Make-A-Story: Visual Memory Conditioned Consistent Story Generation –Supplemental–

Tanzila Rahman[1,3] Hsin-Ying Lee[2] Jian Ren[2] Sergey Tulyakov[2]
Shweta Mahajan[1,3] Leonid Sigal[1,3,4]
[1]University of British Columbia [2]Snap Inc.
[3]Vector Institute for AI [4]Canada CIFAR AI Chair

Method	Reference-text	Char-acc (↑)	Char-F1 (↑)	BG-acc (↑)	BG-F1 (↑)	FID (↓)
Story-LDM (Ours)	✓	69.19	86.59	35.21	**28.80**	69.49
Story-LDM (Ours)	×	**83.29**	**94.61**	**35.54**	27.32	**64.89**

Table 3. **Experimental results using our proposed Story-LDM with and without the reference text on the FlintstonesSV dataset.**

1. Additional Quantitative Results

In Tab. 3, we show the performance of our Story-LDM approach with and without the reference text. We observe that without reference text *i.e.*, with explicit mentions of the characters (by their name), our approach outperforms the prior state-of-the-art VLCStoryGAN [32] (*cf*. Tab. 2, row 1 in the main paper) and the strong LDM [42] baseline on character accuracy by nearly $\sim 3.4\%$.

This demonstrates that introduction of our Memory attention module to the pipeline of the diffusion model with U-Net architecture increases the performance even in the traditional dataset setting. This can be attributed to the fact that even when textual resolution of character names or setting is unnecessary, our memory attention module can still enhance consistency of appearance in the visual domain. When using the descriptions with references, the character accuracy of the model drops by $\sim$ 12% showing the difficulty of the extended task of generating stories from the co-referenced text. Even the strong LDM baseline which outperforms the prior state-of-the-art for story generation, offers limited performance on the complex task of story generation with co-reference resolution illustrating the hardness and complexity of the new task (*cf*. Tab. 2 in the main paper). Our Story-LDM approach outperforms this strong baseline illustrating the benefits of our autoregressive Story-LDM with memory module.

Table 5 shows quantitative results for consistent story generation with the reference text on the PororoSV [27] dataset. We compare our proposed Story-LDM with DUCO-STORYGAN [32] and VLCStoryGAN [31]. Our method outperforms previous baseline models (including LDM baseline) in character evaluation metrics. To be noted, PororoSV dataset has no background information, therefore we only perform character level evaluation on this dataset.

Moreover, we conduct human evaluation. We randomly select 10 examples to compare Story-LDM with LDM. Following [31], we select three evaluation criteria: visual quality, consistence, and relevance for each sample. We conducted forced choice experiment with 13 subjects. Preference rate for our Story-LDM is 74.7%, 65.4% & 69.2% in terms of listed criteria.

2. Additional Qualitative Results

In Fig. 9 we provide example image frames from our extended MUGEN dataset with three characters *Lisa, Tony* and *Jhon*, and six different backgrounds. Our extended MUGEN dataset is thus more complex than the MUGEN dataset in [16] where only one character is considered for two backgrounds.

Fig. 10 shows examples of positive/negative samples. To evaluate fore- and back-ground consistency we propose two evaluation metrics: character and background classification. The generated story is considered accurate (positive sample) if it appropriately resolves character and background references in generation, otherwise it is considered a negative sample. In the figure, the 1st row (*i.e.* positive sample) has 100% character and background accuracy; the 2nd row (*i.e.* negative) has 75% and 50% character and background accuracy.

Figs. 11, 12 and 13 show random samples obtained for story generation with our Story-LDM approach on the MUGEN, FlintstonesSV and PororoSV datasets respectively. Clearly, our approach yields high quality frames with character and background consistency.

Dataset	Method	w/ ref. text	Char-acc (↑)	Char-F1 (↑)	BG-acc (↑)	BG-F1 (↑)	FID (↓)
PororoSV	DUCO-STORYGAN [32]	✓	13.97	38.01	-	-	96.51
	VLCStoryGAN [31]	✓	17.36	43.02	-	-	84.96
	LDM [42]	✓	16.59	56.30	-	-	60.23
	Story-LDM (Ours)	✓	**20.26**	**57.95**	-	-	**36.64**

Table 4. **Quantitative results.** Experimental results on the PororoSV datasets.

