Step-Video Prompt Guildlines

English Version

"Prompt", as the main interactive language for generating videos, will directly affect the video content generated by the model. Therefore, how to write effective prompt words for AI video creation is a skill that every creator hopes to master. As a new breakthrough in AI video generation model, "Step-Video" is still developing and improving, and continuous exploration and innovation are the key to unleashing its maximum potential. In order to help everyone better master video generation technology, we have compiled the prompt word formula for reference.

Prompt = magic word + subject (subject description) + action + environment (environment description) + camera movement + video type + video style

- Subject: The main object in the video, which is the core element of the theme of the picture, may be a person, animal, plant or object, etc.
- Subject description: A detailed description of the subject's appearance details, body posture, etc., which can be listed one by one in short sentences. For example, movement state, hairstyle, clothing, facial features, body movements, etc.
- Subject salience: The visual priority of the subject in the video, which can be sorted according to its importance in the picture. For example, if a man is standing next to a stop sign and eating an apple, then "man" as the subject should be ranked in front.
- Action: Describe the subject's movement state, including static or moving state. The description of the movement should not be too complicated, and the picture that can be presented within 5 seconds should be considered. The smoothness of the movement will also change depending on the details of the action description.
- Environment: The environment where the subject is located, including foreground and background elements.
- Environment description: The detailed description of the subject's environment can be listed in short sentences, but it should not be too much, and it can be consistent with the picture content displayed in the 5-second video. For example, indoor, outdoor or natural environment, etc.

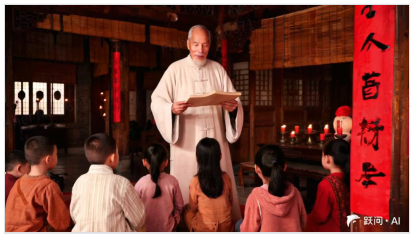
- Camera movement: refers to the shooting techniques of the lens, including fixed lens, mobile shooting, static shooting, etc. The use of different lenses and the connection and switching between lenses can effectively convey the storyline or information.
- Video type: refers to the specific form or category of the video, such as documentary, movie, animation, variety show, interview, concert, etc.
- Video style: refers to the visual presentation of the video, such as realistic, abstract, cartoon, freehand, etc., to reflect the unique artistry and emotional expression of the work.

The core elements of the above formula include magic words, subject, action, environment and camera movement, which constitute the most concise and basic units for describing video images. When we need to describe the subject and environment in more detail, we only need to list multiple descriptive sentences to ensure that all necessary elements are included in the prompt. Step-Video will automatically expand the prompt words according to these expressions and generate a video that meets the expectations.

For example, the original description is: "An old man stands in front of the family tree and tells the children the family story." We can add a detailed description "During the Spring Festival, the candlelight flickered in the old ancestral hall. An old man in a robe stood in front of the family tree, pointing to the family tree and telling the family story to the children sitting in a circle. The children listened attentively with relish, and the sound of firecrackers came faintly outside the ancestral hall." Add the magic words: "UHD, HDR, ambient light, Dolby Atmos, stable picture, smooth action, realistic details, professional-level composition, natural, vivid, ultra-detailed, clear." The quality of the generated video will be further improved, and it is possible to get results beyond expectations.

By adding accurate descriptions of actions and background atmosphere, the generated content can be enriched and aligned.

prompt	An old man stands in front of the family tree and tells family stories to the children.	During the Spring Festival, candlelight flickered in the old ancestral hall. An old man in a long robe stood in front of the family tree, pointing at it and telling the family story to the children sitting around. The children listened attentively and with great interest. The	During the Spring Festival, candlelight flickered in the old ancestral hall. An old man in a robe stood in front of the family tree, pointing at it and telling the children sitting around the family story. The children listened attentively, and the sound of firecrackers

