects.

To generate instructions, we use a simple text template that depends on the instruction number. For example, the second instruction in the sequence will have the following template:

"Add a [object color] [object shape] [relative position: depth] it on the [relative position: horizontal]"

From the third instruction onward, the object position is described relative to two objects. These two objects are chosen randomly from the existing objects in the scene.

F. Additional implementation details

We use 300-dimensional GloVe[1] word embeddings for representing the words in each instruction q_t. These word embeddings are encoded using a bi-directional-GRU to obtain a 1024-dimensional instruction encoding d_t. All state dimensions for the higher level GRU R are set to 1024. The output of the conditioning augmentation module is also 1024-dimensional.

The code for this project was implemented in PyTorch [3]. For the generator and discriminator optimizers, "betas" was set to $(0.0, 0.9)$ and weight decay was set to 0. The learning rates for the image encoding modules were set to 0.006. Gradient norm was clipped at 50. For each training experiment, we used a batch size of 32 over 2 NVIDIA P100 GPUs.

G. Additional language encoder experiments

We experimented with using skip-thought encoding for sentences instead of training the bi-directional-GRU encoder over GloVe embeddings. For the paper, we chose to use the latter since it performed better.

We also experimented with passing the previous image through the language encoder, but observed that it was easier for the model to generate an accurate image when the previous image features are passed to the Generator directly.

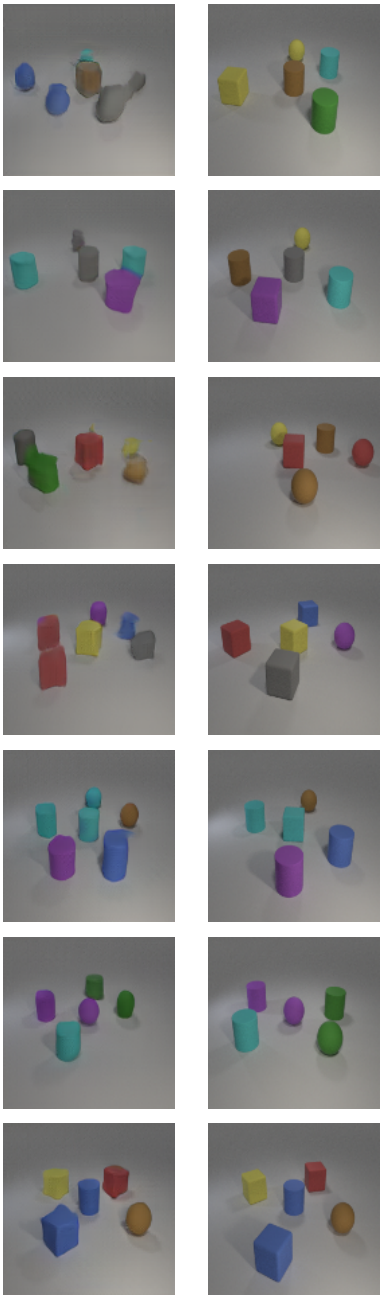
References

[1] Jin-Hwa Kim, Nikita Kitaev, Xinlei Chen, Marcus Rohrbach, Byoung-Tak Zhang, Yuandong Tian, Dhruv Batra, and Devi Parikh, "CoDraw: Collaborative drawing as a testbed for grounded goal-driven communication," in *Proceedings of the 57th Conference of the Association for Computational Linguistics*, 2019, pp. 6495–6513. [Online]. Available: https://www.aclweb.org/anthology/P19-1651

[2] Blender Online Community, "Blender - a 3D modelling and rendering package," 2016. [Online]. Available: http://www.blender.org

[3] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer, "Automatic differentiation in PyTorch," in *NIPS Autodiff Workshop*, 2017.

[1] `http://nlp.stanford.edu/data/glove.840B.300d.zip`



Objects detected in generated image: cube gray, cube brown, sphere gray, sphere blue
Objects detected in ground truth image: cube yellow, sphere yellow, cylinder green, cylinder brown, cylinder cyan
Recall: 0.00
rsim: 0.00
Explanation: None of the correct objects are drawn.

Objects detected in generated image: cube gray, cylinder gray, cylinder purple, cylinder cyan
Objects detected in ground truth image: cube purple, sphere yellow, cylinder gray, cylinder brown, cylinder cyan
Recall: 0.4
rsim: 0.25
Explanation: The cyan and gray cylinders are the only two objects detected in the generated image from the five ground truth objects detected.

Objects detected in generated image: cube red, cube green, sphere brown, cylinder gray
Objects detected in ground truth image: cube red, sphere red, sphere brown, sphere yellow, cylinder brown
Recall: 0.4
rsim: 0.35
Explanation: The red cube and brown sphere are detected common to both images. Most of the relationships of these two and the center are correct.

