











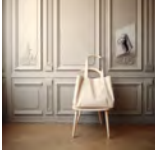

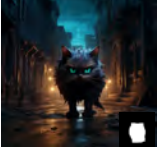
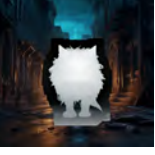




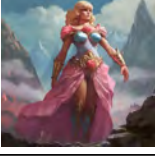
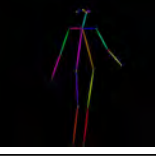



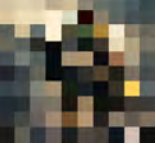
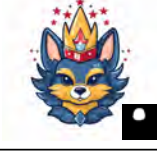
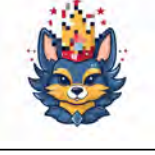






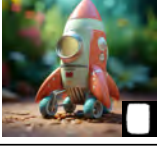



Image Segmentation	<p>Input Image</p> 	<p>Instruction</p> <p>Conduct a segmentation task on (image).</p>	<p>Output Image</p> 	<p>Input Image</p> 	<p>Instruction</p> <p>Set off with the segmentation of (image).</p>	<p>Output Image</p> 
	<p>Input Image</p> 	<p>Instruction</p> <p>Do segmentation on mask of (image).</p>	<p>Output Image</p> 	<p>Input Image</p> 	<p>Instruction</p> <p>Execute segmentation on mask in (image).</p>	<p>Output Image</p> 
	<p>Input Image</p> 	<p>Instruction</p> <p>Produce a depth map for (image) please.</p>	<p>Output Image</p> 	<p>Input Image</p> 	<p>Instruction</p> <p>I need a depth map created from (image).</p>	<p>Output Image</p> 
	<p>Input Image</p> 	<p>Instruction</p> <p>Can you decipher the depth map from the portion of (image) highlighted by mask?</p>	<p>Output Image</p> 	<p>Input Image</p> 	<p>Instruction</p> <p>Generate the depthed depiction of the specified area in (image) as indicated by mask.</p>	<p>Output Image</p> 
Human-pose Estimation	<p>Input Image</p> 	<p>Instruction</p> <p>Can you help distinguish the poses of the figures in (image)?</p>	<p>Output Image</p> 	<p>Input Image</p> 	<p>Instruction</p> <p>Can you provide insight into the pose of the person presented in (image)?</p>	<p>Output Image</p> 
	<p>Input Image</p> 	<p>Instruction</p> <p>Adapt (image) into a mosaic representation.</p>	<p>Output Image</p> 	<p>Input Image</p> 	<p>Instruction</p> <p>Can we get a mosaic version of (image)?</p>	<p>Output Image</p> 
Image Mosaic	<p>Input Image</p> 	<p>Instruction</p> <p>Manipulate (image) to include mosaics in the sections specified by the mask.</p>	<p>Output Image</p> 	<p>Input Image</p> 	<p>Instruction</p> <p>Incorporate a mosaic into the (image), specifically in the mask regions.</p>	<p>Output Image</p> 
	<p>Input Image</p> 	<p>Instruction</p> <p>Can you transform (image) into a black and white one?</p>	<p>Output Image</p> 	<p>Input Image</p> 	<p>Instruction</p> <p>Could you adjust (image) to be black and white?</p>	<p>Output Image</p> 
Image Grayscale	<p>Input Image</p> 	<p>Instruction</p> <p>Would you mind converting the mask section of the (image) into a monochrome hue while keeping the rest as it is?</p>	<p>Output Image</p> 	<p>Input Image</p> 	<p>Instruction</p> <p>Could you please transform the mask area of the (image) into grayscale, whilst leaving the rest as per original?</p>	<p>Output Image</p> 

Figure 1: The ACE's generated visualization of image segmentation, depth estimation, human-pose estimation, image mosaic, and image grayscale in low-level visual analysis.

	Input Image	Instruction	Output Image	Input Image	Instruction	Output Image
Image Degradation		Undertake a reduction procedure on the quality of [image].			Implement a degradation process on [image] to decrease its quality.	
		The mask region in [image] needs to be degraded.			Aim the image degradation process at the mask area of the [image].	
		Get the edge-enhanced result for [image].			Conduct an edge detection operation on [image].	
		I need the result of edge detection for the mask region in [image].			Generate the edge detection map from the mask section of the [image].	
Contour Extraction		Would you be able to make a contour picture from [image] for me?			I'd like you to develop a contour image based on [image].	
		I need your assistance in creating a contour of the designated area in the [image], based on the given mask.			Can we generate a contour dependent on the area highlighted by mask in the prescribed [image]?	
		Generate a scribble of [image], please.			Please illustrate [image] in a scribble style.	
Scribble Extraction		Please derive a local scribble of the [image] referencing the given mask.			Could you assist me in creating a scribble using the mask from the specified image [image]?	

