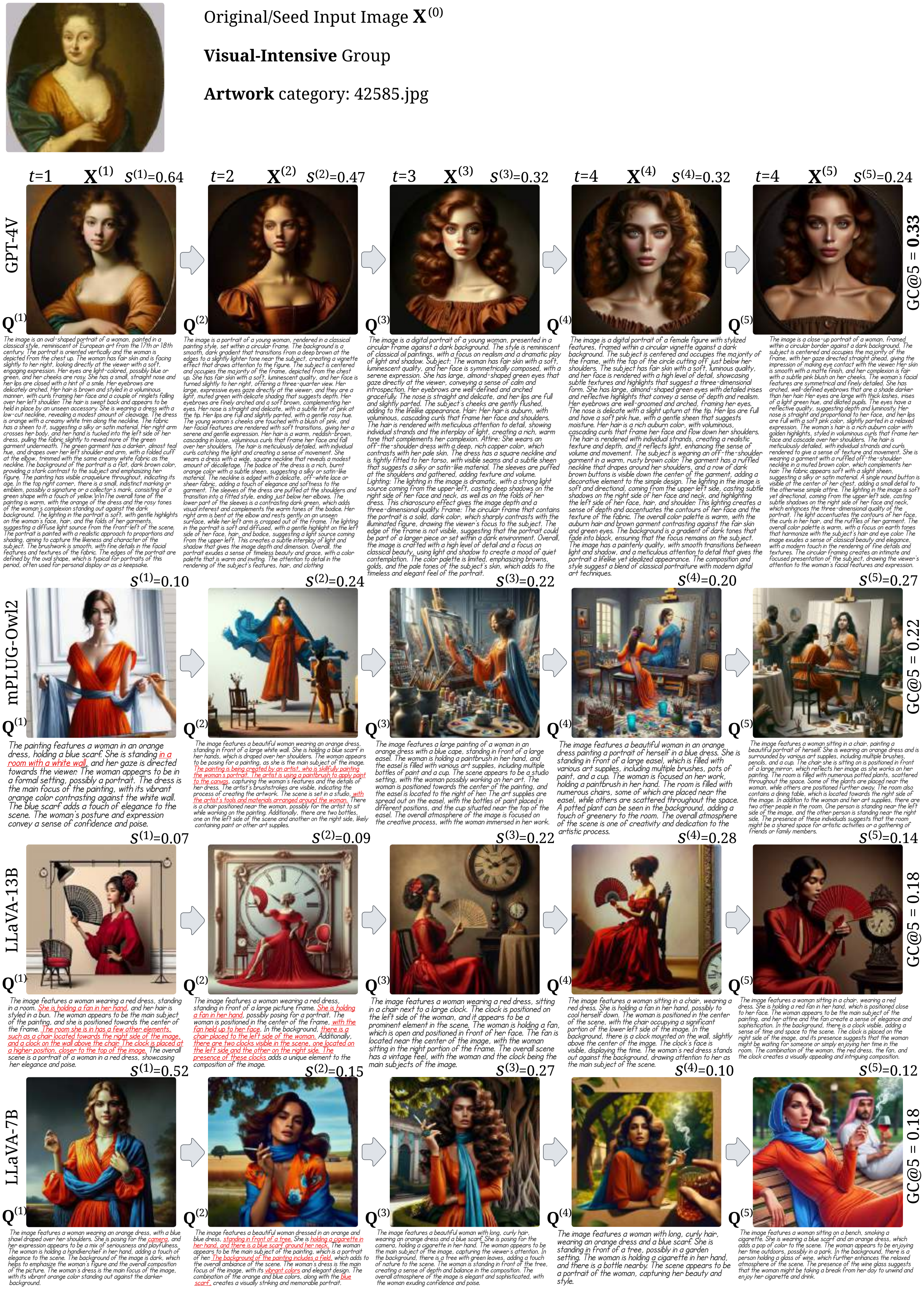


The image is an oval-shaped portrait of a woman, painted in a classical style, reminiscent of European art from the 17th or 18th century. The portrait is oriented vertically and the woman is depicted from the chest up. The woman has fair skin and is facing slightly to her right, looking directly at the viewer with a soft, engaging expression. Her eyes are light-colored, possibly blue or green, and her cheeks are rosy. She has a small, straight nose and her lips are closed with a hint of a smile. Her eyebrows are delicately arched. Her hair is brown and styled in a voluminous manner, with curls framing her face and a couple of ringlets falling over her left shoulder. The hair is swept back and appears to be held in place by an unseen accessory. She is wearing a dress with a low-cut neckline, revealing a modest amount of cleavage. The dress is orange with a creamy white trim along the neckline. The fabric has a sheen to it, suggesting a silky or satin material. Her right arm crosses her body, and her hand is tucked into the left side of her dress, pulling the fabric slightly to reveal more of the green garment underneath. The green garment has a darker, almost teal hue, and drapes over her left shoulder and arm, with a folded cuff at the elbow, trimmed with the same creamy white fabric as the neckline.The background of the portrait is a flat, dark brown color, providing a stark contrast to the subject and emphasizing her figure. The painting has visible craquelure throughout, indicating its age. In the top right corner, there is a small, indistinct marking or emblem, possibly a signature or a collector's mark, consisting of a green shape with a touch of yellow.\n\nThe overall tone of the painting is warm, with the orange of the dress and the rosy tones of the woman's complexion standing out against the dark background. The lighting in the portrait is soft, with gentle highlights on the woman's face, hair, and the folds of her garments, suggesting a diffuse light source from the front-left of the scene. The portrait is painted with a realistic approach to proportions and shading, aiming to capture the likeness and character of the subject. The brushwork is smooth, with fine details in the facial features and textures of the fabric. The edges of the portrait are defined by the oval shape, which is typical for portraits of this period, often used for personal display or as a keepsake.

The image is a portrait of a young woman, rendered in a classical painting style, set within a circular frame. The background is a smooth, dark gradient that transitions from a deep brown at the edges to a slightly lighter tone near the subject, creating a vignette effect that draws attention to the figure. The subject is centered and occupies the majority of the frame, depicted from the chest up. She has fair skin with a soft, luminescent quality, and her face is turned slightly to her right, offering a three-quarter view. Her large, expressive eyes gaze directly at the viewer, and they are a light, muted green with delicate shading that suggests depth. Her eyebrows are finely arched and a soft brown, complementing her eyes. Her nose is straight and delicate, with a subtle hint of pink at the tip. Her lips are full and slightly parted, with a gentle rosy hue. The young woman's cheeks are touched with a blush of pink, and her facial features are rendered with soft transitions, giving her a serene and gentle expression. Her hair is a warm, reddish-brown, cascading in