		sound of firecrackers could be heard faintly outside the ancestral hall.	could be heard faintly outside the ancestral hall. Ultra-high definition, HDR, ambient light, Dolby Atmos, stable picture, smooth movement, realistic details, professional-level composition, natural, vivid, ultra-detailed, clear.
视频			
prompt	A couple sitting on a bench in the park.	A couple was sitting on a bench in the park. The boy took out a bunch of red roses from his back and handed it to the girl. The girl took it with surprise, and then the two of them smiled happily and moved closer to each other.	A couple is sitting on a bench in the park. The boy takes out a bouquet of red roses from his back and hands it to the girl. The girl takes it with surprise, and then the two of them smile happily and get close to each other. Ultra HD, HDR, ambient light, Dolby Atmos, stable picture, smooth motion, realistic details, professional-level composition, natural, vivid, ultra-detailed, clear.
视频			
prompt	A little boy in a suit transforms into Venom	In the video, a little boy in a suit suddenly has a ferocious expression, and his body is gradually covered by black liquid, eventually transforming into Black Venom. This process takes place in a dark environment	In the video, a little boy in a suit suddenly has a ferocious expression, his body is gradually covered by black liquid, and finally transforms into Black Venom. This process takes place in a dark environment with a blurred

		with a blurred background, highlighting every detail of the boy's transformation. The video is shot in close-up, with a sci-fi style, clearly showing every detail of the transformation, giving people a sense of shock.	background, highlighting every detail of the boy's transformation. The video is shot in close-up, with a sci-fi style, clearly showing every detail of the transformation, giving people a sense of shock. Ultra-high definition, HDR, ambient light, Dolby Atmos, stable picture, smooth movement, realistic details, professional-level composition, natural, vivid, ultra-detailed, clear.
视频			

By adding descriptions of camera movements and camera angles, the overall presentation can be made more visually impactful.

prompt	A skier is skiing on a snowy mountain. The skier is wearing a dark ski suit, goggles and a light-colored helmet. He slides down the hillside, and a lot of snow is kicked up behind him. The background is a clear blue sky and a snowy mountain. The sun is strong, forming a starburst effect in the picture. The whole video is full of movement, showing the excitement and fun of skiing.	A skier skis on a snowy mountain, captured from a first-person perspective using a selfie stick. The skier is wearing a dark ski suit, goggles, and a light-colored helmet, and holding a selfie stick. He slides down the slope, with a lot of snow behind him. In the background is a clear blue sky and a snowy mountain. The sun is strong, creating a starburst effect in the frame. The skier has an excited facial expression, his mouth is open, and he waves to the camera with one hand free. The whole video is full of movement, showing the excitement and fun of skiing.
视频		

prompt	A woman in a light-colored gauze dress and an elaborate headdress is smiling, and then the camera switches to a man wearing a transparent veil and staring ahead.	A woman in light-colored gauze, wearing an exquisite headdress, smiling. The camera closes up to show her beautiful face and gentle temperament. Then, the camera switches to a man wearing a transparent veil, staring forward. A horizontal close-up shows the texture of the veil and the man's thoughtful expression. The scene switches again, and the man wears a black robe with dark patterns on the lapels and cuffs, standing in a courtyard-like environment with a serious expression. The medium shot shows the whole body of the character and the surrounding environment, creating a mysterious atmosphere.
视频		

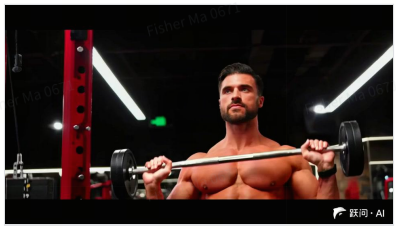
Increasing the description of the range of motion can make the overall presentation more realistic and the range of motion larger.

prompt	The video shows a yellow racing car driving on a high-speed track floating in the clouds. The track winds and	The video shows a yellow racing car speeding on a high-speed track floating in the clouds. The track winds and

	stretches above the boundless white sea of clouds, with the edge clearly marked by red and white warning tape. The camera follows the car from multiple angles.	stretches above the boundless white sea of clouds, and the edges are clearly marked with red and white warning tape, adding visual tension. The streamlined body of the racing car reflects a metallic luster in the sun, and the tires stir up cloud-like ripples. With every turn, the track seems to leave a faint light mark in the air. The camera uses a follow-up shooting method, following the movement of the racing car from multiple angles, accurately capturing the details of its rapid turning and straight-line acceleration. The overall picture has both a sense of dynamics and a surreal dream effect.
视频		