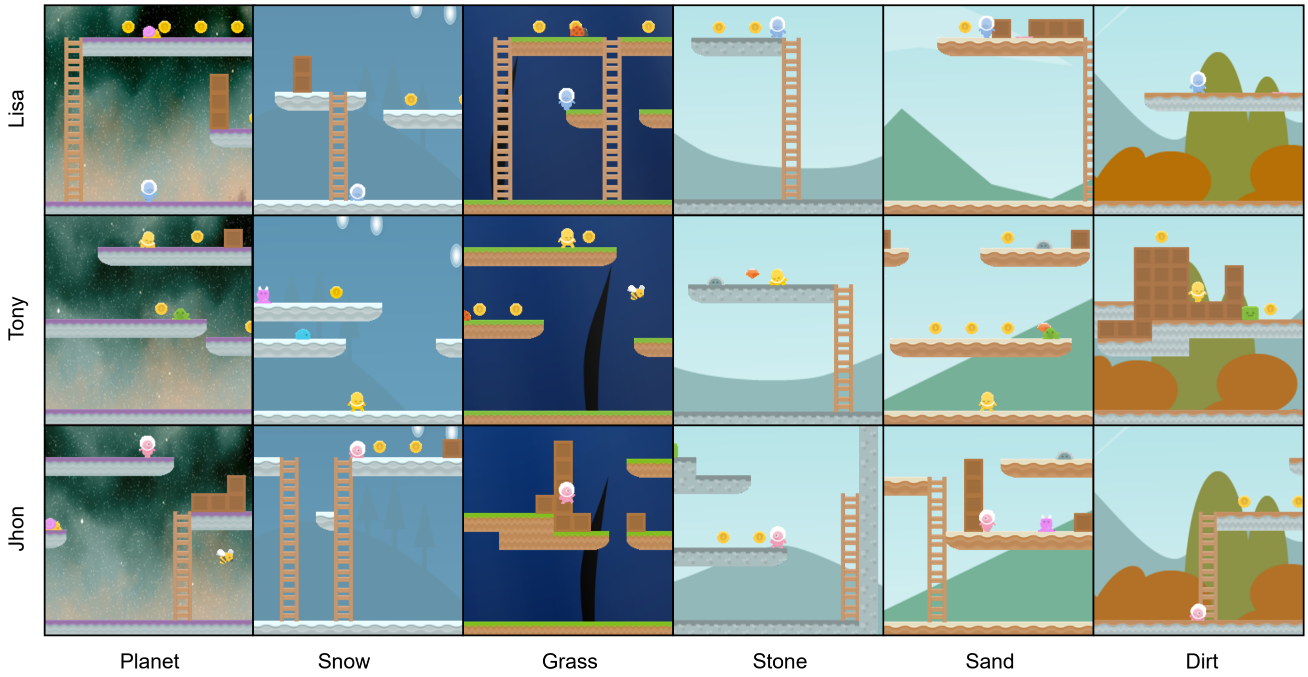

Figure 9. **Combination of different characters and backgrounds from the MUGEN dataset [16].**

Dataset	Method	w/ ref. text	Char-acc (↑)	Char-F1 (↑)	BG-acc (↑)	BG-F1 (↑)	FID (↓)
Flintstones	VLCStoryGAN [31]	×	27.73	42.01	4.83	16.49	120.85
	LDM [42]	×	79.86	92.33	**48.02**	**37.86**	**61.40**
	Story-LDM (Ours)	×	**83.29**	**94.61**	35.54	27.32	64.89
MUGEN	LDM [42]	×	95.25	97.04	21.10	23.98	123.69
	Story-LDM (Ours)	×	**97.60**	**98.44**	**74.72**	**80.51**	**79.41**

Table 5. **Experimental results using non-referential text on the Flintstones and Mugen dataset.**

