

Supplementary Material for Length-Controllable Image Captioning

I Setting of [EOS] decay factor γ

We perform cross validation on MS COCO Caption Karpathy’s val split to determine the value of γ . We set γ to different values for different levels in 4-level LaBERT. Specifically, γ is set to 0.88 and 0.95 for the second and third level of 4-level LaBERT, respectively. While for the first and the last level, we find the model performs well without [EOS] decay, *i.e.*, $\gamma = 1.0$.



		
Ori	A group of men on a field playing baseball.	VLP
Lv1	Some baseball players are playing a baseball game.	
Lv2	A group of men on a field playing baseball.	
Lv3	A baseball game in progress with the batter up to plate waiting for the pitch.	
Lv4	A baseball game in progress with the batter up to plate and the catcher ready to catch the ball.	
<hr/>		
Ori	A baseball player holding a bat next to home plate.	AoANet
Lv1	A baseball player standing next to home plate.	
Lv2	A baseball player holding a bat on a field.	
Lv3	A baseball player holding a bat on a field with a crowd of people watching.	
Lv4	Baseball game with a batter at home plate and a catcher and umpire behind him and a crowd watching.	
<hr/>		
Lv1	A baseball game in progress in a stadium.	LaBERT
Lv2	A baseball player holding a bat on a field.	
Lv3	A baseball player holding a bat on top of a field during a game.	
Lv4	A baseball game with a batter at home plate and a catcher behind him ready to catch the ball.	
<hr/>		
		
Ori	A group of people that are sitting under umbrellas.	VLP
Lv1	A group of people sitting under red umbrella.	
Lv2	A group of people that are sitting under umbrellas.	
Lv3	A group of people sitting under umbrellas next to a body of water.	
Lv4	A group of people sitting under umbrellas next to a body of water on a cloudy day.	
<hr/>		
Ori	A group of people sitting at tables under umbrellas.	AoANet
Lv1	A group of people sitting under red umbrellas.	
Lv2	A group of people sitting under umbrellas on a boat.	
Lv3	A group of people sitting under umbrellas on the side of a street under a tree.	
Lv4	A group of people sitting under umbrellas on the side of the road with a red umbrella over them.	
<hr/>		
Lv1	A group of people sitting on a boat.	LaBERT
Lv2	A group of people sitting at tables by the water.	
Lv3	A group of people sitting at a table on the side of a river.	
Lv4	A group of people sitting at the side of a boat with palm trees and buildings in the background.	

Fig. 1. Examples of Length-Controllable Image Captioning from test split. “Ori” denotes the original results of AoANet or VLP. “Lv n ” denotes results on level n .

II Examples of Length-Controllable Image Captioning

We show more examples of our 4-level AoANet, 4-level VLP, and 4-level LaBERT in Fig. 1-4.

III Examples of the Refine Process in LaBERT

We show examples of the refine process of our 4-level LaBERT in Fig. 5 and Fig. 6.

	
Ori	A white toilet sitting next to a sink in a bathroom. VLP
Lv1	A bathroom with a toilet sink and mirror.
Lv2	A white toilet sitting next to a bathroom sink.
Lv3	A bathroom with a white toilet sitting next to a sink under a mirror.
Lv4	A bathroom with a white toilet sitting next to a sink and a bathtub with a shower curtain.
Ori	A white toilet sitting next to a bath tub. AoANet
Lv1	A bathroom with a toilet sink and shower.
Lv2	A white toilet sitting next to a bath tub.
Lv3	A bathroom with a toilet a sink and a shower with a shower curtain.
Lv4	A bathroom with a white toilet sitting next to a white sink and a shower with a white curtain.
Lv1	A bathroom with white walls and brown floor. LaBERT
Lv2	A bathroom with a sink toilet and a shower.
Lv3	A bathroom with a white toilet a white sink and a white shower curtain.
Lv4	A bathroom with a brown tile and wood floor and a white bathtub with a white shower curtain.
	
Ori	A man riding on the back of a brown horse. VLP
Lv1	A man riding a horse down a street.
Lv2	A man riding on the back of a brown horse.
Lv3	A man riding on the back of a brown horse down a city street.
Lv4	A man in a green jacket riding a brown horse in front of a group of people on a sidewalk.
Ori	Two men and a woman are riding horses. AoANet
Lv1	A group of people sitting on some horses.
Lv2	A group of people riding on the backs of horses.
Lv3	A group of people riding on the back of horses next to a building
Lv4	A group of people sitting on horses in front of a building with a group of people standing around.
Lv1	A man that is sitting on a horse. LaBERT
Lv2	A man riding on the back of a brown horse.
Lv3	A man riding on the back of a brown horse next to other people.
Lv4	A man in a green coat and hat sitting on a horse with a woman standing next to him.
	