Figure 2: The ACE's generated visualization of image degradation, edge extraction, contour extraction, and scribble extraction in low-level visual analysis.

	Input Image	Instruction	Output Image	Input Image	Instruction	Output Image
Segmentation-based Generation		Based on the primitives from [image], synthesize a real image that fits the context and details provided in "a charming girl, long black straight hair, sky background, in yumeji style, realistic portrait".			"rugged man wearing aviator sunglasses on deep red background". Transforming [image]	
		Following the segmentation outcome in mask of [image], develop a real-life image using the explanatory note in "water-colour, dreamy, stuido ghibli, art nouveau".			"Character design sheet of a asian woman, round face, round spectacles, messy cute hair, freckles, small nose, white background". Develop a detailed image from the image segmentation represented by mask of [image].	
Depth-based Generation		Could you use the depth map [image] and the text caption "a young elf-like character with pointed ears, green eyes, and blonde hair, adorned in a red and brown outfit with green accents. A bokeh effect with green hues and hints of light reflecting through foliage" to create a corresponding graphic [image]?			Create character reference illustrations of Oliver, an adventurous 8-year-old, has tousled brown, wind-blown, outdoor lifestyle, hazel eyes, freckles. Oliver wears, worn green adventurer, vest". Could you bring to life an image using the depth map [image]?	
		A cartoon dinosaur, likely a Tyrannosaurus Rex, dressed in a white astronaut suit, set against a backdrop of outer space. The dinosaur is colored in vibrant shades of orange, with a lighter shade on its underbelly. Please restore the region of the photo highlighted by mask using the information from the depth map [image].			"A close-up illustration of a penguin wearing large, orange sunglasses. The sunglasses have a thick, orange frame and dark blue lenses that reflect the penguin's surroundings.". Utilizing the depth map [image], I'd like you to reestablish the local zones as indicated in mask.	
Pose-based Generation		Could you please help translate this posture schema [image] into a colored image based on the context I provided "dwarf character dungeons & dragons with blue eyes, long barb and mustache, total body"?			I'm hoping to turn this pose guide [image] into a full color image with your help, using my description provided in the "a cartoon-style eagle dressed in a black leather jacket, round sunglasses, a feathered chest piece, and accessorized with a necklace and earrings".	
Mosaic-based Generation		Transform and generate an image using mosaic [image] and "Osamu Tezuka middle aged astro man" description			A woman representing the herb lavender, photorealistic, beautiful, detailed. Could you bring back the integrity of this mosaic art [image] by transforming it into its original photograph?	
		A whimsical, steampunk-inspired goldfish, a fascinating fusion of transparent glass and gleaming metal, reveals an intricate network of gears, pipes, and wires, hinting at a complex internal mechanism. Would you be able to eliminate the mosaic in areas indicated by mask on [image] to uncover the underlying image part.			Several wooden balconies adorn the exterior of the treehouse, each lined with potted plants and small windows.". Recover the image [image] by taking away the mosaic on the mask area.	
Grayscale-based Generation		Can you make this [image] colorful as per the "retrofuturism exploration on mars"?			"Futuristic robotic astronaut floating, dynamic posture, with a sense of design, cartoon style, cartoon astronaut, cartoon proportions, anime aesthetics, with ultra fine textures, with ultra fine textures. 3D, black background". Please transform this grayscale [image] into full color.	
		Make the mask part of my [image] to be color processed based on the notes "a whimsical, steampunk-inspired airship, seemingly cobbled together from various salvaged parts. It's a vibrant orange and white, with exposed pipes, gears, and wires adding to its ramshackle charm".			The chimpanzee is wearing a yellow astronaut helmet, and its head and shoulders are visible. The chimpanzee's fur is black and its eyes are brown. Its mouth is slightly open, and its teeth are visible. Kindly apply color to the mask portion of the [image], while the rest of it stays in grayscale.	

Figure 3: The ACE’s generated visualization of segmentation-based, depth-based, pose-based, mosaic-based, and grayscale-based generation in controllable generation.