loose, voluminous curls that frame her face and fall over her shoulders. The hair is meticulously detailed, with individual curls catching the light and creating a sense of movement. She wears a dress with a wide, square neckline that reveals a modest amount of décolletage. The bodice of the dress is a rich, burnt orange color with a subtle sheen, suggesting a silky or satin-like material. The neckline is edged with a delicate, off-white lace or sheer fabric, adding a touch of elegance and softness to the garment. The sleeves of the dress are puffed at the shoulders and transition into a fitted style, ending just below her elbows. The lower part of the sleeves is a contrasting dark green, which adds visual interest and complements the warm tones of the bodice. Her right arm is bent at the elbow and rests gently on an unseen surface, while her left arm is cropped out of the frame. The lighting in the portrait is soft and diffused, with a gentle highlight on the left side of her face, hair, and bodice, suggesting a light source coming from the upper left. This creates a subtle interplay of light and shadow that gives the image depth and dimension. Overall, the portrait exudes a sense of timeless beauty and grace, with a color palette that is warm and inviting. The attention to detail in the rendering of the subject's features, hair, and clothing

The image is a digital portrait of a young woman, presented in a circular frame against a dark background. The style is reminiscent of classical oil paintings, with a focus on realism and a dramatic play of light and shadow. Subject: The woman has fair skin with a soft, luminescent quality, and her face is symmetrically composed, with a serene expression. She has large, almond-shaped green eyes that gaze directly at the viewer, conveying a sense of calm and introspection. Her eyebrows are well-defined and arched gracefully. The nose is straight and delicate, and her lips are full and slightly parted. The subject's cheeks are gently flushed, adding to the lifelike appearance. Hair: Her hair is auburn, with voluminous, cascading curls that frame her face and shoulders. The hair is rendered with meticulous attention to detail, showing individual strands and the interplay of light, creating a rich, warm tone that complements her complexion. Attire: She wears an off-the-shoulder dress with a deep, rich copper color, which contrasts with her pale skin. The dress has a square neckline and is tightly fitted to her torso, with visible seams and a subtle sheen that suggests a silky or satin-like material. The sleeves are puffed at the shoulders and gathered, adding texture and volume. Lighting: The lighting in the image is dramatic, with a strong light source coming from the upper left, casting deep shadows on the right side of her face and neck, as well as on the folds of her dress. This chiaroscuro effect gives the image depth and a three-dimensional quality. Frame: The circular frame that contains the portrait is a solid, dark color, which sharply contrasts with the illuminated figure, drawing the viewer's focus to the subject. The edge of the frame is not visible, suggesting that the portrait could be part of a larger piece or set within a dark environment. Overall, the image is crafted with a high level of detail and a focus on classical beauty, using light and shadow to create a mood of quiet contemplation. The color palette is limited, emphasizing browns, golds, and the pale tones of the subject's skin, which adds to the timeless and elegant feel of the portrait.