Ori	A red airplane sitting on top of an airport runway. VLP
Lv1	A person standing next to a small plane.
Lv2	A red and white plane sitting on top of a runway.
Lv3	A red and white plane sitting on top of an airport tarmac.
Lv4	A red and white plane sitting on top of a tarmac with a man standing next to it .
Ori	A red and white plane sitting on a runway. AoANet
Lv1	A small red plane on a run way.
Lv2	A red and white plane sitting on a runway.
Lv3	A red and white plane sitting on top of a runway next to a runway.
Lv4	A red and white airplane with a red tail on the tail and a person standing in front of it.
Lv1	A small red airplane sitting on a runway. LaBERT
Lv2	A small red and white airplane sitting on a runway.
Lv3	A red and white plane parked on a runway with a person beside it.
Lv4	A red and white plane sitting on top of a tarmac with a person standing next to it.
	
Ori	A group of birds sitting on top of a pond. VLP
Lv1	A bird that is standing in the water.
Lv2	A close up of a bird on a leafy plant.
Lv3	A close up of a bird on a leafy plant in a pond.
Lv4	A close up of a bird in a pond with leaves and water lilies in the background.
Ori	A bird is sitting on a branch of a tree. AoANet
Lv1	A bird that is sitting in a tree.
Lv2	A bird is perched on the branch of a tree.
Lv3	A bird is perched on the branch of a tree with leaves and leaves.
Lv4	A bird is perched on the branch of a tree with leaves and leaves on the side of it.
Lv1	A black bird is sitting in the water. LaBERT
Lv2	A black bird is swimming in a pool of water.
Lv3	A black bird sitting on a planter next to a body of water.
Lv4	A black bird with its wings spread feeding on a plant on the side of a body of water.

Fig. 2. Examples of Length-Controllable Image Captioning from **test** split. “Ori” denotes the original results of AoANet or VLP. “Lv n ” denotes results on level n .

		
Ori	A plane is parked behind a chain link fence.	VLP
Lv1	A plane that is sitting behind a fence.	
Lv2	A plane is parked behind a chain link fence.	
Lv3	A blue and white plane sitting on top of a field behind a fence.	
Lv4	A blue and white plane sitting on top of a field next to a chain link fence and a fence.	
Ori	A small plane sitting on top of an airport tarmac.	AoANet
Lv1	A small plane is parked behind a fence.	
Lv2	A small airplane sitting on top of a tarmac.	
Lv3	A small plane is parked on the fence and a fence is behind it.	
Lv4	a small airplane is parked on a fence and a fence with a chain link fence in the background.	
Lv1	A small airplane is parked behind a fence.	LaBERT
Lv2	A small airplane is parked behind a chain link fence.	
Lv3	A plane with propellor parked in front of a chain link fence.	
Lv4	A small blue propellor airplane sitting on top of a wooden floor behind a chain link fence.	
		
Ori	A woman sitting on a bench next to a dog.	VLP
Lv1	A woman and a dog on a bench.	
Lv2	A woman sitting on a bench next to a dog.	
Lv3	A woman sitting on a bench with a dog sitting on top of it.	
Lv4	A woman sitting on a bench with a dog sitting on it in front of a chain link fence.	
Ori	A woman sitting on a bench with a dog.	AoANet
Lv1	A woman is sitting on a wooden bench.	
Lv2	A woman sitting on a bench with a dog.	
Lv3	A woman sitting on a bench with a dog and a dog on a leash.	
Lv4	A woman sitting on a bench with a dog on her lap and a dog sitting on a bench.	
Lv1	A woman and a dog on a bench.	LaBERT
Lv2	A woman and a dog sitting on a wooden bench.	
Lv3	A woman sitting on a bench next to a dog sitting on a log.	
Lv4	A woman sitting on a bench with a dog on top of it sitting in front of a fence.	
		