Objects detected in generated image: cube gray, cube red, cube yellow, sphere purple
Objects detected in ground truth image: cube gray, cube red, cube blue, cube yellow, sphere purple
Recall: 0.8
rsim: 0.45
Explanation: Only the blue cube is not detected in the generated image. Several spatial relationships of the common objects and the center are incorrect.

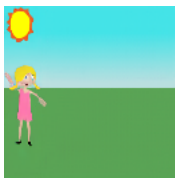
Objects detected in generated image: sphere brown, sphere cyan, cylinder blue, cylinder purple, cylinder cyan
Objects detected in ground truth image: cube cyan, sphere brown, cylinder blue, cylinder purple, cylinder cyan
Recall: 0.8
rsim: 0.675
Explanation: Cyan cube detected in ground truth image is missing from the generated image. Most spatial relationships of the common objects and center are correct.

Objects detected in generated image: sphere green, sphere purple, cylinder green, cylinder purple, cylinder cyan
Objects detected in ground truth image: sphere green, sphere purple, cylinder green, cylinder purple, cylinder cyan
Recall: 1.0
rsim: 0.76
Explanation: All the objects are detected correctly but some of the spatial relationships are incorrect.

Objects detected in generated image: cube red, cube blue, cube yellow, sphere brown, cylinder blue
Objects detected in ground truth image: cube red, cube blue, cube yellow, sphere brown, cylinder blue
Recall: 1.00
rsim: 1.00
Explanation: All the objects are detected correctly and are in the correct relative positions.

Figure 1. **Column 1:** Generated final image; **Column 2:** Ground truth final image; **Column 3:** List of objects detected in the generated and ground truth image, the recall on object detection, the value of the relational similarity (rsim) metric. The examples have been selected to qualitatively show examples with diverse score values between the minimum (0) and the maximum (1) values of the rsim metric.

Drawer: ready to draw ? **Teller:** Medium sun is on the left corner fully visible.



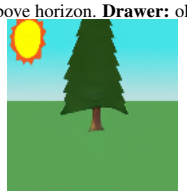
Teller: Below sun sits a mad girl with legs on front she faces right and hand touches the left border a 1 2 head is above horizon. **Drawer:** ok.

Teller: The girl is big. A fire is on front feet of girl like 1 2 “. **Drawer:** ok.

Teller: A grill is just next to fire the grill is a little lower than top flame. **Drawer:** ok.

Teller: A small pine is on right side 1 4 “ left side is cut also the tip is cut. **Drawer:** ok

Teller: In top left hand corner is medium sun cut off on top and on side. **Drawer:** I am a patient worker ready to start.



Teller: In middle of screen is a medium pine tree trunk starts dead middle of screen. **Drawer:** Got it.

Teller: A large boy is sitting cross legged almost in left corner slightly higher and to right he is facing right . **Drawer:** ok.

Teller: Drink in right hand hot dog in left, to left of hot dog is a fire. **Drawer:** ok.

Teller: On right side is small girl angry face running her foot is cut off head just touches horizon. **Drawer:** ok.

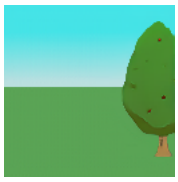
Teller: Small hot air balloon in upper left corner touching left edge very top of balloon off top edge. **Drawer:** ok.

Teller: Med tree with hole in trunk at right side about 1 3 of it below horizon and right edge of it off screen. **Drawer:** ok.

Teller: Big crying girl sitting on ground legs outstretched facing left top of her head touching bottom left corner of tree trunk. **Drawer:** ok.

Teller: Big standing boy arms in air facing right toes touching bottom edge of page slightly left of center. **Drawer:** ok.

Teller: Soccer ball on ground between boy and girl about level with boy ’s hips . **Drawer:** ok.



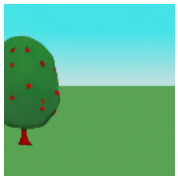
Drawer: Ready. **Teller:** Medium bushy tree right side top 2 3 above horizon right side touching right side of tree.

Teller: Left side medium slide facing right half ladder cut off at left one step is in sky. **Drawer:** Done.

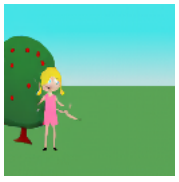
Teller: head covering bottom of slide part bum by bottom sitting legs out happy girl facing right. **Drawer:** Done.

Teller: By her back hand is a medium beach ball her hand is touching the right side of it from hand holding medium party ballo. **Drawer:** Done.

Teller: Actual balloons part is over horizon back of head touching left side of tree running happy boy facing girl wearing rainbow hat. **Drawer:** Done.



Drawer: Ready. **Teller:** On the left is a small apple tree slightly cut at the left edge. **Drawer:** Done.



Teller: In front of the tree facing right stands a happy girl waving **Drawer:** Done.

