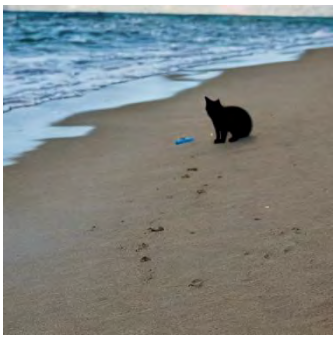
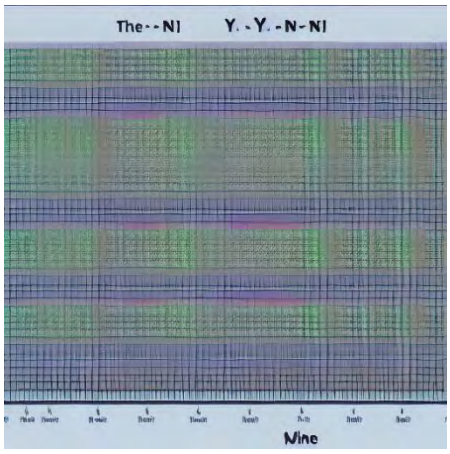
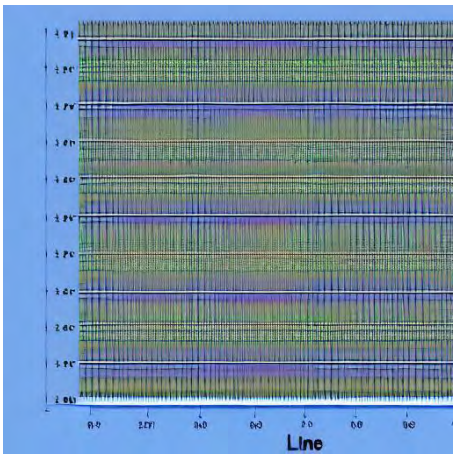
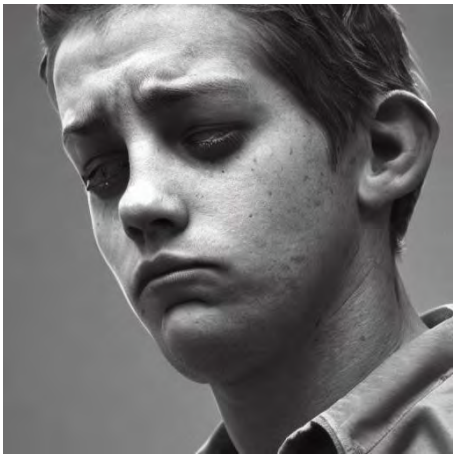
	Origin picture	Change stucture	Change word
	A cat on the beach	a beach with a cat	A cat on the lawn
No Seed			
Seed			
	a photograph of an astronaut riding a horse	An astronaut on horseback is featured in a photograph	a cartoon of an astronaut riding a horse
No Seed			
Seed			

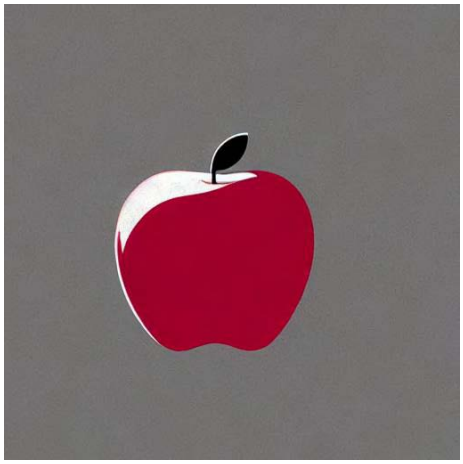
	A girl with short hair in a long purple dress	A girl with short hair wears a long purple dress	A girl with short hair in a long blue dress
No Seed			
Seed			

	a photograph of A man flying a kite	A man engaged in kite-flying in a photograph	a photograph of A cat flying a kite
No Seed			
Seed			

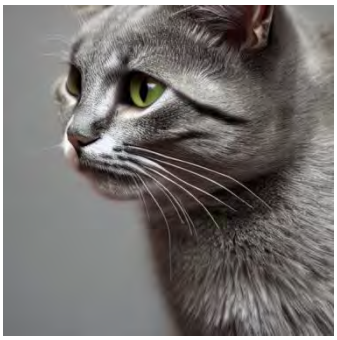
When the initial noise is the same, the structure picture has a higher similarity, and the background or the main body or the layout are more similar.
However, when the noise is different, the change is very obvious, only in line with the prompt description.
When the initial noise was the same, the similarity between the two could still be seen by changing the prompt. If the background is changed, the overall image of the subject is

similar, if the image type is changed, such as from pictures to cartoons, both the subject and the background are similar to only changing the material or adding filters. There was confusion when modifying the subject, perhaps because there were fewer or only one sample of kite flying during training.		
	A house is on the left of a river	A house is on the right of a river
Position change		
	The graph of the line y=1	The graph of the line y=x+1
math		
	this is a car?	this is a car!
Tone change/ Punctuation marks change		

	A big dog and a big mouse	A small dog and a big mouse
size change		
	A man sat outside the boat	A man sat inside the boat
In out change		
	a happy boy	a sad boy
Mood change		