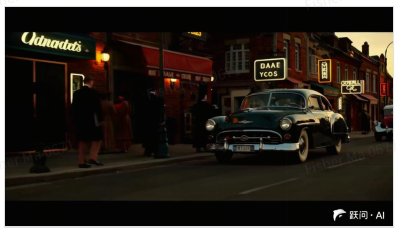
Excellent Cases

prompt	In the video, a kid in a suit takes off his coat in front of the camera to show off his muscular body. The kid's expression is focused and proud, and his muscle lines are clearly visible under the light. The background is a simple white wall, which forms a sharp contrast with the kid's muscles. The whole picture is shot with a fixed lens, which clearly shows every detail of the kid's	Ultra HD, HDR, ambient light, Dolby Atmos, stable image, smooth motion, realistic details, professional-grade composition, natural, vivid, ultra-detailed, clear. A man in professional running gear runs fast on a spacious stadium track. The camera follows his pace, clearly showing his tight leg muscles and smooth running movements. The background is neatly arranged spectator	Ultra-high definition, HDR, ambient light, Dolby Atmos, stable image, smooth movements, realistic details, professional-grade composition, natural, vivid, ultra-detailed, clear. A muscular fitness enthusiast concentrates on lifting dumbbells in a brightly lit gym. He is wearing tight sportswear that highlights the lines of every muscle. During the movement, the

	movements, with a realistic style.	seats and the finish line in the distance, creating a tense and energetic atmosphere.	dumbbells rise and fall smoothly between his arms, and each lift is accompanied by the tightening and stretching of the muscles. The lens uses close-up shots, focusing on the interaction between the dumbbells and muscles, creating a visual impact of strength and beauty. The overall style is realistic and full of power.
视频			
prompt	Ultra HD, HDR, ambient light, Dolby Atmos, stable image, smooth motion, realistic details, professional-grade composition, natural, vivid, ultra-detailed, clear. A skier in bright ski clothes glides at an amazing speed on the snowy slopes. His skis rub against the snow to create bright sparks, and his movements are smooth and elegant. The camera uses a push lens, zooming in from a distance to a close-up, capturing the skier's focused expression and the moment of rapid sliding, creating a strong sense of speed and sports beauty.	In the video, a panda is wearing a suit, a white shirt, black suit pants, a red tie, and a pair of black-framed glasses, with a serious expression. It holds a document with its front paw and supports the table with its other front paw, as if it is attending an important meeting. Behind it is a white wall with a world map hanging on it. The entire scene is shot with a fixed lens, and the picture is clear and realistic.	Ultra-high definition, HDR, ambient light, Dolby Atmos, stable picture, smooth motion, realistic details, professional-level composition, natural, vivid, ultra-detailed, clear. The camera slowly pushes from the sea level to the distance, showing a vast blue ocean, sparkling under the sunlight. In the distance, where the sea and the sky meet, a few white sails are looming. The sea breeze blows gently, and the waves gently hit the rocks on the coast, making a pleasant sound. A few seagulls are circling at low altitude, blending with the blue sky and the blue ocean, creating a peaceful and magnificent picture.

视频			
prompt	In the video, a little boy in a suit suddenly transforms into a magical girl. The little boy is wearing a black suit, a red tie, and a focused expression. He stretches out his hands, closes his eyes, and instantly transforms into a magical girl in a pink skirt and a crown. The girl opens her arms, emitting a dazzling light, with a mysterious starry sky in the background.	Ultra-high definition, HDR, ambient light, Dolby Atmos, stable picture, smooth motion, realistic details, professional-grade composition, natural, vivid, ultra-detailed, clear. Deep in a lush forest, a modern research laboratory stands quietly. Inside, the experimenters, dressed in neat white coats, are concentrating on operating various high-tech equipment, which are meticulously analyzing the surrounding natural resources. Outside the window, the lush natural landscape is in full view, as if silently telling the secrets of nature. On the large screen in the laboratory, the data changes about the forest are jumping in real time, clearly showing the beautiful picture of the harmonious coexistence of nature and technology. The lens adopts fixed shooting, showing the rigorous and orderly working atmosphere in the laboratory and the picturesque natural beauty outside the window.	Ultra-high definition, HDR, ambient light, Dolby Atmos, stable picture, smooth movement, realistic details, professional-grade composition, natural, vivid, ultra-detailed, clear. A doctor in a white coat sits at the table in the consulting room, concentrating on reviewing the patient's medical records. He occasionally smiles, comforts the patient with a soft expression and nod. The doctor patiently explains the treatment plan to the patient in a gentle tone, and every detail is clear. The patient sits opposite with peace of mind, listening carefully, and his expression reveals trust and gratitude for the doctor. The whole scene is quiet and professional, and the camera uses a combination of close-ups and close-ups to show the warm and trusting interaction between doctors and patients.
视频			