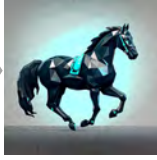
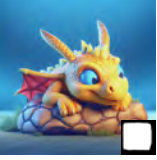
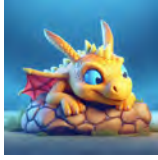

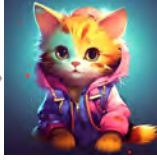

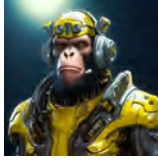


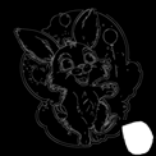
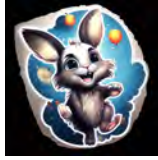




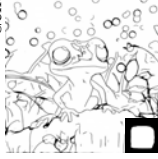






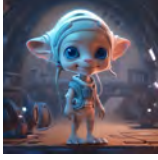
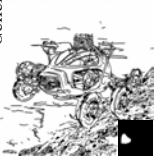




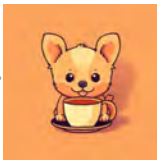

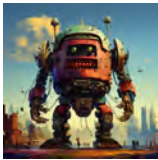
Degradation-based Generation	<p>Input Image</p> 	<p>Instruction</p> <p>"Handsome boy, age 20 years old, chest and gluteal muscles developed, six pack abs, white sports shorts, smile, short hair, less chest hair, white sports shoes, sitting on the grass". Eliminate noise interference in (image) and maximize the crispness to obtain superior high-definition quality.</p>	<p>Output Image</p> 	<p>Input Image</p> 	<p>Instruction</p> <p>Adhere to "Ultra clear, high resolution, high detail, Chinese style, Ancient Chinese, face view, long black hair, brown eyes, frontal, horizontal Angle, facing camera, outdoor, street, serious, upper body" to clean the noise from (image) and develop a clearer visual.</p>	<p>Output Image</p> 
	<p>Input Image</p> 	<p>Instruction</p> <p>The horse stands against a gradient background that ranges from light to dark, emphasizing the contrast of the horse's colors. The background could be interpreted as either the light of day or the darkness of night, depending on the viewpoint of the observer". Denoise the mask segment in the (image) to improve clarity.</p>	<p>Output Image</p> 	<p>Input Image</p> 	<p>Instruction</p> <p>The dragon is yellow with orange accents, featuring spikes along its back and spikes on each cheek. Its eyes are large and bright blue, adding to its expression of contentment. Escalate high resolution refinement on the mask area of the (image).</p>	<p>Output Image</p> 
Edge-based Generation	<p>Input Image</p> 	<p>Instruction</p> <p>Take the edge conscious (image) and the written guideline "cat wearing colorful shirt, brown eyes, bright brown eyes, chibi, detailed fur", high resolution, 4k, soft fur" and produce a realistic image.</p>	<p>Output Image</p> 	<p>Input Image</p> 	<p>Instruction</p> <p>"Create a cybernetic monkey character positioned at a 45-degree angle, existing within the vast expanse of the universe. The monkey wears a sleek yellow-trimmed suit adorned with futuristic elements and one earring in the ear.". Craft a genuine image by leveraging the edge representations in (image).</p>	<p>Output Image</p> 
	<p>Input Image</p> 	<p>Instruction</p> <p>The head of the robot mimics a large, white, cartoon-like animal with a simple smiling face, featuring large expressive eyes and minimal design features like a cute nose and a small triangle shape for the mouth.". Interpret the aspects of the mask region in the (image) into an authentic image</p>	<p>Output Image</p> 	<p>Input Image</p> 	<p>Instruction</p> <p>Rebuild a lifelike image by using the edges of (image) under the guidance of the mask.</p>	<p>Output Image</p> 
Doodle-based Generation	<p>Input Image</p> 	<p>Instruction</p> <p>"Card captor sakura". Please instruct on creating an equivalent image of the doodle (image).</p>	<p>Output Image</p> 	<p>Input Image</p> 	<p>Instruction</p> <p>"A young man with glowing eyes, hair in a pony tail, epic pose, holding a blue flame Katana, GLOWING samurai armor, in the mountains, painterly style". Create an image that accurately represents the doodle depicted in (image).</p>	<p>Output Image</p> 
	<p>Input Image</p> 	<p>Instruction</p> <p>Create a regional image from the doodle (image), based around the text descriptor "A bright blue, anthropomorphic frog is depicted with a large, expressive eye and several smaller protrusions on its skin." marked by the mask.</p>	<p>Output Image</p> 	<p>Input Image</p> 	<p>Instruction</p> <p>Please provide me with instructions on how to regenerate a portion of the doodle (image) that matches with the description "A young girl with brown hair wearing an orange hoodie." from the masked area mask.</p>	<p>Output Image</p> 
Contour-based Generation	<p>Input Image</p> 	<p>Instruction</p> <p>An artistic illustration of a woman in a bright red jacket and black shorts stands prominently in front of a semi-transparent wall. Behind her, there are gym equipment and what appears to be the silhouette of a tennis court. For the contour (image), create a suitable matching image.</p>	<p>Output Image</p> 	<p>Input Image</p> 	<p>Instruction</p> <p>Please follow the description "a 3D digital art style image of an animated child character, showing a alien child in a space suit.His expression was friendly, and his eyes showed curiosity and excitement". Process the contour map (image) and output the appropriate image.</p>	<p>Output Image</p> 
	<p>Input Image</p> 	<p>Instruction</p> <p>The vehicle is a futuristic concept with a sleek, rounded design, featuring multiple rotors for lift and propulsion. Its exterior is primarily an iridescent orange, with details that suggest advanced materials and technology. By amalgamating the contour image (image) we can produce a localized image in the mask mask area.</p>	<p>Output Image</p> 	<p>Input Image</p> 	<p>Instruction</p> <p>Using the specified mask region and taking "a charming, whimsical stone house, seemingly plucked from a fairytale. The house is constructed from large, unevenly shaped, light brown stones" as the reference, create an image partial corresponding to the (image) contour.</p>	<p>Output Image</p> 
Scribble-based Generation	<p>Input Image</p> 	<p>Instruction</p> <p>"A cute little brown and white dog with erect ears and playful mouth. He's enjoying his coffee and looks like he's enjoying it.". please create the corresponding local image in the (image) draft according to the area indicated by mask.</p>	<p>Output Image</p> 	<p>Input Image</p> 	<p>Instruction</p> <p>"A robotic monster by Edmund McMillen". Could you please generate the image that corresponds to the given scribble (image)?"</p>	<p>Output Image</p> 