Ori	A group of zebra standing next to each other.	VLP
Lv1	Three zebras are standing near water.	
Lv2	A group of zebra standing next to each other.	
Lv3	A herd of zebra standing next to each other on a field.	
Lv4	A herd of zebra standing on top of a lush green field next to a watering hole.	
Ori	A group of zebras standing next to a watering hole.	AoANet
Lv1	A group of zebras standing near a water.	
Lv2	A herd of zebra standing next to each other on a field.	
Lv3	A herd of zebra standing next to each other on top of a dirt field.	
Lv4	A herd of zebra standing next to each other on a dirt field next to a body of water.	
Lv1	A group of zebra standing near a lake.	LaBERT
Lv2	A group of zebras standing next to a watering hole.	
Lv3	A group of zebras standing next to a herd of antelope.	
Lv4	Three zebras standing next to a body of water with a couple of other animals in the background.	
		
Ori	A woman riding a wave on top of a surfboard.	VLP
Lv1	A woman surfing a wave in the ocean.	
Lv2	A woman riding a wave on top of a surfboard.	
Lv3	A woman riding a wave on top of a surfboard in the ocean.	
Lv4	A woman in a wetsuit riding a wave on top of a surfboard in the ocean.	
Ori	A person riding a surf board on a wave.	AoANet
Lv1	A person on a surfboard riding a wave.	
Lv2	A person riding a surf board on a wave.	
Lv3	A surfer in a wet suit riding a wave on top of a surfboard.	
Lv4	A man in a wet suit surfing on a wave in the ocean with a surfboard in the background.	
Lv1	A surfer riding the crest of the wave.	LaBERT
Lv2	A person on a surfboard riding a wave in the ocean.	
Lv3	woman riding a surfboard on top of a wave in the ocean.	
Lv4	A person in a wetsuit riding a surfboard on the top of a wave in the ocean.	

Fig. 3. Examples of Length-Controllable Image Captioning from **test** split. “Ori” denotes the original results of AoANet or VLP. “Lv n ” denotes results on level n .

	
Ori	A bedroom with a desk and a bed.
Lv1	A bedroom with a bed and a desk.
Lv2	A bedroom with a bed desk and a chair.
Lv3	A bedroom with a bed and a desk with a computer on top of it.
Lv4	A bedroom with a bed and a desk with a chair and a desk with a computer on it.
<hr/>	
Ori	A bedroom with a bed desk chair and computer.
Lv1	A bedroom with a bed and a desk.
Lv2	A bedroom with a bed a desk and a chair.
Lv3	A bedroom with a bed a desk a chair and a computer on it.
Lv4	A bedroom with a bed a desk and a desk with a computer on it and a chair.
<hr/>	
Lv1	A bedroom with a bed and a desk.
Lv2	A bedroom with a bed a desk and a chair.
Lv3	A bedroom with a bed a desk a chair and a window in it.
Lv4	A bedroom with a bed a computer desk a chair and a window and some pictures on the wall.
<hr/>	
	
Ori	A couple of men standing on top of a snow covered slope.
Lv1	Two people on skis in the snow.
Lv2	A couple of people standing on top of a snow covered slope.
Lv3	Two people standing on skis on top of a snow covered ski slope.
Lv4	Two people standing on skis on a snowy mountain top with a ski lift in the background.
<hr/>	
Ori	Two people standing on skis in the snow.
Lv1	Two people standing on skis in the snow.
Lv2	A couple of people standing on top of a snow covered slope.
Lv3	A couple of people standing on top of a snow covered ski slope.
Lv4	Two people standing on top of a snow covered ski slope with a ski lift in the background.
<hr/>	
Lv1	Two people are standing on a ski slope.
Lv2	Two people on skis standing on a snowy mountain.
Lv3	Two skiers standing on a snowy mountain in front of a ski lift.
Lv4	Two skiers standing on top of a snow covered ski slope with a ski lift in the background.
<hr/>	
	
Ori	A rusted out truck sitting in a field of flowers.
Lv1	An old rusty truck sitting in a field.
Lv2	A rusted out truck sitting in a field of flowers.
Lv3	A rusted out truck sitting in a field of wildflowers.
Lv4	A rusted out truck sitting in a field of wildflowers next to a bunch of trees.
<hr/>	
Ori	An old rusted out truck sitting in a field of flowers.
Lv1	An old rusty truck sitting in a field.
Lv2	An old rusty truck is sitting in a field.
Lv3	An old rusty truck with a rusted out of it sitting in a field of flowers.
Lv4	An old rusty truck with a rusted out of it sitting in the middle of a field of yellow flowers.
<hr/>	
Lv1	An old truck is parked in a field.
Lv2	An old truck sitting in a field of flowers.
Lv3	An old truck sitting in a field of flowers next to a forest area.
Lv4	An old truck parked in a field of weeds with a lot of flowers and trees in the background.
<hr/>	
	