The image is a digital portrait of a female figure with stylized features, framed within a circular vignette against a dark background. The subject is centered and occupies the majority of the frame, with the top of the circle cutting off just below her shoulders. The subject has fair skin with a soft, luminous quality, and her face is rendered with a high level of detail, showcasing subtle textures and highlights that suggest a three-dimensional form. She has large, almond-shaped green eyes with detailed irises and reflective highlights that convey a sense of depth and realism. Her eyebrows are well-groomed and arched, framing her eyes. The nose is delicate with a slight upturn at the tip. Her lips are full and have a soft pink hue, with a gentle sheen that suggests moisture. Her hair is a rich auburn color, with voluminous, cascading curls that frame her face and flow down her shoulders. The hair is rendered with individual strands, creating a realistic texture and depth, and it reflects light, enhancing the sense of volume and movement. The subject is wearing an off-the-shoulder garment in a warm, rusty brown color. The garment has a ruffled neckline that drapes around her shoulders, and a row of dark brown buttons is visible down the center of the garment, adding a decorative element to the simple design. The lighting in the image is soft and directional, coming from the upper left side, casting subtle shadows on the right side of her face and neck, and highlighting the left side of her face, hair, and shoulder. This lighting creates a sense of depth and accentuates the contours of her face and the texture of the fabric. The overall color palette is warm, with the auburn hair and brown garment contrasting against the fair skin and green eyes. The background is a gradient of dark tones that fade into black, ensuring that the focus remains on the subject. The image has a painterly quality, with smooth transitions between light and shadow, and a meticulous attention to detail that gives the portrait a lifelike yet idealized appearance. The composition and style suggest a blend of classical portraiture with modern digital art techniques.

The image is a close-up portrait of a woman, framed within a circular border against a dark background. The subject is centered and occupies the majority of the frame, with her gaze directed straight ahead, giving the impression of making eye contact with the viewer. Her skin is smooth with a matte finish, and her complexion is fair with a subtle pink blush on her cheeks. The woman's facial features are symmetrical and finely detailed. She has arched, well-defined eyebrows that are a shade darker than her hair. Her eyes are large with thick lashes, irises of a light green hue, and dilated pupils. The eyes have a reflective quality, suggesting depth and luminosity. Her nose is straight and proportional to her face, and her lips are full with a soft pink color, slightly parted in a relaxed expression. The woman's hair is a rich auburn color with golden highlights, styled in voluminous curls that frame her face and cascade over her shoulders. The hair is meticulously detailed, with individual strands and curls rendered to give a sense of texture and movement. She is wearing a garment with a ruffled off-the-shoulder neckline in a muted brown color, which complements her hair. The fabric appears soft with a slight sheen, suggesting a silky or satin material. A single round button is visible at the center of her chest, adding a small detail to the otherwise simple attire. The lighting in the image is soft yet directional, coming from the upper left side, casting subtle shadows on the right side of her face and neck, which enhances the three-dimensional quality of the portrait. The light accentuates the contours of her face, the curls in her hair, and the ruffles of her garment. The overall color palette is warm, with a focus on earth tones that harmonize with the subject's hair and eye color. The image exudes a sense of classical beauty and elegance, with a modern touch in the rendering of fine details and textures. The circular framing creates an intimate and focused presentation of the subject, drawing the viewer's attention to the woman's facial features and expression.

The painting features a woman in an orange dress, holding a blue scarf. She is standing in a room with a white wall, and her gaze is directed towards the viewer. The woman appears to be in a formal setting, possibly a portrait. The dress is the main focus of the painting, with its vibrant orange color contrasting against the white wall. The blue scarf adds a touch of elegance to the scene. The woman's posture and expression convey a sense of confidence and poise.

The image features a beautiful woman wearing an orange dress, standing in front of a large white wall. She is holding a blue scarf in her hands, which is draped over her shoulders. The woman appears to be posing for a painting, as she is the main subject of the image. The painting is being created by an artist, who is skillfully painting the woman's portrait. The artist is using a paintbrush to apply paint to the canvas, capturing the woman's features and the details of her dress. The artist's brushstrokes are visible, indicating the process of creating the artwork. The scene is set in a studio, with the artist's tools and materials arranged around the woman. There is a chair positioned near the woman, possibly for the artist to sit while working on the painting. Additionally, there are two bottles, one on the left side of the scene and another on the right side, likely containing paint or other art supplies.