Teller: In the middle of the field standing facing right his hands to the right. **Drawer:** Who is standing in the middle and what expression?

Teller: Is a boy he is angry. **Drawer:** Done. **Teller:** On the right is a medium pine tree slightly cut at the right and at the top. **Drawer:** Done.

Teller: In front of the tree is a dog facing left. **Drawer:** Done.

Drawer: Start. **Teller:** Bushy tree in middle. **Drawer:** Ok.

Teller: Full sun right corner. **Drawer:** Yes.

Teller: Soccer ball bottom right corner. **Drawer:** Done.

Teller: Plane top left corner . **Drawer:** Ok.

Teller: Boy between sun and soccer ball. **Drawer:** Boy direction. **Teller:** Hands out facing left. **Drawer:** Ok.

Figure 2. Generation examples from our best model (D Subtract) for the CoDraw dataset.

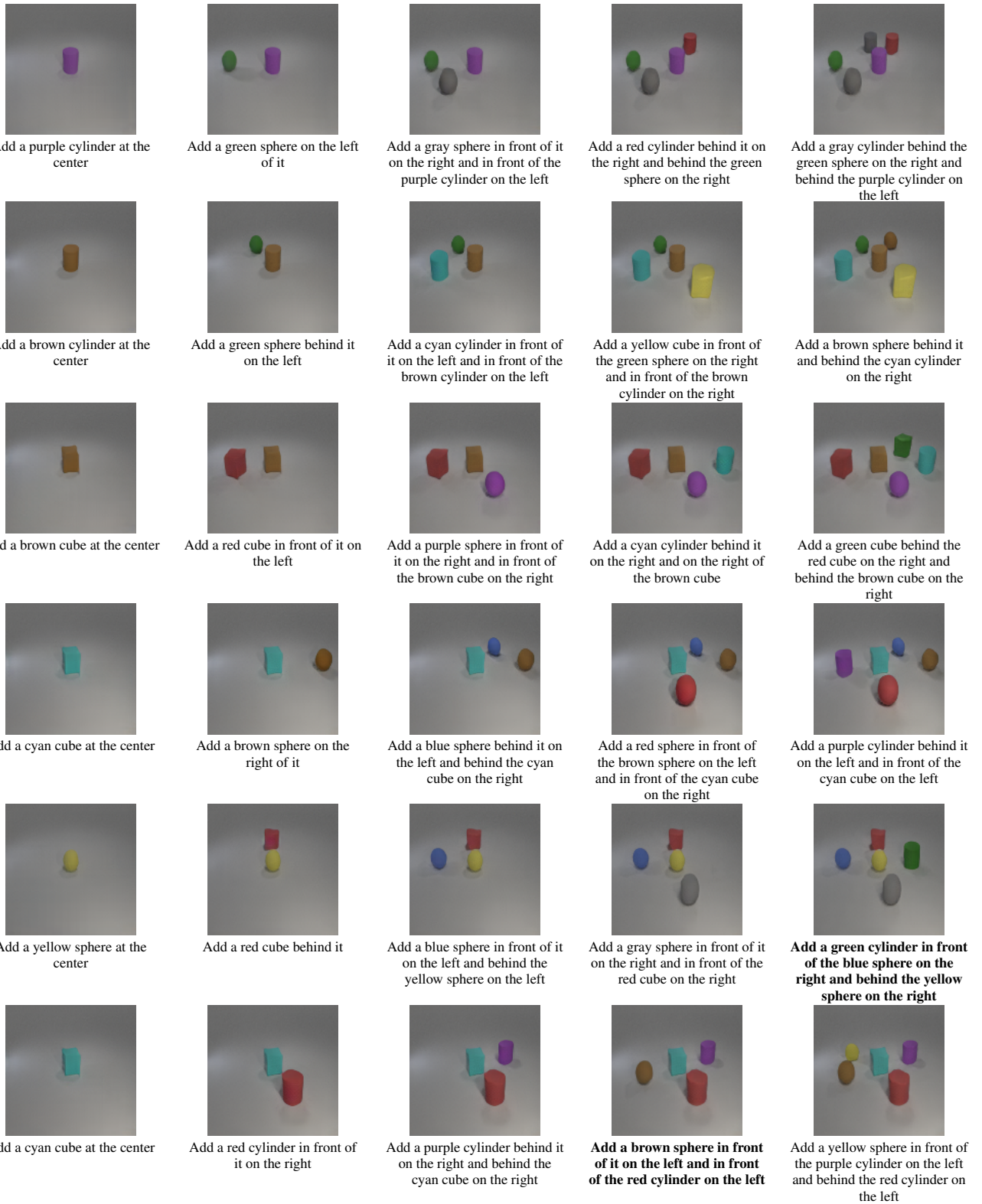
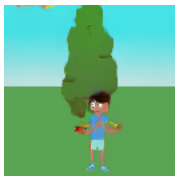


Figure 3. Generation examples from our best model (D Subtract) for the i-CLEVR dataset. Instructions where the model made a mistake are marked in bold.