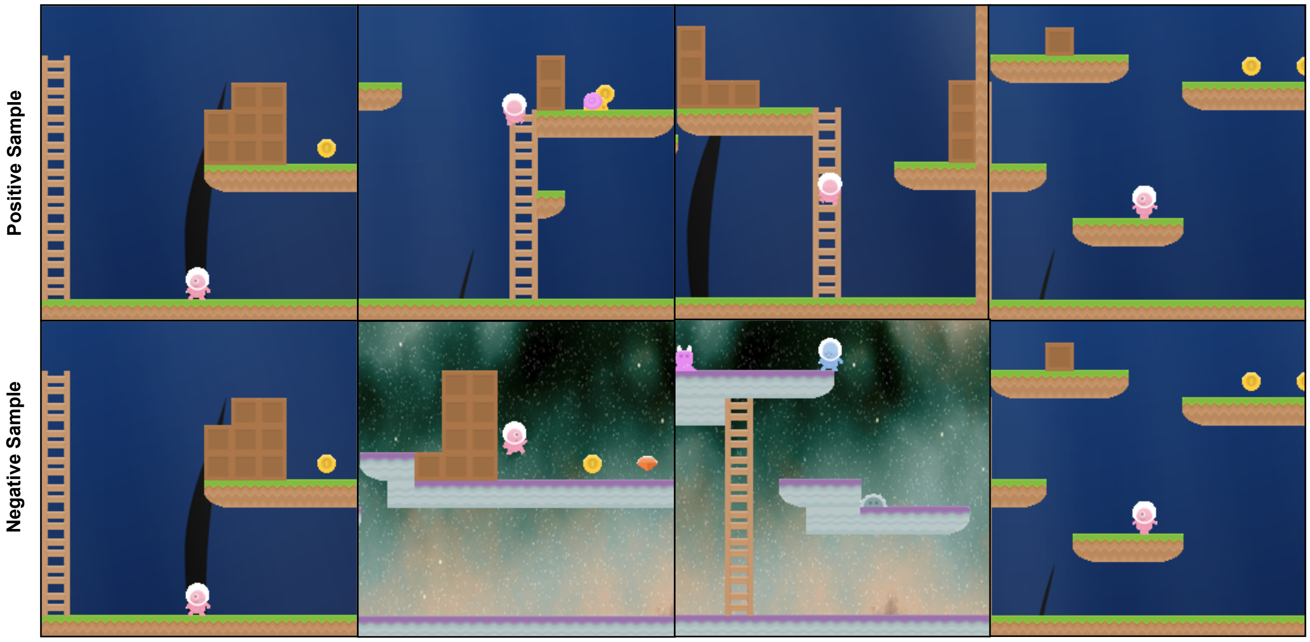

Figure 10. **Positive (1st row) and negative (2nd row) examples.**

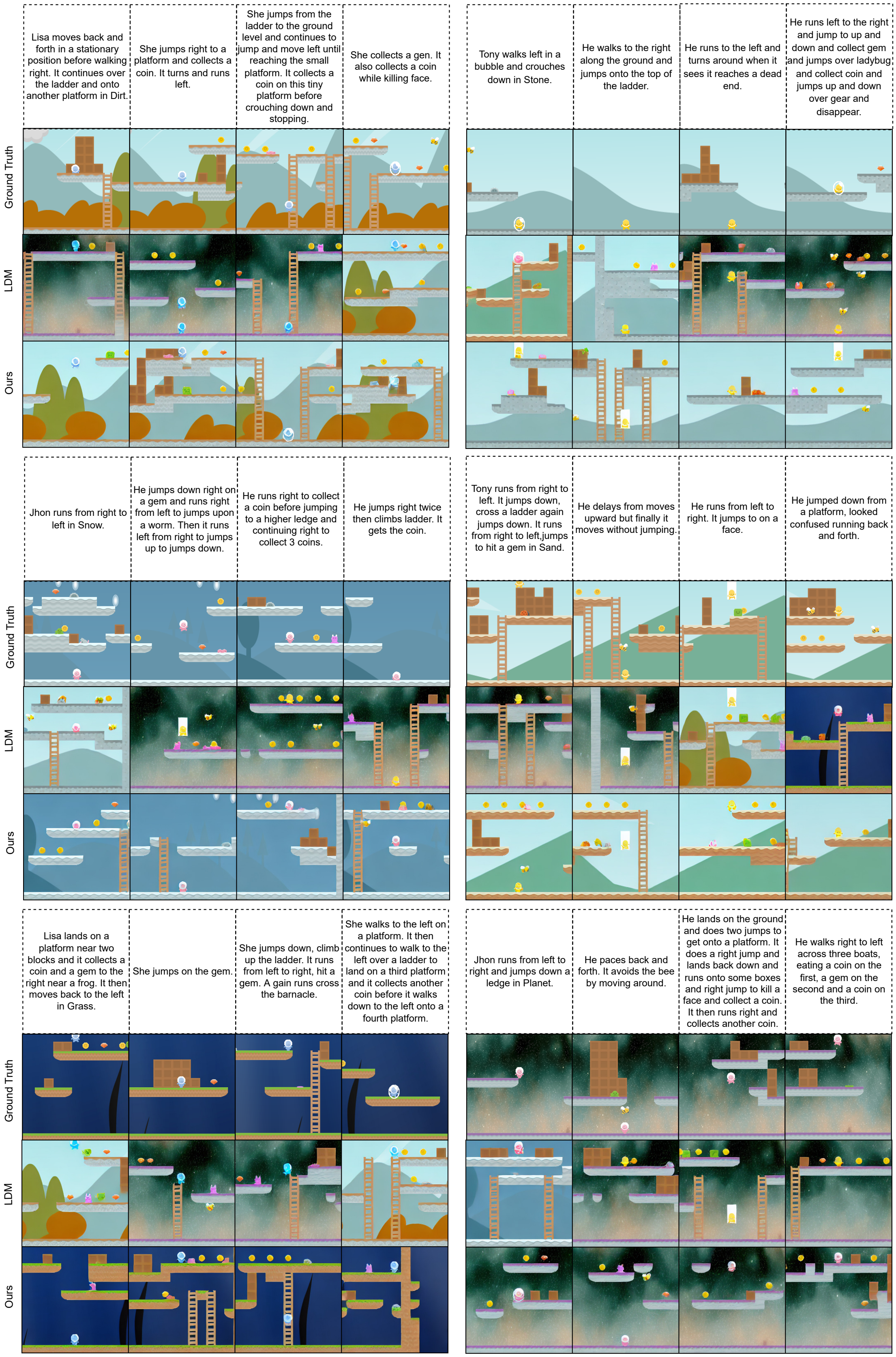
Lisa moves back and forth in a stationary position before walking right. It continues over the ladder and onto another platform in Dirt.
She jumps right to a platform and collects a coin. It turns and runs left.
She jumps from the ladder to the ground level and continues to jump and move left until reaching the small platform. It collects a coin on this tiny platform before crouching down and stopping.
She collects a gen. It also collects a coin while killing face.
Tony walks left in a bubble and crouches down in Stone.
He walks to the right along the ground and jumps onto the top of the ladder.
He runs to the left and turns around when it sees it reaches a dead end.
He runs left to the right and jump to up and down and collect gem and jumps over ladybug and collect coin and jumps up and down over gear and disappear.
Ground Truth
LDM
Ours
Jhon runs from right to left in Snow.
He jumps down right on a gem and runs right from left to jumps upon a worm. Then it runs left from right to jumps up to jumps down.
He runs right to collect a coin before jumping to a higher ledge and continuing right to collect 3 coins.
He jumps right twice then climbs ladder. It gets the coin.
Tony runs from right to left. It jumps down, cross a ladder again jumps down. It runs from right to left,jumps to hit a gem in Sand.
He delays from moves upward but finally it moves without jumping.
He runs from left to right. It jumps to on a face.
He jumped down from a platform, looked confused running back and forth.
Ground Truth
LDM
Ours
Lisa lands on a platform near two blocks and it collects a coin and a gem to the right near a frog. It then moves back to the left in Grass.
She jumps on the gem.
She jumps down, climb up the ladder. It runs from left to right, hit a gem. A gain runs cross the barnacle.
She walks to the left on a platform. It then continues to walk to the left over a ladder to land on a third platform and it collects another coin before it walks down to the left onto a fourth platform.
Jhon runs from left to right and jumps down a ledge in Planet.
He paces back and forth. It avoids the bee by moving around.
He lands on the ground and does two jumps to get onto a platform. It does a right jump and lands back down and runs onto some boxes and right jump to kill a face and collect a coin. It then runs right and collects another coin.
He walks right to left across three boats, eating a coin on the first, a gem on the second and a coin on the third.
Ground Truth
LDM
Ours