The image features a large painting of a woman in an orange dress with a blue cape, standing in front of a large easel. The woman is holding a paintbrush in her hand, and the easel is filled with various art supplies, including multiple bottles of paint and a cup. The scene appears to be a studio setting, with the woman possibly working on her art. The woman is positioned towards the center of the painting, and the easel is located to the right of her. The art supplies are spread out on the easel, with the bottles of paint placed in different positions, and the cup situated near the top of the easel. The overall atmosphere of the image is focused on the creative process, with the woman immersed in her work.

The image features a beautiful woman in an orange dress painting a portrait of herself in a blue dress. She is standing in front of a large easel, which is filled with various art supplies, including multiple brushes, pots of paint, and a cup. The woman is focused on her work, holding a paintbrush in her hand. The room is filled with numerous chairs, some of which are placed near the easel, while others are scattered throughout the space. A potted plant can be seen in the background, adding a touch of greenery to the room. The overall atmosphere of the scene is one of creativity and dedication to the artistic process.

The image features a woman sitting in a chair, painting a beautiful portrait of herself. She is wearing an orange dress and is surrounded by various art supplies, including multiple brushes, pencils, and a cup. The chair she is sitting on is positioned in front of a large mirror, which reflects her image as she works on her painting. The room is filled with numerous potted plants, scattered throughout the space. Some of the plants are placed near the woman, while others are positioned further away. The room also contains a dining table, which is located towards the right side of the image. In addition to the woman and her art supplies, there are two other people in the room. One person is standing near the left side of the image, and the other person is standing near the right side. The presence of these individuals suggests that the room might be a shared space for artistic activities or a gathering of friends or family members.

The image features a woman wearing a red dress, standing in a room. She is holding a fan in her hand, and her hair is styled in a bun. The woman appears to be the main subject of the painting, and she is positioned towards the center of the frame. The room she is in has a few other elements, such as a chair located towards the right side of the image, and a clock on the wall above the chair. The clock is placed at a higher position, closer to the top of the image. The overall scene is a portrait of a woman in a red dress, showcasing her elegance and poise.

The image features a woman wearing a red dress, standing in front of a large picture frame. She is holding a fan in her hand, possibly posing for a portrait. The woman is positioned in the center of the frame, with the fan held up to her face. In the background, there is a chair placed to the left side of the woman. Additionally, there are two clocks visible in the scene, one located on the left side and the other on the right side. The presence of these clocks adds a unique element to the composition of the image.

The image features a woman wearing a red dress, sitting in a chair next to a large clock. The clock is positioned on the left side of the woman, and it appears to be a prominent element in the scene. The woman is holding a fan, which is open and positioned in front of her face. The fan is located near the center of the image, with the woman sitting in the right portion of the frame. The overall scene has a vintage feel, with the woman and the clock being the main subjects of the image.

The image features a woman sitting in a chair, wearing a red dress. She is holding a fan in her hand, possibly to cool herself down. The woman is positioned in the center of the scene, with the chair occupying a significant portion of the lower left side of the image. In the background, there is a clock mounted on the wall, slightly above the center of the image. The clock's face is visible, displaying the time. The woman's red dress stands out against the background, drawing attention to her as the main subject of the scene.

The image features a woman sitting in a chair, wearing a red dress. She is holding a red fan in her hand, which is positioned close to her face. The woman appears to be the main subject of the painting, and her attire and the fan create a sense of elegance and sophistication. In the background, there is a clock visible, adding a sense of time and space to the scene. The clock is placed on the right side of the image, and its presence suggests that the woman might be waiting for someone or simply enjoying her time in the room. The combination of the woman, the red dress, the fan, and the clock creates a visually appealing and intriguing composition.

The image features a woman wearing an orange dress, with a blue shawl draped over her shoulders. She is posing for the camera, and her expression appears to be a mix of seriousness and playfulness. The woman is holding a handkerchief in her hand, adding a touch of elegance to the scene. The background of the image is dark, which helps to emphasize the woman's figure and the overall composition of the picture. The woman's dress is the main focus of the image, with its vibrant orange color standing out against the darker background.

The image features a beautiful woman dressed in an orange and blue dress, standing in front of a tree. She is holding a cigarette in her hand, and there is a blue scarf around her neck. The woman appears to be the main subject of the painting, which is a portrait of her. The background of the painting includes a field, which adds to the overall ambiance of the scene. The woman's dress is the main focus of the image, with its vibrant colors and elegant design. The combination of the orange and blue colors, along with the blue scarf, creates a visually striking and memorable portrait.