Ori	Three people sitting on a couch holding game controllers.
Lv1	Three people are playing a video game together.
Lv2	Three people sitting on a couch holding wii remotes.
Lv3	Three people sitting on a couch holding nintendo wii remotes.
Lv4	Three people sitting on a couch one holding a nintendo wii controller and the other holding a nintendo wii.
<hr/>	
Ori	A group of young men playing a game with nintendo wii controllers.
Lv1	A group of people playing a video game.
Lv2	A group of young men playing a game with nintendo wii controllers.
Lv3	A group of young men standing next to each other holding nintendo wii controllers.
Lv4	A group of young men standing next to each other in front of a couch holding a nintendo wii game
<hr/>	
Lv1	Three young men playing a video game together.
Lv2	Three young men playing a game with remote controllers.
Lv3	Three men sitting in a living room with two them holding wii remotes.
Lv4	Three young men sitting in a room with a yellow wall and two of them holding wii remotes.
<hr/>	

Fig. 4. Examples of Length-Controllable Image Captioning from **test** split. “Ori” denotes the original results of AoANet or VLP. “Lv n ” denotes results on level n .





	Level 1 refine process (total steps 10):
	Step 2 A signs signs signs attached on brick wall.
	Step 4 A red sign attached on a brick wall.
	Step 6 Two red signs signs on a brick wall.
	Step 8 Two street signs hanging on a brick wall.
	Step 10 Two street signs mounted on a brick wall.
	Level 2 refine process (total steps 15):
	Step 3 A plate plate food with bread bread bread bread a.
	Step 6 A plate of food with bread on a a a.
	Step 9 A plate of food with bread bread and bread it.
	Step 12 A plate of food with bread and eggs on it.
	Step 15 A plate of food with bread and eggs on it.
	Level 3 refine process (total steps 20):
	Step 4 A elephants of elephants a a a a a in the background.
	Step 8 A group of elephants standing in a a of a a in the background.
	Step 12 A group of elephants walking in a water with with trees in the background.
	Step 16 A group of elephants walking in a shallow hole with trees in the background.
	Step 20 A group of elephants walking through a watering hole with trees in the background.
	Level 4 refine process (total steps 25):
	Step 5 A man standing standing a a a a a a a in the the.
	Step 10 A man standing standing the side of a a a a a cow a the mountains in the background.
	Step 15 A man standing a a cow a a a cow of a a road with mountains in the background.
	Step 20 A man standing a a shirt next to a cow on a green field with mountains in the background.
	Step 25 A man standing on a sidewalk next to a cow on a green field with mountains in the background.

Fig. 5. Examples of the Refine Process from COCO test split.

	Level 1 refine process (total steps 10):	
	Step 2	A man riding a motorcycle on a road.
	Step 4	A man riding a motorcycle on a road.
	Step 6	A person riding a motorcycle down a road.
	Step 8	A person riding a motorcycle on a road.
	Step 10	A person riding a motorcycle on a road.
	Level 2 refine process (total steps 15):	
	Step 3	A train train covered covered on on on on tracks.
	Step 6	A train with covered on on on on tracks.
	Step 9	A train with covered in is sitting on the tracks.
	Step 12	A train with graffiti on is sitting on the tracks.
	Step 15	A train with colorful graffiti is sitting on the tracks.
	Level 3 refine process (total steps 20):	
	Step 4	A wooden room with with with a a and vase vase vase in flowers.
	Step 8	A wooden table with a large of and and vase flowers flowers the it.
	Step 12	A wooden table with a vase of flowers sitting it bottles of of windows.
	Step 16	A wooden table with a vase of flowers on it next to large windows.
	Step 20	A wooden table with a vase of flowers on it next to large windows.
	Level 4 refine process (total steps 25):	
	Step 5	A polar bear in in in water water water ball ball ball ball ball ball ball the the.
	Step 10	A polar bear in in the water water water ball ball ball ball ball ball ball in the water.
	Step 15	A polar bear is in the paw paw paw paw paw paw with a a ball in the water.
	Step 20	A polar bear laying on its back legs its paw and playing with a blue ball in the water.
	Step 25	A polar bear laying on its back on one paw and playing with a blue ball in the water.

Fig. 6. Examples of the Refine Process from COCO `test` split.