Figure 11. **Qualitative results on the MUGEN dataset [16].**

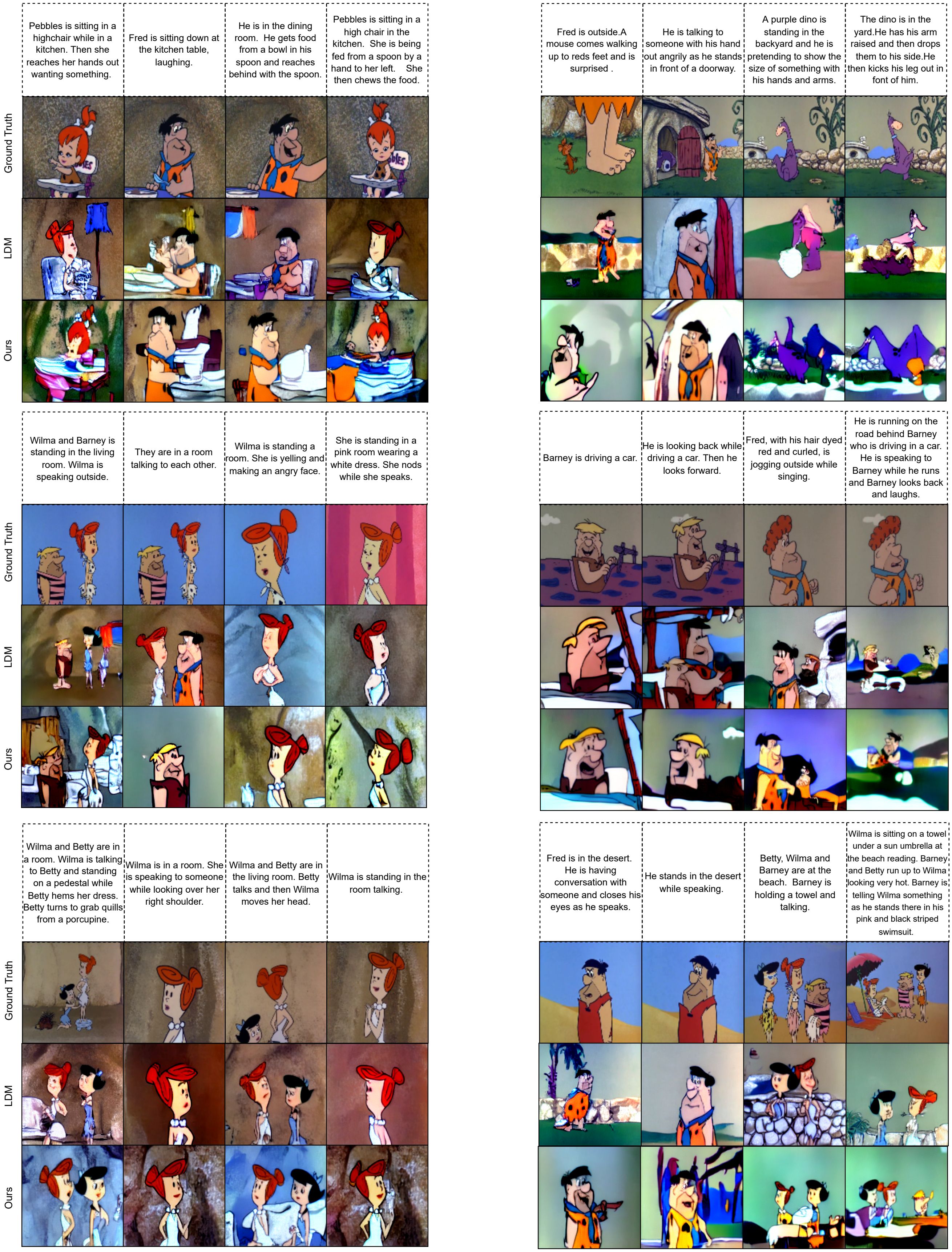
Pebbles is sitting in a highchair while in a kitchen. Then she reaches her hands out wanting something.
Fred is sitting down at the kitchen table, laughing.
He is in the dining room. He gets food from a bowl in his spoon and reaches behind with the spoon.
Pebbles is sitting in a high chair in the kitchen. She is being fed from a spoon by a hand to her left. She then chews the food.
Fred is outside.A mouse comes walking up to reds feet and is surprised .
He is talking to someone with his hand out angrily as he stands in front of a doorway.
A purple dino is standing in the backyard and he is pretending to show the size of something with his hands and arms.
The dino is in the yard.He has his arm raised and then drops them to his side.He then kicks his leg out in font of him.
Ground Truth
LDM
Ours
Wilma and Barney is standing in the living room. Wilma is speaking outside.
They are in a room talking to each other.
Wilma is standing a room. She is yelling and making an angry face.
She is standing in a pink room wearing a white dress. She nods while she speaks.
Barney is driving a car.
He is looking back while driving a car. Then he looks forward.
Fred, with his hair dyed red and curled, is jogging outside while singing.
He is running on the road behind Barney who is driving in a car. He is speaking to Barney while he runs and Barney looks back and laughs.
Ground Truth
LDM
Ours
Wilma and Betty are in a room. Wilma is talking to Betty and standing on a pedestal while Betty hems her dress. Betty turns to grab quills from a porcupine.
Wilma is in a room. She is speaking to someone while looking over her right shoulder.
Wilma and Betty are in the living room. Betty talks and then Wilma moves her head.
Wilma is standing in the room talking.
Fred is in the desert. He is having conversation with someone and closes his eyes as he speaks.
He stands in the desert while speaking.
Betty, Wilma and Barney are at the beach. Barney is holding a towel and talking.
Wilma is sitting on a towel under a sun umbrella at the beach reading. Barney and Betty run up to Wilma looking very hot. Barney is telling Wilma something as he stands there in his pink and black striped swimsuit.
Ground Truth
LDM
Ours