Figure 4: The ACE's generated visualization of degradation-based, edge-based, doodle-based, contour-based, and scribble-based generation in controllable generation.

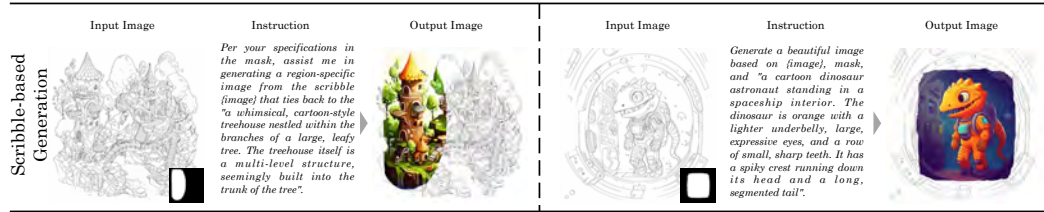


Figure 5: The ACE's generated visualization in scribble-based controllable generation.

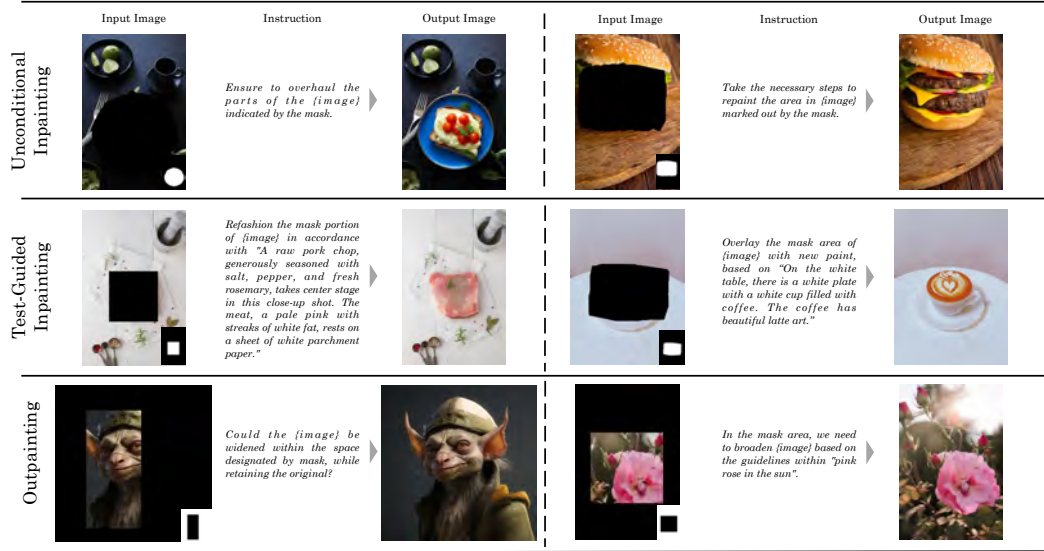


Figure 6: The ACE's generated visualization of repainting.