The image features a beautiful woman with long, curly hair, wearing an orange dress and a blue scarf. She is posing for the camera, holding a cigarette in her hand. The woman appears to be the main subject of the image, capturing the viewer's attention. In the background, there is a tree with green leaves, adding a touch of nature to the scene. The woman is standing in front of the tree, creating a sense of depth and balance in the composition. The overall atmosphere of the image is elegant and sophisticated, with the woman exuding confidence and poise.

The image features a woman with long, curly hair, wearing an orange dress and a blue scarf. She is standing in front of a tree, possibly in a garden setting. The woman is holding a cigarette in her hand, and there is a bottle nearby. The scene appears to be a portrait of the woman, capturing her beauty and style.

The image features a woman sitting on a bench, smoking a cigarette. She is wearing a blue scarf and an orange dress, which adds a pop of color to the scene. The woman appears to be enjoying her time outdoors, possibly in a park. In the background, there is a person holding a glass of wine, which further enhances the relaxed atmosphere of the scene. The presence of the wine glass suggests that the woman might be taking a break from her day to unwind and enjoy her cigarette and drink.

Figure 1: Illustration of a 5-iteration GenCeption procedure run on artwork images to evaluate 4 VLLMs. Each iteration t shows the generated image $\mathbf{X}^{(t)}$, the description $\mathbf{Q}^{(t)}$ of the preceding image $\mathbf{X}^{(t-1)}$, and the similarity score $s^{(t)}$ relative to $\mathbf{X}^{(0)}$. The GC@5 metric for each VLLM is also presented. Hallucinated elements within descriptions $\mathbf{Q}^{(1)}$ and $\mathbf{Q}^{(2)}$ as compared to the seed image are indicated with red underlining.