prompt	Ultra HD, HDR, Ambient Light, Dolby Atmos, Picture Stabilization, Smooth Motion, Realistic Details, Professional Composition, Natural, Vivid, Ultra Detail, Sharp. A lush green leaf blocks most of the frame, revealing a girl admiring flowers in a sun-dappled garden. She is wearing a pale pink dress and white sneakers, and her long hair moves gently in the breeze. The shot is shot from a low angle, creating a dreamy, tranquil atmosphere.	Ultra HD, HDR, Ambient Light, Dolby Atmos, Picture Stabilization, Smooth Motion, Realistic Details, Professional Composition, Natural, Vivid, Ultra Detail, Clear. In a bright studio, an artist is concentrating on painting colorful colors on the canvas with a brush. He is wearing a dark painting coat with paint on his cuffs. Occasionally, he stops his work, steps back, stares at his work, frowns slightly, and falls into deep thought, thinking about his next painting plan. The whole picture is shrouded in soft light, and the lens uses a fixed lens to create a focused and tranquil artistic atmosphere.	Ultra-high definition, HDR, ambient light, Dolby Atmos, stable images, smooth movements, realistic details, professional-grade composition, natural, vivid, ultra-detailed, clear. In the vast farmland, a high-tech automated agricultural machine is busy sowing and harvesting. Its mechanical arm is precise and powerful, and every operation seems so smooth and natural. At the same time, a drone is hovering in the blue sky, its sharp camera is always paying attention to the growth of crops and transmitting real-time data back to the control center. The camera uses a tracking method, slowly rising from the ground to the air, showing the spectacular scene in the process of agricultural modernization.
视频			
prompt	Ultra-high definition, HDR, ambient light, Dolby Atmos, stable picture, smooth motion, realistic details,	Ultra-high definition, HDR, ambient light, Dolby Atmos, stable images, smooth movements, realistic details,	Ultra HD, HDR, ambient light, Dolby Atmos, stable image, smooth motion, realistic details, professional-grade

	professional-grade composition, natural, vivid, ultra-detailed, clear. A disheveled, unshaven man sat alone in a messy room. The surroundings were filled with debris and unwashed dishes, and there was a faint smell of mold in the air. On the table in front of the man, there were a few crumpled newspapers and a pile of empty beverage cans. He was eating a bucket of instant noodles, and the steaming soup spilled on the greasy table, blending in with the messy scene around him. The whole scene was dim and depressing, as if time had stopped here.	professional-grade composition, natural, vivid, ultra-detailed, and clear. In this future city full of high-tech elements, the towering skyscrapers are made of shining metal and transparent glass, and the streamlined design highlights its futuristic sense. In the air, silent flying cars shuttle through the boundless sky, and on the ground, pedestrians wear simple and smooth clothes, walking like messengers of the future. On both sides of the street, the digital images on the advertising screens are ever-changing, adding a touch of technology to the city. The picture has a cold tone, and the metallic luster shines in the light, creating an atmosphere that is both modern and mysterious.	composition, natural, vivid, ultra-detailed, clear. A classic 50s sedan drives slowly in the afterglow of the sunset, with its body exuding a retro luster. On both sides of the streets of the old town, retro neon signs flicker slightly in the orange halo, contrasting with the brick buildings. Pedestrians on the street are dressed in sophisticated retro clothes and stroll leisurely. In the background, the slightly flickering neon lights add a touch of fantasy to the whole scene. The camera uses panning to shoot, telling this nostalgic picture.
视频			

Some practical tips:

- Try to choose concise words and sentence structures, and avoid using overly complex language.
- Keep the content of the picture as simple as possible, and ensure that it can be fully presented within 5 to 10 seconds.
- Use words such as "Eastern artistic conception", "China", and "Asia" to more easily generate pictures with Chinese style or Chinese characters.

- The current video model does not handle numbers accurately enough. For example, "10 puppies on the beach" is difficult to keep the number consistent.
- For split-screen scenes, you can use the prompt word: "4 cameras, spring, summer, autumn, and winter" to help generate different picture distributions.
- At this stage, it is difficult for the model to generate complex physical movements, such as ball bouncing, high-altitude throwing, etc.
- Change special effects can also be generated, and the details of the change process need to be improved, such as suddenly becoming a hideous expression, the body is gradually wrapped in black liquid, and finally transformed into black venom.