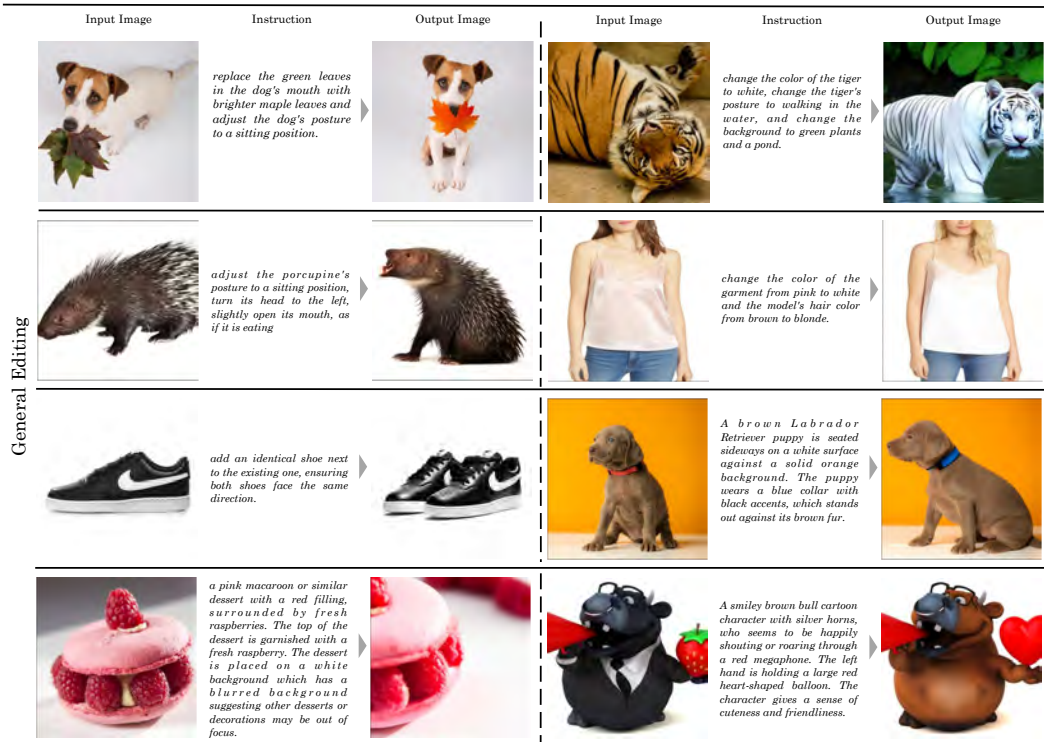


Figure 7: The ACE's generated visualization of general editing in semantic editing.

	Input Image	Instruction	Output Image	Input Image	Instruction	Output Image
Facial Attributes Preservation		Maintain the same face in (image), you need to revert the background to a clear indoor or shaded setting, using bright lights and increasing contrast. The clothing can be adjusted back to a large ruffled white garment or jacket, maintaining its original style.			Restyle the characters from (image) according to "the woman has straight, light brown hair with a side part, which is swept away from her face. She is wearing a white top with cut-out details, and her makeup appears more natural." and make sure their facial attributes remain the same.	
		You need to change the background color from pink elements and stairs or step patterns to a blurry background with a repeating pattern. Remove the shadows and restore the soft glow atmosphere. Also, adjust the background color back to blue. The faces in the two images are the same, usually changing their pose.			Modify the (image) as per the "has her hair flowing more dynamically, suggesting movement or a breeze, and the background has more defined greenery, creating a natural and serene setting. The light is more diffuse and the color tones are rich, which adds to the realistic feel of the scene." while preserving facial identity.	
		Replace the long-sleeved blouse with a high neckline featuring a pattern of red chains against a pale background with a sleeveless top with lace detailing at the straps. Additionally, the large hoop earrings and textured clutch purse should be replaced with a broad smile.			Correspond the composition of (image) with another style taking into account the "The girl has a light blue, long-sleeved shirt with a floral pattern, which seems comfortable and casual. The setting is outdoors, with greenery softly in the background." but keeping the facial aspects constant.	
		Decrease light on the well-lit side to create appropriate shadows. Conceal some parts of the shoulder and arm. Also, adjust the stripe pattern on the collar edge. The faces in the two images are the same, usually changing their pose.			Keep the same facial feature in (image), change the woman's clothing from a white jacket to a white turtleneck sweater and adjust her posture so that she is pulling the collar of the sweater with both hands. Other aspects, such as background, hairstyle, facial expression, etc., remain unchanged.	
		Keep the facial features of the character in (image), based on "change the background to solid pink. Turn the character's shirt black, change their hairstyle to short hair, and make their gaze more determined."			Keep the same facial feature in (image). Transform the girl's outfit into a formal dress with suspenders, spread her hair out, and hold a large bouquet of red roses in her hand. The overall lighting becomes dim, creating a sense of atmosphere	
		Transform the faces of the character in (image) to capture genuine smiles.			Adjust the expressions of the character in (image) to reflect natural, friendly smiles.	
Facial Attributes Transformation		The person appears to be wearing more subtle makeup to enhance his facial features. The skin appears to be smoother, with possibly a touch of foundation to even out the complexion. A light application of blush might have been used to give a healthy glow to the cheeks.			The person in (image) is wearing makeup. The makeup enhances the lips, making them appear fuller and more vivid. The eyebrows are neatly shaped with color added, while the skin looks smoother and perhaps given a lighter tone from the makeup application.	
		Add natural beards to the characters in the (image).			Generate natural facial hair for the character in the (image).	
		The character's hair color appear as vibrant pink, with the same styled bun and loose tendrils framing her face. The character continues to wear the same clothes and accessories. The makeup remains largely unchanged, focusing attention on her facial features.			The woman's hair has transformed to a deep purple hue, with a more pronounced wave similar to that in the left image. The length and style appear to be the same as the left image. The overall effect is still soft and ethereal, but distinctly different in the hair color.	

