

CoIn Supplementary Material

For story visualization and continuation tasks, comparing a large number of synthesized story images conditioned on different story descriptions is the most direct and accurate way to evaluate the performance. Thus, in supplementary material, we compare more story images synthesized by WL-SV [3], StoryDALL-E [4], and our CoIn on the Pororo-SV [2] and Flintstones-SV [1] datasets.

Figures 1 and 2 show examples of visual comparisons between our CoIn and the baselines on story visualization and continuation tasks, respectively. As shown in Figure 1, the characters synthesized by WL-SV contain break or wrong shapes. Moreover, there are some wrong characters synthesized in the results. However, the story images synthesized by our CoIn have correct shapes and clear backgrounds. For story continuation, the StoryDALL-E cannot synthesize target frames that are visually consistent with the first frame. But our CoIn can synthesize a more consistent visual appearance. And the story images synthesized by our CoIn are more realistic.

1. fred is on the sidewalk . a bunch of children are running past him .
2. fred is standing in a room frowning .
3. fred and barney are standing outside next to the stone wall . barney is wearing an outfit that makes him look like a boy scout . fred says something to barney and then points at him .
4. fred and barney are standing on a sidewalk . barney is speaking to fred , while fred listens silently with his hands on his hips .
5. barney is outside talking .



1. fred is sitting at a table in a dining room . then a hand reaches across the table to adjust the white flower pinned on the front of fred 's tunic .
2. fred is sitting in a chair at a restaurant pulled up to a table with two coffee mugs . an off screen character reaches over and pulls a white object off of his chest and he then blinks three times .
3. fred and barney argue in the dining room .
4. fred and wilma are at a table in the restaurant . fred says something to wilma and then looks to his left . wilma looks mad .
5. wilma is standing in a room and talking .



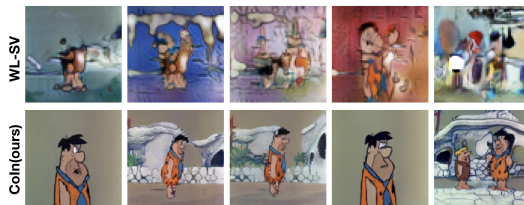
1. fred is talking standing in the kitchen . he is standing by the fridge .
2. fred is in the yard . fred is carrying a ladder .
3. fred is talking while in a car .
4. fred and barney are in a car . fred is driving with his eyes closed while barney sits in the passenger seat .
5. fred is driving a car while talking with barney in the passenger seat .



1. an old man with white beard and bald head dressed in white is speaking to fred in a room . the old man is leaning across a table while fred is sitting in a chair .
2. fred is in the kitchen . he is pointing at something to his right and talking
3. fred is in an office . an older man is walking past him talking .
4. fred is sitting in a chair in a dining room , and looking away from the table .
5. fred and barney are in a room . fred is sitting in a chair listening to barney talk while barney is sitting on the table .



1. fred is standing in a room with a blank face talking to someone on his side .
2. fred is walking down the sidewalk , talking to someone off camera right .
3. fred is on the sidewalk . a bunch of children are running past him .
4. fred is standing in a room frowning .
5. fred and barney are standing outside next to the stone wall . barney is wearing an outfit that makes him look like a boy scout . fred says something to barney and then points at him .



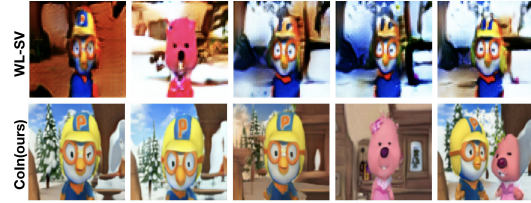
1. pororo imagines pororo dancing for loopy .
2. pororo visits loopy to sing for loopy .
3. loopy is mad at pororo .
4. pororo is wearing a pig mask .
5. pororo says sorry with pororo pig mask .



1. pororo is satisfied with pororo .
2. loopy approaches with angry face .
3. pororo and pororo friends say hi to loopy .
4. loopy says hello with angry face .
5. loopy asks something to everyone .



1. pororo and pororo friends are talking about pororo being secret friend .
2. pororo and pororo friends are talking about pororo being secret friend .
3. pororo is satisfied with pororo .
4. loopy approaches with angry face .
5. pororo and pororo friends say hi to loopy.



1. poby finds his house under the cliff .
2. poby knows the way to his house . poby smiles.
3. the bulldozer arrives safely at poby 's house.
4. pororo and his friends are relaxed to arrive at poby 's house . eddy and rody look back.
5. rody looks at eddy . eddy asks did they have fun.



1. pororo imagines that crong is drawing a plenty of pictures on the floor with a happy . pororo cannot repress pororo 's astonishment pororo shakes his head passionately .
2. pororo ran to his house with an anger face . pororo 's friends look at the pororo 's back.
3. there is pororo 's house . crong is sitting down on the couch and reading a book.
4. crong is sitting down on the couch and reading a book . pororo is watching the floors and walls with a look filled with doubt.
5. crong is sitting down on the couch and reading a book . pororo is watching the floors and walls with a look filled with doubt.