Teller: small pine on right most of tree cut off on right and top big shocked running mike on right facing left his elbow on the left edge of tree

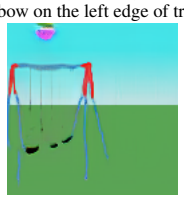


Teller: big sun in middle 1 2 cut off on top small oak tree on left hole facing right 1 4 from top

Teller: big slide on left facing right slide is in front of tree

Teller: big angry sitting with legs out jenny is on ground in front of end of slide

Teller: big cat is under slide mike 's head is over horizon jenny is 1 2 below horizon



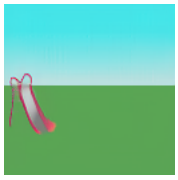
Drawer: ready when you are describe away **Teller:** large swing on the left facing right

Teller: looks like a happy girl standing on the 2nd swing facing right no teeth

Teller: sad facing right feet almost at the bottom an inch from the left **Drawer:** are his arms out or left one up

Teller: one arm in the air **Drawer:** got it **Teller:** there is a sun on the top right a little bit of it is seen it 's behind the tree

Teller: under the tree is a bee facing left **Drawer:** got it



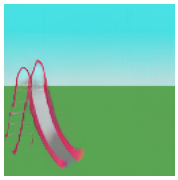
Teller: there is a small slide on the left side facing right tip of the top of the slide is above horizon

Teller: the boy with sad face is sliding down two legs kicked out there 's a baseball near the end of the slide **Drawer:** ready

Teller: the boy is probably medium size there is also a medium girl on the right edge smiling and jumping

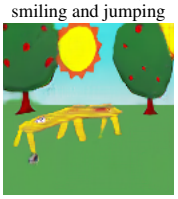
Teller: she is facing left and the top of her head is touching the horizon she has a baseball glove on her left hand

Teller: there is an airplane facing right at the top right of the picture let me know when you 're ready for me to check



Drawer: ready **Teller:** medium slide on left of screen facing right about 1 2 " from left surprised boy sitting at bottom of slide

Teller: small tree in back of slide to right a bit and then an apple tree to the right of that tree about 1 2 " apart



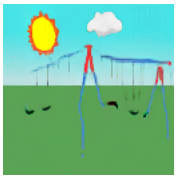
Teller: sun between the trees bottom left of sun blocked by tree on left **Drawer:** got it

Teller: surprised girl on right of screen about 1 2 inch from right border hands in the air facing left **Drawer:** got it

Teller: tennis ball beneath her left foot **Drawer:** got it

Teller: left side big swing set left behind leg cut off edge and left corner small sun covered partially by small plain cloud

Teller: right side medium sand basket dune facing left in front of dune sad cross leg sitting crying girl facing right **Drawer:** ready

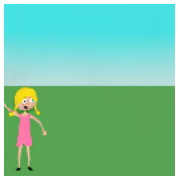


Teller: next to girl cat sitting looking at her her hand is hidden by cat 's head

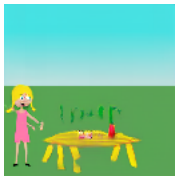
Teller: in front of the sandbox near the girl in the corner boy sad cross leg looking right

Teller: behind boy and behind girl 's hand spring bee body hidden by boy facing right and i will check

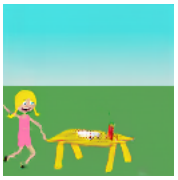
Figure 4. Random selection of examples generated by our Baseline model for the CoDraw dataset.



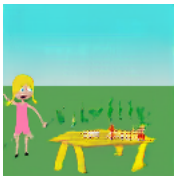
Drawer: ready **Teller:** 1 girl happy running facing right 0 2 inch from bottom to top and 0 2 inches from left to right with a chef



Teller: hat on her a table 1 and a half inches from left to right 1 2 inches from bottom to top with a pizza in the middle



Teller: and and a hot dog facing left to the right of the pizza a fire 0 1 inches from bottom to top 0 4 inches from right to left and above a



Teller: tent facing left and top of the tent is above the horizon line 0 1 inches and right is cover a little bit and above

Teller: a air balloon small like 1 2 inches from right to left and 0 2 inches from top to bottom and that 's it

Drawer: what 's in the picture **Teller:** in the top right is a sun covered partially by clouds **Drawer:** is it a large sun

Teller: in the middle left is a helicopter **Drawer:** which way is it facing

Teller: yes a large sun **Teller:** heli is facing to the right **Teller:** tail is to the left **Teller:** on bottom right is a girl in pink with left arm raised