Figure 8: The ACE's generated visualization of facial editing in semantic editing.

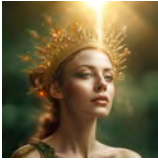









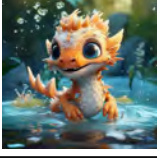
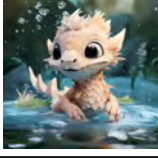


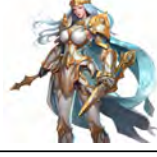



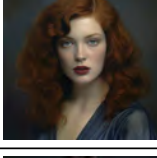

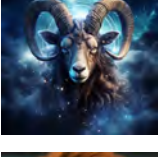

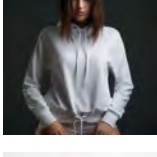
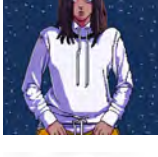


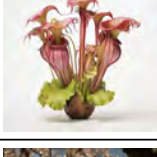
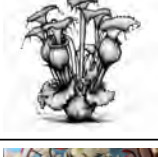

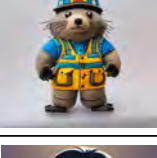


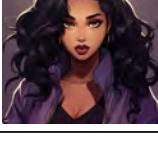
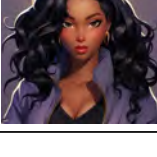
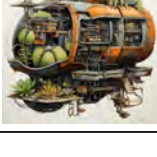
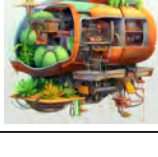
	Input Image	Instruction	Output Image	Input Image	Instruction	Output Image
Schools of Painting		Could you please make [image] an Impressionism painting.			Let [image] be in the style of Fauvism.	
Animation Studios		Create a new image of Walt Disney Animation referring to [image].			Adjust the [image] to capture the essence of Ghibli Studio's style.	
Paper Art		Change the style of [image] to paper cut craft.			Convert [image] into a paper cutting.	
Drawing		Transform [image] into pencil painting.			Give [image] a one line drawing look	
Materials		Make [image] a picture made of stained glass			Make [image] in Play-Doh clay style	
Special Effects		Change [image] to match Low poly style.			Change [image] to match 8-bit pixel art style	
		Re-style [image] to risograph ISO format			Make [image] in ink render style	
Fabrics		Apply felt doll style to [image].			Generate a quilted art style image using the [image] content.	
3D		Could you please make [image] 3D cartoon.			Make [image] a Pixar Animation.	

Figure 9: The ACE's generated visualization of style editing in semantic editing.

	Input Image	Instruction	Output Image	Input Image	Instruction	Output Image
Text Render		On the (image), superimpose the text "LAKE" according to the area identified by the mask			Put the text "CARD" at the position marked by mask in the (image)	
		Embed the text "EVER" in the (image) at the location indicated by mask			By using the mask as a guide, you can position text "RPG" on (image)	
		Stamp text "FRIDAY" into the (image), as defined by the mask coordinates			In the image (image), position the text "Food" according to the guidance of the mask	
		text "STRONG" should be applied to the (image) at the position marked by mask			The mask is to be used as a marker to incorporate the text "DANGEROUS" into the (image)	
		add white text 'UFES' near the center of the image.			add the light brown text 'NVZ' at the bottom right of the picture	
		Rub out any text found in the mask sector of the (image).			Obiterate the text from the specified mask slice of the (image)	
Text Remove		Vacate the text from the identified mask spot on the (image).			Obiterate the text in the mask in the (image) image	
		Aim to remove any textual element in (image)			Let's make the (image) completely devoid of any text	
		Scrub out every text snippet from (image)			Eradicate all text on (image)	

Figure 10: The ACE's generated visualization of text editing in element editing.