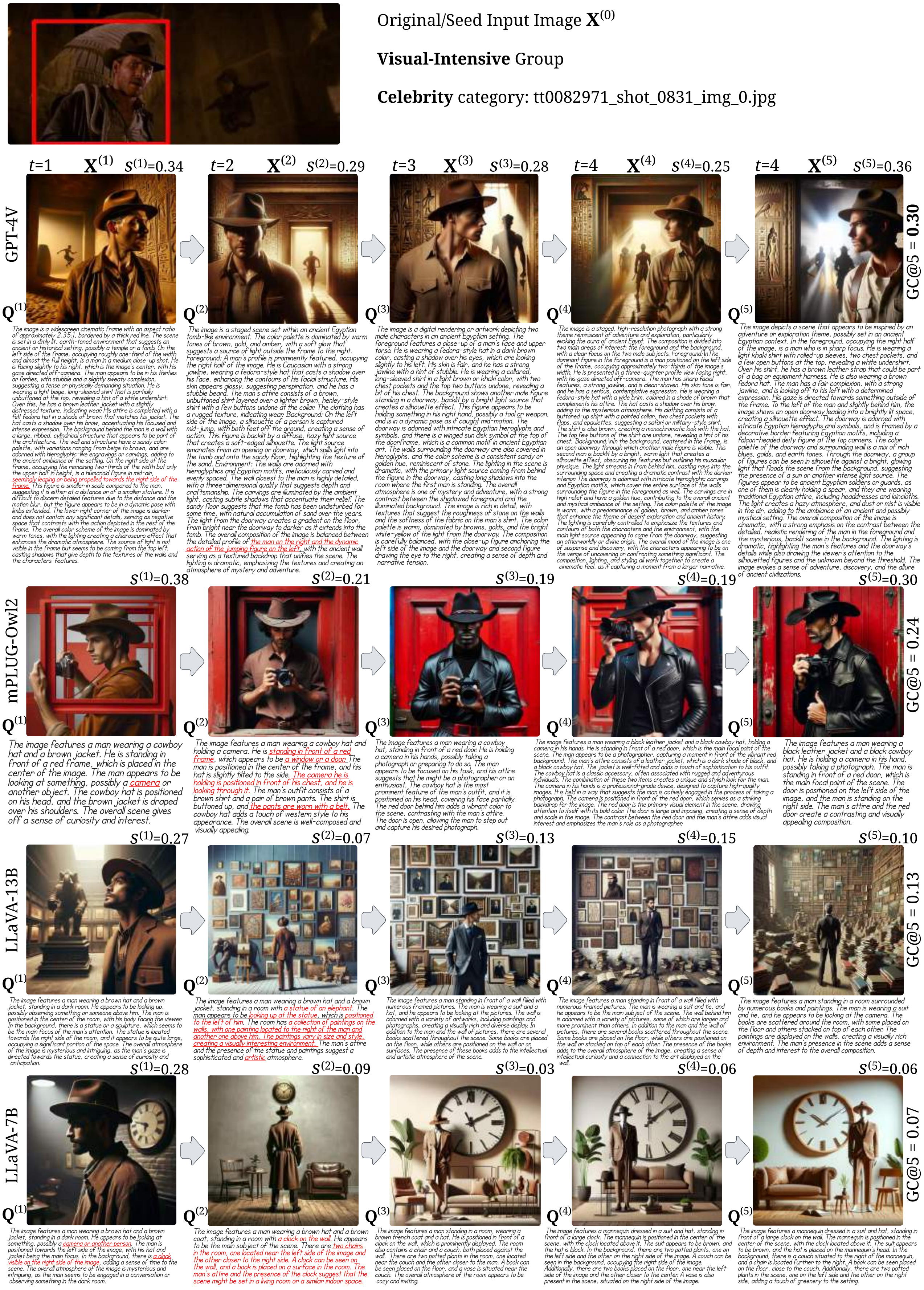
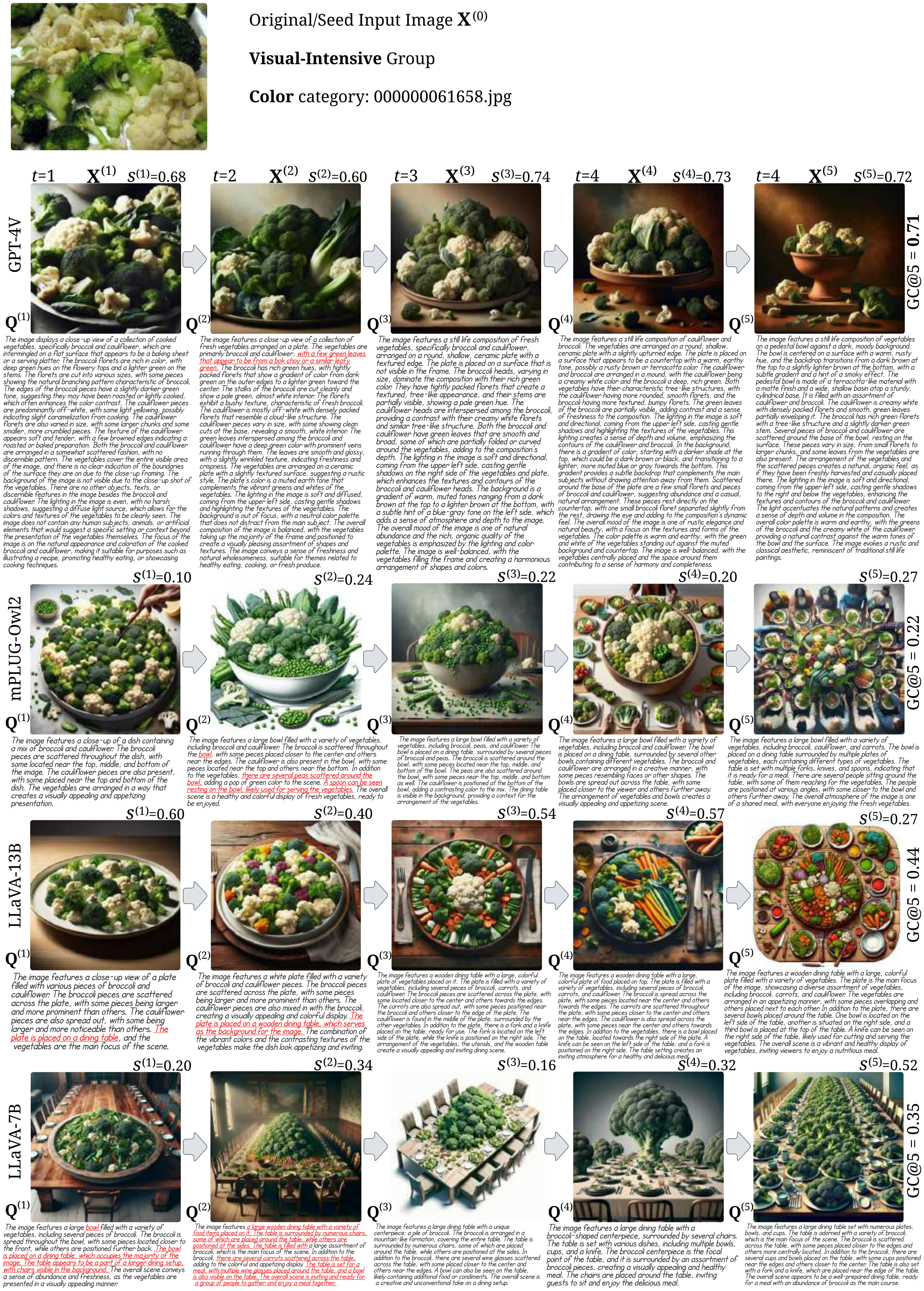