Teller: on his mouth is an o shape **Teller:** to the right of the boy is a dog with a blue collar

Teller: above the boys left hand in the blue sky is a yellow ball

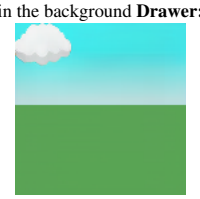
Drawer: may you please tell the first thing to draw **Teller:** there is a small tree on the right side of the scene sort of in the background **Drawer:**

Teller: there is a bear to the left but in the foreground of the tree **Drawer:** what next

Teller: the bear is small **Drawer:** **Teller:** the girl is to the left facing left looking at the bear with her leg out scared facing right **Drawer:**

Teller: there is a small dog below the girl and a angry boy to left facing left with a racket in the left hand **Drawer:** what is next

Teller: there is a small helicopter at the top in the sky in between the boy and girl **Drawer:** right above the bear



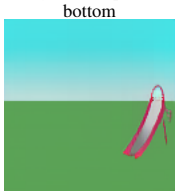
Teller: there is a cloud in the top left corner it is cut off on the top and left sides two puffs on the right and 3 puffs on the bottom

Teller: an inch from the right of the cloud are small balloons the orange balloon is on the right

Teller: airplane 1 4 inch to the right of the balloon the nose of the plane is in line with the yellow balloon it faces left

Teller: angry boy sitting below the plane facing left about 1 2 inch grass above and below him

Teller: a girl sits to the left of the boy facing him feet almost touching surprised wearing viking hat top brown part of hat touches horizon



Drawer: ready **Teller:** a small bear close to the right side one finger off picture small sliding on his left side bear left foot touching the sliding

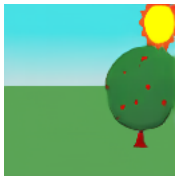
Teller: table medium in the center front

Teller: medium pine tree behind the table one inch

Teller: sad girl standing far left part hand missing sad face medium

Teller: boy sitting on her right side look mad

Figure 5. Random selection of examples generated by our Mismatch model for the CoDraw dataset.



Teller: medium sun on your right hand with a small half apple tree under it

Teller: small snake left of apple tree snake facing left

Teller: small bushy tree is on your left hand side hole facing left 1 2 inch from side 1 inch from top

Teller: large table in left hand corner slanted right north with medium owl on right end facing right

Teller: bushy tree looks like it 's sitting on table

Drawer: i 'm ready **Teller:** boy flying a kite

Teller: wearing blue t shirt **Teller:** blue shoes **Teller:** lighter blue shorts **Drawer:** got it **Teller:** has black hair **Drawer:** i have all of him **Teller:** sunny outside

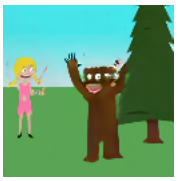
Teller: a girl next to a grill

Teller: i am sorry only one girl in the scene **Teller:** next to a girl **Teller:** on the grill **Drawer:** is a burger **Teller:** there are hamburgers

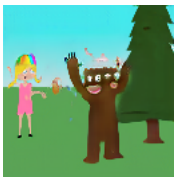
Teller: 3 of them **Teller:** she is wearing pink overall holding ketchup in her hand and she is wearing glasses on

Drawer: ready when you are **Teller:** right side medium pine tree cut in half by right edge and cut top **Drawer:** got it

Teller: left from pine tree big size bear facing left head above horizon



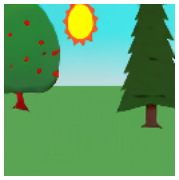
Teller: left side running angry big size girl facing right horn hat holding football in right hand our right



Teller: now behind girl they overlap running angry boy facing right purple glasses witch hat on facing right half body above horizon **Drawer:** got it

Teller: movie girl down a bit move boy right a bit lines with girl next to girls what is a frisbee

Drawer: what is in the scenery of the image **Teller:** medium pine tree on right trunk is about 1 inch from bottom and trunk is 1 2 from right top is cut off on top and right **Drawer:** what 's next



Teller: on left is medium apple tree cut off on left trunk is halfway down grass big sun in center of two trees **Drawer:** what 's next

Teller: on the right side of apple tree trunk is sad boy sitting legs out facing right hes 1 inch from bottom very close to trunk **Drawer:** what 's next

Teller: to right of boy align with his hand is a fire to right of fire is girl kneeling smiling one arm up girl and boy small **Drawer:** what 's next

Teller: girls right hand overlaps pine tree a little left side of sun is overlapped by apple tree below fire is ketchup left mustard right **Drawer:** next

Teller: ready **Drawer:** what what is the first object and location **Teller:** small table middle of green **Drawer:** next

Teller: there is a pine to the left med size upper peck can not see **Drawer:** next

Teller: med sun in upper right corner **Drawer:** any of it cut off

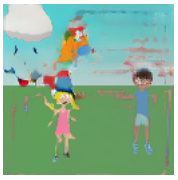
Teller: no sun is whole straight down from sun is a boy standing with a laugh on face **Drawer:** facing left

Teller: yes boy is almost at bottom of screen **Drawer:** what else **Teller:** there is a girl standing at corner of table looks likes she is running smiling **Drawer:** next

Figure 6. Random selection of examples generated by our G prior model for the CoDraw dataset.