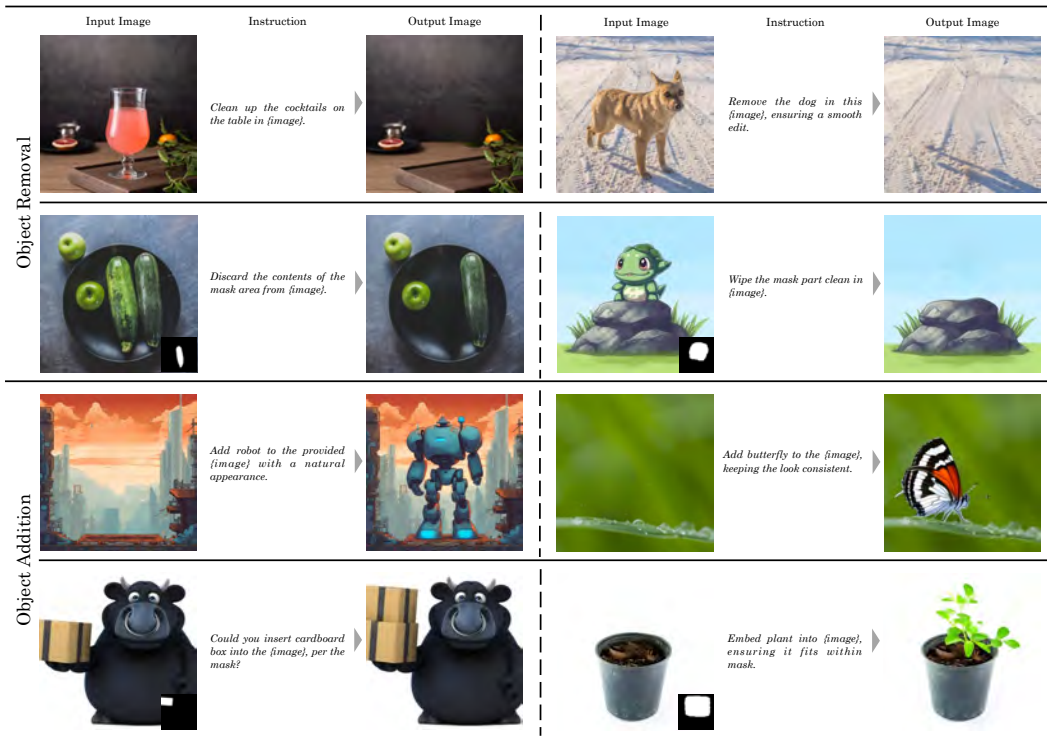


Figure 11: The ACE's generated visualization of object editing in element editing.

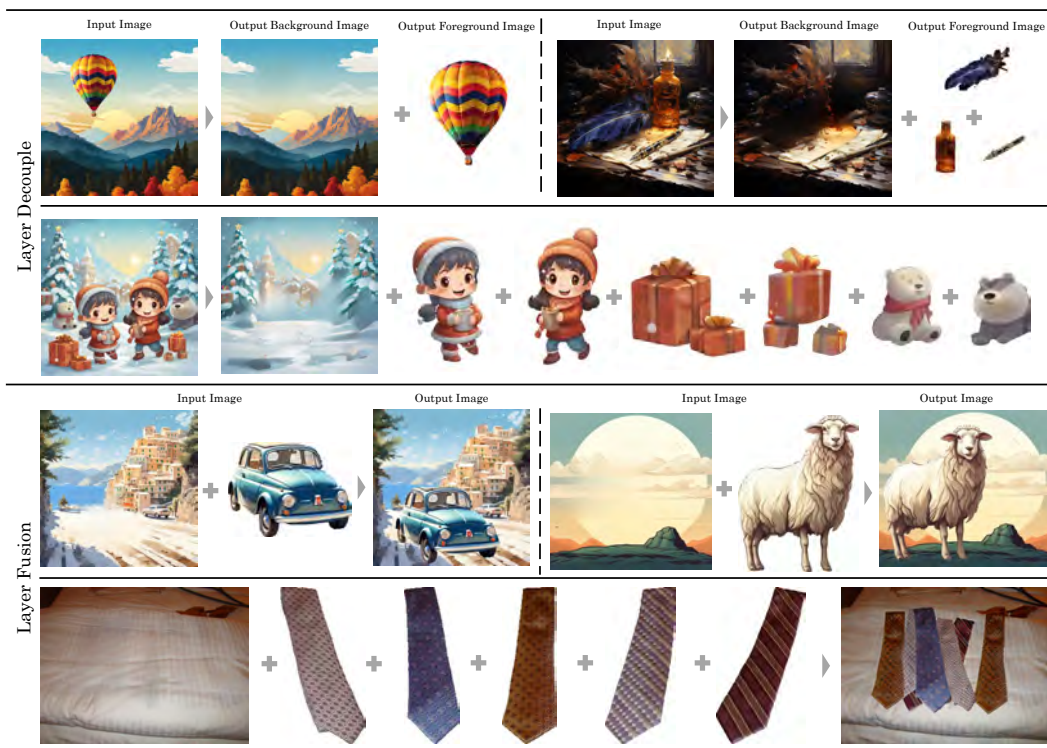


Figure 12: The ACE's generated visualization of layer decouple and layer fusion in layer editing.