Figure 12. **Qualitative results on the FlintstonesSV dataset [14].**

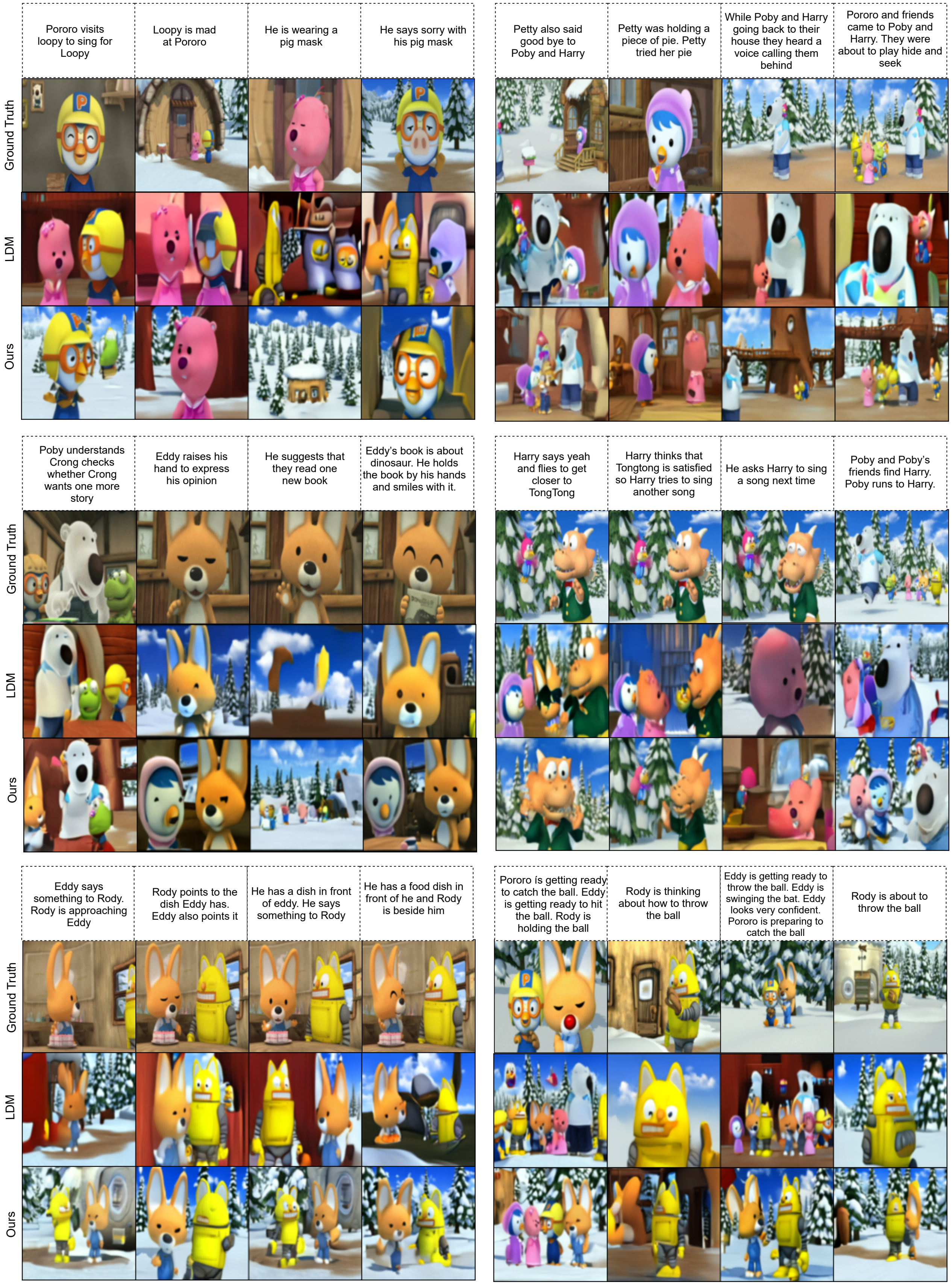
Pororo visits loopy to sing for Loopy
Loopy is mad at Pororo
He is wearing a pig mask
He says sorry with his pig mask
Petty also said good bye to Poby and Harry
Petty was holding a piece of pie. Petty tried her pie
While Poby and Harry going back to their house they heard a voice calling them behind
Pororo and friends came to Poby and Harry. They were about to play hide and seek
Ground Truth
LDM
Ours
Poby understands Crong checks whether Crong wants one more story
Eddy raises his hand to express his opinion
He suggests that they read one new book
Eddy's book is about dinosaur. He holds the book by his hands and smiles with it.
Harry says yeah and flies to get closer to TongTong
Harry thinks that Tongtong is satisfied so Harry tries to sing another song
He asks Harry to sing a song next time
Poby and Poby's friends find Harry. Poby runs to Harry.
Ground Truth
LDM
Ours
Eddy says something to Rody. Rody is approaching Eddy
Rody points to the dish Eddy has. Eddy also points it
He has a dish in front of eddy. He says something to Rody
He has a food dish in front of he and Rody is beside him
Pororo is getting ready to catch the ball. Eddy is getting ready to hit the ball. Rody is holding the ball
Rody is thinking about how to throw the ball
Eddy is getting ready to throw the ball. Eddy is swinging the bat. Eddy looks very confident. Pororo is preparing to catch the ball
Rody is about to throw the ball
Ground Truth
LDM
Ours

Figure 13. **Qualitative results on the PororoSV dataset [27].**