Drawer: go **Teller:** large rain cloud left corner touches side cut off on top drops almost touch grass **Drawer:** next

Teller: large rocket on right tip of cloud flying left with very small girl sad legs out sitting on its upper wing **Drawer:** sitting on rocket the rocket is middle scene



Teller: rocket overcloud large regular cloud on right side cut off on top and side a bit surprised boy legs out facing left under cloud **Drawer:** next

Teller: cat facing boy 1 2 inch to left of his feet **Drawer:** next

Teller: i will check and send adjustments **Drawer:** yes i do n't have the girl tell me where is the girl **Teller:** she is sitting on the rockets upper wing her back arm is under the window **Drawer:** check

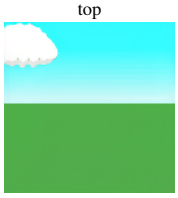
Drawer: where is jenny and mike **Teller:** on left hand side of the screen 2 inches above bottom is a pine tree cut off at top

Teller: straight down from tree is jenny legs crossed facing right right arm in the air

Teller: next to jenny is mike same level standing facing right with arms out mouth open

Teller: on the right hand side inch from the bottom is a duck facing jenny and mike

Teller: straight above duck is soccer ball



Teller: large cloud in left corner top and left are off screen

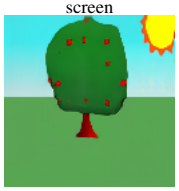
Teller: large apple tree right side top of trunk lines up with horizon right side and top of screen

Teller: in front of trunk little over to the right of trunk by right side is a soccer ball

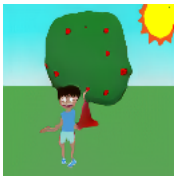
Teller: left side large girl angry face sitting cross legged facing right

Teller: neck on horizon line she 's holding a baseball in her up hand and wearing rainbow hat

Drawer: what do we have **Teller:** med sun right corner **Drawer:** and



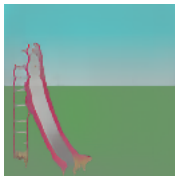
Teller: middle of green with trees half in blue half in green is a apple tree med size **Drawer:** and



Teller: med boy standing on right of tree **Drawer:** face expression and where are his hands **Teller:** he has a tennis ball in right hand he is smiling showing teeth **Drawer:** and

Teller: right arm sticking out across from him is a girl almost to the left edge left arm down right hand with a ball glove **Drawer:** smiling please give more elaborations

Teller: yes smiling looking right she is standing her middle is on the line of green **Drawer:** and



Teller: large slide to the left facing right with owl sitting on top of platform **Drawer:** top of handles or where we stand

Teller: small cloud in top center a bit to the right owl is on the platform **Drawer:** go

Teller: there is a medium to small apple tree on right half in blue half in green right side of tree cut off a bit **Drawer:** go

Teller: a dog directly under tree facing left **Drawer:** size

Teller: a girl standing in front of slide arms up smiling dog looks small

Figure 7. Random selection of examples generated by our Aux model for the CoDraw dataset.

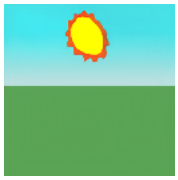
Teller: large pine tree on right trunk 1 4 inch from bottom cut off top and right

Teller: large sand box on left mound on left half bottom edge cut off and left corner cut off mound fully visible

Teller: middle of sandbox large girl running right big toothy smile

Teller: bottom left corner of sandbox large cat looking right top right corner sandbox is baseball

Teller: girl is wearing pirate cap will check when ready



Teller: there is a large sun in the left hand corner

Teller: there is a small girl facing left she has one leg in the air 3 inches from the right 1 2 an inch from the bottom

Teller: small duck facing girl bill at her foot like she is going to kick the duck but move the duck 1 4 inch away about an inch from bottom

Teller: soccer ball to the left of the duck to the left 1 4 inch and up 2 inches

Teller: boy like he is running facing the duck and girl they playing keep away from duck his arms up and looks like he is running mouth open **Teller:** he has glasses on

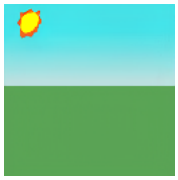
Teller: top left corner is large sun half of it hidden **Drawer:** what else **Teller:** top right corner is large cloud half of it hidden **Drawer:**

Teller: on left side of screen is a large girl standing with her arms in front of her facing right her eyes are even with the horizon line **Drawer:**