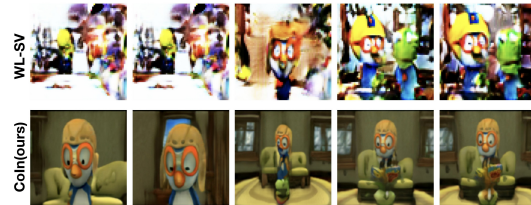


Figure 1: Comparison of story visualization results between WL-SV and our proposed Coln.

1. fred is sitting in a chair at a restaurant pulled up to a table with two coffee mugs . an off screen character reaches over and pulls a white object off of his chest and he then blinks three times .
2. fred and wilma argue in the dining room .
3. fred and wilma are at a table in the restaurant . fred says something to wilma and then looks to his left . wilma looks mad .
4. wilma is standing in a room and talking .
5. wilma is standing in the room talking to someone off screen left . she has bones in her hair .



1. fred explains his reason for being angry to barney while driving the car .
2. barney and fred are outside driving in a car . barney is talking .
3. fred , betty and wilma are all standing in a room next to each other . fred is talking to both of them while the women seem concerned .
4. wilma and betty are talking to a woman with yellow hair at the amusement park .
5. fred and barney are crawling on their knees in a room with their eyes closed . fred speaks , then stops speaking . after that , barney speaks .



1. betty is standing in the room talking . wilma is standing next to betty .
2. wilma and betty are standing in a room . betty is holding her arms up while talking and wilma turns to look at her .
3. betty and wilma gossiping in the room .
4. fred and barney are standing in front of a cafe . fred is reading the cafe menu . then barney nods his head in assent .
5. a paper boy calls for the news in a room .



1. wilma and betty are having tea in the kitchen while listening to the radio .
2. wilma and betty look out a window while wilma talks and betty listens .
3. wilma and betty are at a window . wilma speaks .
4. fred is in the living room . he looks dejected as he speaks .
5. fred sitting in chair in living room



1. fred is in a room marking on a calendar .
2. wilma is in the living room . she appears to be singing and dancing across the room .
3. fred is marking of the calendar and shaking his head in a room .
4. fred is sitting at the dining room table holding a knife and fork , ready to eat a large chicken leg on the table .
5. wilma is in a room talking to someone over her shoulder .



1. eddy is in diving suit and equipment at the sea shore with tongtong and pororo . eddy is smiling . the sky is blue and all the ground mountains and trees are covered in snow .
2. pororo says something and tongtong is nodding his head . the sky is blue and all the ground mountains and trees are covered in snow .
3. eddy is in diving suit and equipment . eddy is smiling . the sky is blue and all the ground mountains and trees are covered in snow .
4. eddy is in diving suit and equipment . eddy is smiling . the sky is blue and all the ground mountains and trees are covered in snow .
5. pororo is smiling as well making the cheer up gesture . tongtong is beside pororo . the sky is blue and all the ground mountains and trees are covered in snow .



1. pororo and crong are hearing someone who calls them . pororo is curious about that so asks to the crong what it is .
2. finally eddy is facing pororo 's front face with his telescope . pororo 's front face is well captured with full size .
3. pororo and crong are curious about what they heard . therefore pororo and crong are shaking their heads with murmuring .
4. there isn 't any person in this scene . only for two snowballs thrown by someone .
5. maybe pororo is struck by snowballs which was thrown by someone . pororo seems angry and crong looks shocked . pororo and crong are quickly running away .



1. crong pororo harry rody and eddy are cheering while sitting around the table .
2. petty is holding plates on both hands and smiling .
3. eddy rody harry crong pororo and loopy gather around petty who is holding plates on both hands .
4. eddy rody harry crong pororo and loopy are surrounding petty . petty smiles while holding plates on petty both hands .
5. poby is opening up pororo mouth and smiles .



1. rody is picking up the ball . rody is about to throw the ball .
2. eddy stops rody from throwing the ball . rody is listening to eddy .
3. eddy is explaining to rody how to throw the ball . rody is listening to it . rody is trying to understand how to throw the ball .
4. rody is holding the ball . rody is looking at the ball . rody is thinking how to throw the ball .
5. eddy is preparing to catch the ball . rody is holding the ball on the other side .



1. loopy decides to accept present .
2. loopy is angry about present . pororo is surprised .
3. pororo is making an excuse to loopy . there is a spring inside the box .
4. loopy is angry . loopy turns back and walks away . pororo is startled .
5. pororo looks sad . pororo and pororo friends are at playground .



Figure 2: Comparison of story continuation results between StoryDALL-E and our proposed Coln.

References

- [1] Tanmay Gupta, Dustin Schwenk, Ali Farhadi, Derek Hoiem, and Aniruddha Kembhavi. Imagine this! scripts to compositions to videos. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 598–613, 2018. [1](#)
- [2] Kyung-Min Kim, Min-Oh Heo, Seong-Ho Choi, and Byoung-Tak Zhang. Deepstory: Video story qa by deep embedded memory networks. *arXiv preprint arXiv:1707.00836*, 2017. [1](#)
- [3] Bowen Li. Word-level fine-grained story visualization. In *European Conference on Computer Vision*, pages 347–362. Springer, 2022. [1](#)
- [4] Adyasha Maharana, Darryl Hannan, and Mohit Bansal. Storydall-e: Adapting pretrained text-to-image transformers for story continuation. *arXiv preprint arXiv:2209.06192*, 2022. [1](#)