Original/Seed Input Image $\mathbf{X}^{(0)}$
Visual-Intensive Group
Celebrity category: tt0082971_shot_0831_img_0.jpg
t=1 $\mathbf{X}^{(1)}$ $s^{(1)}$=0.34
t=2 $\mathbf{X}^{(2)}$ $s^{(2)}$=0.29
t=3 $\mathbf{X}^{(3)}$ $s^{(3)}$=0.28
t=4 $\mathbf{X}^{(4)}$ $s^{(4)}$=0.25
t=4 $\mathbf{X}^{(5)}$ $s^{(5)}$=0.36
GPT-4V
GC@5 = 0.30
$\mathbf{Q}^{(1)}$ $\mathbf{Q}^{(2)}$ $\mathbf{Q}^{(3)}$ $\mathbf{Q}^{(4)}$ $\mathbf{Q}^{(5)}$
$s^{(1)}$=0.38 $s^{(2)}$=0.21 $s^{(3)}$=0.19 $s^{(4)}$=0.19 $s^{(5)}$=0.30
mPLUG-Owl2
GC@5 = 0.24
$s^{(1)}$=0.27 $s^{(2)}$=0.07 $s^{(3)}$=0.13 $s^{(4)}$=0.15 $s^{(5)}$=0.10
LLaVA-13B
GC@5 = 0.13
$s^{(1)}$=0.28 $s^{(2)}$=0.09 $s^{(3)}$=0.03 $s^{(4)}$=0.06 $s^{(5)}$=0.06
LLaVA-7B
GC@5 = 0.07

Figure 2: Illustration of a 5-iteration GenCeption procedure run on celebrity images to evaluate 4 VLLMs. Each iteration t shows the generated image $\mathbf{X}^{(t)}$, the description $\mathbf{Q}^{(t)}$ of the preceding image $\mathbf{X}^{(t-1)}$, and the similarity score $s^{(t)}$ relative to $\mathbf{X}^{(0)}$. The GC@5 metric for each VLLM is also presented. Hallucinated elements within descriptions $\mathbf{Q}^{(1)}$ and $\mathbf{Q}^{(2)}$ as compared to the seed image are indicated with red underlining.


Original/Seed Input Image $\mathbf{X}^{(0)}$
Visual-Intensive Group
Color category: 000000061658.jpg
t=1 $\mathbf{X}^{(1)}$ $s^{(1)}$=0.68
t=2 $\mathbf{X}^{(2)}$ $s^{(2)}$=0.60
t=3 $\mathbf{X}^{(3)}$ $s^{(3)}$=0.74
t=4 $\mathbf{X}^{(4)}$ $s^{(4)}$=0.73
t=4 $\mathbf{X}^{(5)}$ $s^{(5)}$=0.72
GPT-4V
GC@5 = **0.71**
mPLUG-Owl2
$s^{(1)}$=0.10 $s^{(2)}$=0.24 $s^{(3)}$=0.22 $s^{(4)}$=0.20 $s^{(5)}$=0.27
GC@5 = 0.22
LLaVA-13B
$s^{(1)}$=0.60 $s^{(2)}$=0.40 $s^{(3)}$=0.54 $s^{(4)}$=0.57 $s^{(5)}$=0.27
GC@5 = 0.44
LLaVA-7B
$s^{(1)}$=0.20 $s^{(2)}$=0.34 $s^{(3)}$=0.16 $s^{(4)}$=0.32 $s^{(5)}$=0.52
GC@5 = 0.35