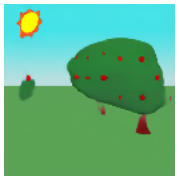
Teller: girl is wearing a chefs hat sunglasses and is holding a pink shovel in her right hand **Drawer:**

Teller: on right side of screen is large boy one hand on hip facing the girl with angry expression his neck is even with horizon line **Drawer:** got it what else

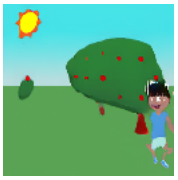
Teller: between them is a large grill **Drawer:**



Teller: big sun on left only top cut a little bit the yellow part



Teller: on right sun small apple tree cut top



Teller: on right of tree happy mike hand front stand facing left head touches the tree

Teller: below sun small sand mike is medium box mound on left left corner is hidden a small duck inside the sandbox

Teller: medium jenny sits on right side of sandbox crossed legs and one hand up facing right head cover right corner of sandbox

Teller: sun top right almost touching top 1 " from left medium size

Teller: big bear at left facing right arm pits at horizon

Teller: medium to small size girl about 3 " away from bear facing left angry arms out holding a bat in right hand head in the blue area

Teller: boy behind her about 1 5 " away arms up startled glove on left hand rainbow hat on his eyes slightly below horizon

Teller: move boy down about an inch face him to left and his striped hat and we are good

Figure 8. Random selection of examples generated by our *D* Concat model for the CoDraw dataset.

Teller: big cloud top left side **Drawer:** got it

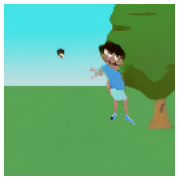
Teller: on the right side is a swing big size **Drawer:** any parts cut off

Teller: girl on the left side neg horizon one arm up facing right happy face **Drawer:** got it

Teller: one part cut from swing just a bit from the right **Drawer: Teller:** next to the right of the cloud is a basketball a bit over the cloud **Drawer:**

Teller: a boy is on the swing the right sit legs cross surprised face facing left color hat baseball glove **Drawer:** got it

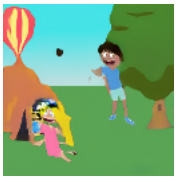
Drawer: ready **Teller:** big oak on right hole facing right hole almost touching horizon **Drawer:** if its large it is huge how much is cut off on the right



Teller: smiling big hands out mike on right left trunk point touching his back hair above horizon

Teller: small hot balloon on left 1 4 from top 1 in from left big tent on left facing right cut off slightly on back top above horizon

Teller: smiling big hands out jenny is in front of left opening tent



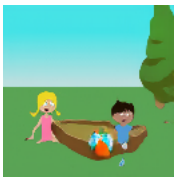
Teller: she has an owl sitting on her left wrist her hand is in the dark opening **Drawer:** she is facing him and large owl right

Drawer: go **Teller:** small bushy tree facing left owl on right middle of tree

Teller: to the left of tree is a medium sandbox mound on right close to bottom **Drawer:** next

Teller: girl sitting in left corner indian style smiling one arm up **Drawer:** next

Teller: boy in right corner sitting indian style smiling with arms open both facing right **Drawer:** next



Teller: under boys right hand is cup in sand straw to left to left of cup is medium beach ball

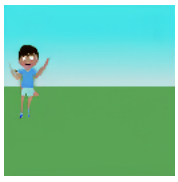
Teller: tree hole facing left cut off from right side a little bit top hiding a bit of the sun

Teller: bumblebee with ear touching the bottom left of tree trunk facing to the right side **Drawer:** got it

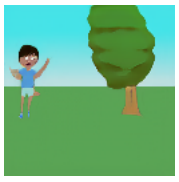
Teller: girl sitting smiling facing right hand behind her one inch from side wearing crown

Teller: crown almost touches the horizon **Drawer:** got it **Teller:** boy faces girl sitting smiling his feet r half inch from hers and raised up a little he wears a beanie with top of it just at horizon

Teller: duck between the two with ducks feet level to boys top foot

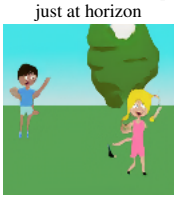


Teller: on the left an inch from the edge is a boy **Drawer:** what is he doing **Teller:** he is facing right standing one hand up teeth showing and holding a racket with one hand that is in the air

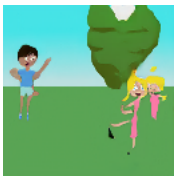


Teller: next to him a medium tree hole facing left head touches the tip of last branch truck aligns with his waist

Teller: on the right 1 inch from edge is a girl sad looking left one hand in the air head aligns with the boy 's **Drawer:** her left hand cut off



Teller: above her is a small cloud right above her can i check



Teller: no about 2 cm from the edge the hand **Teller:** move her 1 more cm from the edge she is holding a yellow small ball in the hand in air

Figure 9. Random selection of examples generated by our D Subtract model for the CoDraw dataset.