Figure 13: The ACE's generated visualization of multi-reference generation and reference-guided editing.

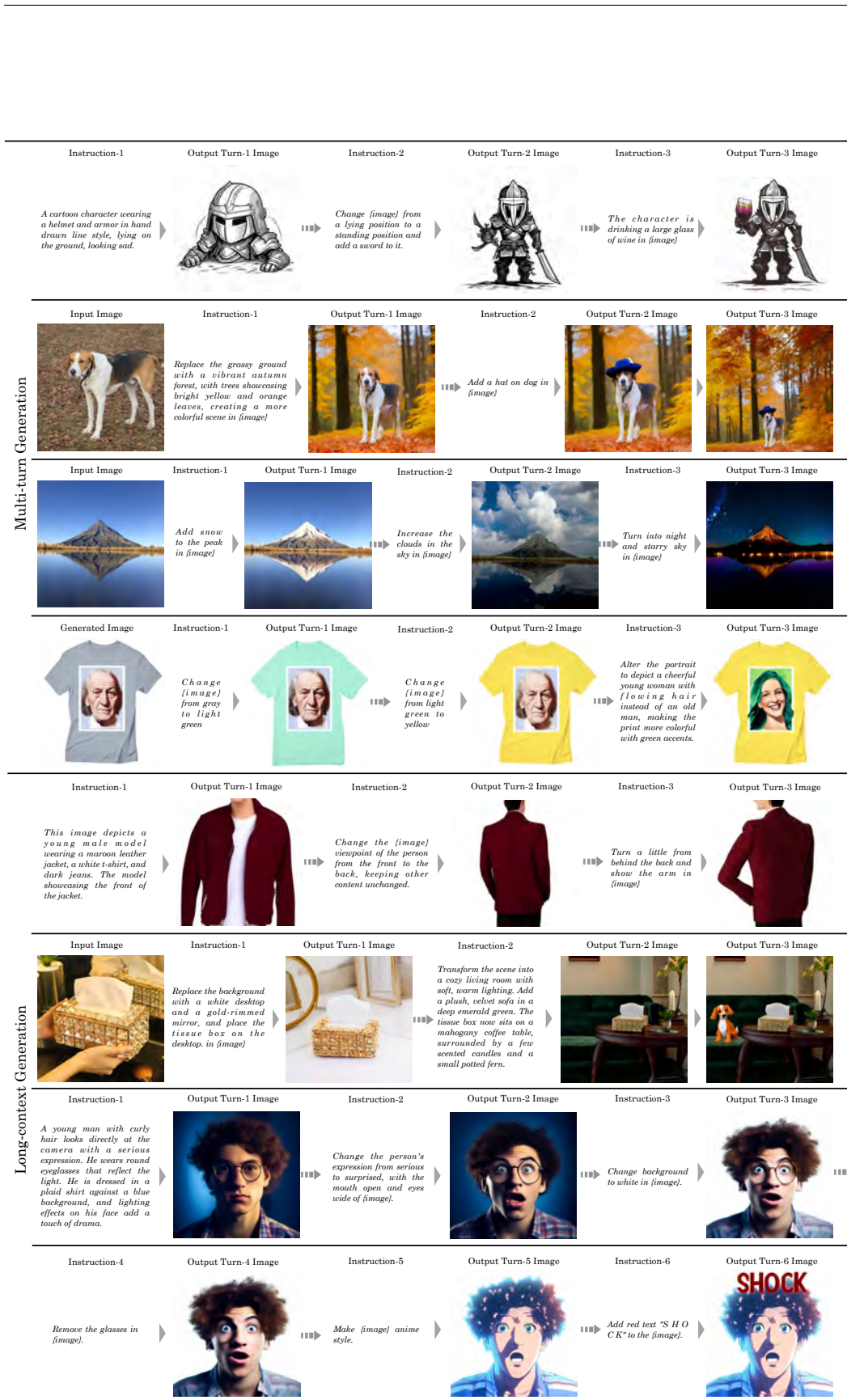


Figure 14: The ACE’s generated visualization of multi-turn and long-context generation.