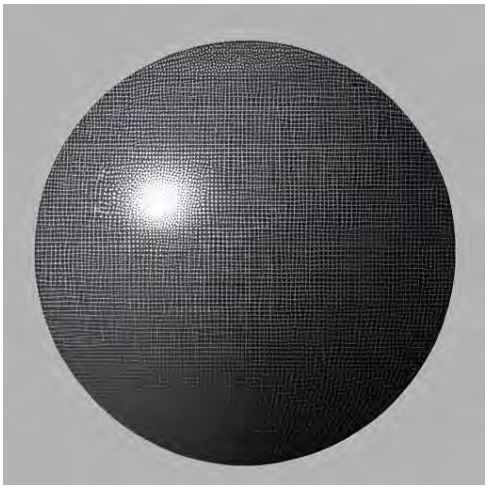
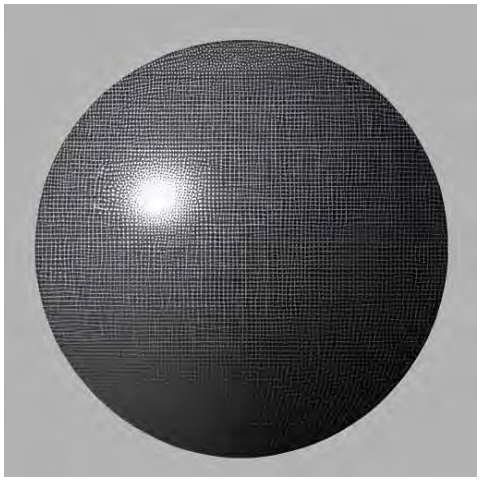
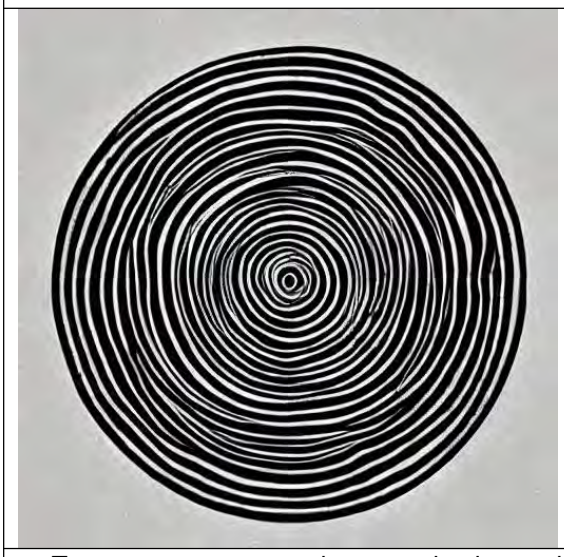
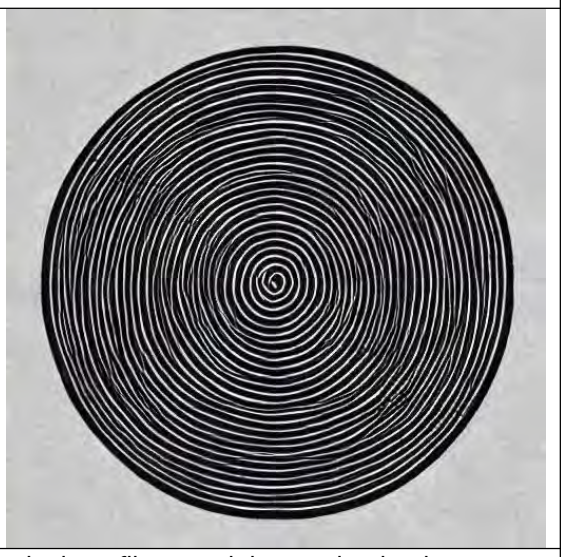
	An open book	An closed book
State change		
	one apple	two apple
Number change		

When the image design to position, mathematics, quantity, state, tone, the effect is not ideal, for quantity and number problems, should not be a simple convolution operation can be solved. Information about location and status may be learned with a large sample size. In addition, it should be impossible to learn the relevant information of tone in the training of the model, because it is difficult to represent a picture with emotional colors through punctuation marks or word order. However, when it comes to special emotional words such as sad, the model can achieve the effect, probably because the combination of emotion and face has specific forms of expression in reality. As for "a sad car" and the like that can't be trained.

A cat	A sad cat	A crying cat
		
A car	A sad cat	A crying car
		
A man	A sad man	A crying man
		
happy	sad	cry
		

When a subject is described by an emotional word, divided by the person, there is no significant change in the rest of the subject, whether it is an object capable of emotional expression or not. And the model almost equates cry and sad. When only emotional words are used as input, the generation is very confusing.

Internally contradictory sentences	
A curved and straight road	A straight and curved road
	
A sunny night	A night sunny
	
A big small ball	A big small ball

	
A circle square	A square circle
	
Try to enter a prompt that contains internal logical conflicts, and the production is more chaotic, but it is not difficult to see that in a pair of conflicting words, one word plays a decisive role, and is not affected by positional factors, in such a prompt, the structure change does not cause a large change.	

A photograph of a lion walking in the jungle	A photograph of a leopard walking in the jungle	A photograph of a tiger walking in the jungle

		
A poster of a superman standing in the middle of a cyber city	A poster of a teddybear standing in the middle of a cyber city	A poster of a kid standing in the middle of a cyber city
		
A photograph of a smiling face of an old man wearing glasses	A photograph of a smiling face of a man wearing glasses	A photograph of a smiling face of a young man wearing glasses
		

The quality of this set of pictures is very good, which should be prompt selected better, although the sentences are much longer on average than my previous prompt, and contain a lot of modifier words, perhaps because the long sentences guide the pictures better. The part of the picture that was not given when generating short sentences did not have clear guidance, and the content of the picture was confused because the iteration did not move in a reasonable direction. Then the correlation between words and words in diffusion might need to be considered.