Drawer: what is there **Teller:** big thunderbolt on left bolt facing right touching on left edge close to top **Drawer:** and **Teller:** big raindrop cloud to the right of thunderbolt cloud cutting it off on right side a little big shocked jenny with arms up **Drawer:** where is she **Teller:** head right below horizon sad big sitting mike with legs facing right is beside her **Drawer:** and **Teller:** he 's wearing a star hat soccer ball is covering his left foot and shin **Drawer:** and

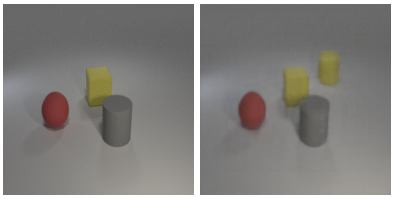
Teller: ready **Drawer:** and ready **Teller:** upper right corner large sun with right edges a bit cut off and top cut off **Teller:** under sun happy boy standing facing left with right arm up his shoulders just above horizon line **Teller:** he is wearing a pirate hat it touches on of the sun tips on the left side **Teller:** happy girl kicking on left side she is about 1 5 inches in from left side her mouth is at the horizon line **Drawer:** got it **Teller:** just a tiny bit off of girls kicking foot is a beach ball a cloud is over the girl towards the right center **Drawer:** is the cloud on the right or the sun you said sun upper right corner

Teller: boy left side kicking leg facing right his half torso aligns with horizon he is shocked **Drawer:** go **Teller:** finger away from his leg soccer ball its bottom part touches horizon **Teller:** right side medium tree 1 4 cut off right side and trunk half way in grass with slight cut off as well right side hole facing left **Drawer:** go **Teller:** plain cloud top middle top part cut off big size in front of tree dog its legs behind completely cut off and it 's facing left **Teller:** near dog is a big cat its tail cover 's dog 's front leg slightly and facing right and then girl sitting smiling facing right

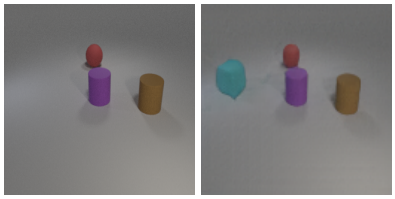
Drawer: ready **Teller:** top left facing left one 1 4 inch from side blade touch top small helicopter facing left **Drawer:** what 's in the left the helicopter **Teller:** nothing it is a 1 4 inch from side flying left **Drawer:** got it it 's tiny right **Teller:** yes **Teller:** below copter is large boy facing right arms out mouth open neck at horizon **Teller:** right of boy his top hand is on first plank is a large picnic table left top corner is highest point pie is there in corner **Drawer:** which side is the pie **Teller:** right of pie is large girl facing left standing with smile no teeth one arm up and one down pie top left corner **Drawer:** where is she to the horizon and she is in front of the table **Teller:** girl in front of table nose at horizon top right corner of table is ketchup **Drawer:** got it **Teller:** 1 2 inch from right side and 1 2 inch from horizon is large grill

Drawer: ready **Teller:** there is a medium in the center of the sky just below the top edge **Drawer:** medium cloud **Teller:** oh sorry medium sun the medium cloud is down and to the right in the sky **Teller:** there is a small oak tree on the left an inch away from the left edge hole facing right 2 3s of the leaves are above the horizon **Teller:** on the right side the kids are both medium sized and facing left jenny is happy jumping half inch from the right edge

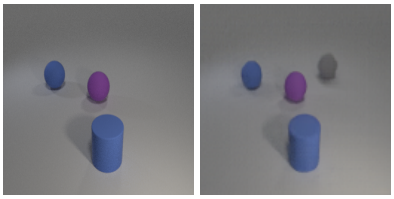
Figure 10. Random selection of examples generated by our Non-iterative model for the CoDraw dataset.



(a) **Left:** Initial Image **Right:** Final Image **Instruction:** Add a yellow cylinder behind the gray cylinder on the right and behind the yellow cube on the right



(b) **Left:** Initial Image **Right:** Final Image **Instruction:** Add a cyan cube behind the brown cylinder on the left and behind the purple cylinder on the left



(c) **Left:** Initial Image **Right:** Final Image **Instruction:** Add a gray sphere behind the blue cylinder on the right and behind the purple sphere on the right

Figure 11. When GeNeVA-GAN is provided with an initial image different from the background image used during training, it still adds the desired object with the right properties at the correct location. The model was not trained in this setting and the success of this experiment demonstrates that it has learnt to preserve the existing canvas, understand the existing objects, and add new objects with the correct relationships to existing objects.