$\mathbf{Q}^{(1)}$ The image displays a close-up view of a collection of cooked vegetables, specifically broccoli and cauliflower, which are intermingled on a flat surface that appears to be a baking sheet or a serving platter. The broccoli florets are rich in color, with deep green hues on the flowery tops and a lighter green on the stems. The florets are cut into various sizes, with some pieces showing the natural branching pattern characteristic of broccoli. The edges of the broccoli pieces have a slightly darker green tone, suggesting they may have been roasted or lightly cooked, which often enhances the color contrast. The cauliflower pieces are predominantly off-white, with some light yellowing, possibly indicating slight caramelization from cooking. The cauliflower florets are also varied in size, with some larger chunks and some smaller, more crumbled pieces. The texture of the cauliflower appears soft and tender, with a few browned edges indicating a roasted or baked preparation. Both the broccoli and cauliflower are arranged in a somewhat scattered fashion, with no discernible pattern. The vegetables cover the entire visible area of the image, and there is no clear indication of the boundaries of the surface they are on due to the close-up framing. The background of the image is not visible due to the close-up shot of the vegetables. There are no other objects, texts, or discernible features in the image besides the broccoli and cauliflower. The lighting in the image is even, with no harsh shadows, suggesting a diffuse light source, which allows for the colors and textures of the vegetables to be clearly seen. The image does not contain any human subjects, animals, or artificial elements that would suggest a specific setting or context beyond the presentation of the vegetables themselves. The focus of the image is on the natural appearance and coloration of the cooked broccoli and cauliflower, making it suitable for purposes such as illustrating a recipe, promoting healthy eating, or showcasing cooking techniques.

$\mathbf{Q}^{(2)}$ The image features a close-up view of a collection of fresh vegetables arranged on a plate. The vegetables are primarily broccoli and cauliflower, with a few green leaves that appear to be from a bok choy or a similar leafy green. The broccoli has rich green hues, with tightly packed florets that show a gradient of color from dark green on the outer edges to a lighter green toward the center. The stalks of the broccoli are cut cleanly and show a pale green, almost white interior. The florets exhibit a bushy texture, characteristic of fresh broccoli. The cauliflower is mostly off-white with densely packed florets that resemble a cloud-like structure. The cauliflower pieces vary in size, with some showing clean cuts at the base, revealing a smooth, white interior. The green leaves interspersed among the broccoli and cauliflower have a deep green color with prominent veins running through them. The leaves are smooth and glossy, with a slightly wrinkled texture, indicating freshness and crispness. The vegetables are arranged on a ceramic plate with a slightly textured surface, suggesting a rustic style. The plate's color is a muted earth tone that complements the vibrant greens and whites of the vegetables. The lighting in the image is soft and diffused, coming from the upper left side, casting gentle shadows and highlighting the textures of the vegetables. The background is out of focus, with a neutral color palette that does not distract from the main subject. The overall composition of the image is balanced, with the vegetables taking up the majority of the frame and positioned to create a visually pleasing assortment of shapes and textures. The image conveys a sense of freshness and natural wholesomeness, suitable for themes related to healthy eating, cooking, or fresh produce.

$\mathbf{Q}^{(3)}$ The image features a still life composition of fresh vegetables, specifically broccoli and cauliflower, arranged on a round, shallow, ceramic plate with a textured edge. The plate is placed on a surface that is not visible in the frame. The broccoli heads, varying in size, dominate the composition with their rich green color. They have tightly packed florets that create a textured, tree-like appearance, and their stems are partially visible, showing a pale green hue. The cauliflower heads are interspersed among the broccoli, providing a contrast with their creamy white florets and similar tree-like structure. Both the broccoli and cauliflower have green leaves that are smooth and broad, some of which are partially folded or curved around the vegetables, adding to the composition's depth. The lighting in the image is soft and directional, coming from the upper left side, casting gentle shadows on the right side of the vegetables and plate, which enhances the textures and contours of the broccoli and cauliflower heads. The background is a gradient of warm, muted tones ranging from a dark brown at the top to a lighter brown at the bottom, with a subtle hint of a blue-gray tone on the left side, which adds a sense of atmosphere and depth to the image. The overall mood of the image is one of natural abundance and the rich, organic quality of the vegetables is emphasized by the lighting and color palette. The image is well-balanced, with the vegetables filling the frame and creating a harmonious arrangement of shapes and colors.

$\mathbf{Q}^{(4)}$ The image features a still life composition of cauliflower and broccoli. The vegetables are arranged on a round, shallow, ceramic plate with a slightly upturned edge. The plate is placed on a surface that appears to be a countertop with a warm, earthy tone, possibly a rusty brown or terracotta color. The cauliflower and broccoli are arranged in a mound, with the cauliflower being a creamy white color and the broccoli a deep, rich green. Both vegetables have their characteristic tree-like structures, with the cauliflower having more rounded, smooth florets, and the broccoli having more textured, bumpy florets. The green leaves of the broccoli are partially visible, adding contrast and a sense of freshness to the composition. The lighting in the image is soft and directional, coming from the upper left side, casting gentle shadows and highlighting the textures of the vegetables. This lighting creates a sense of depth and volume, emphasizing the contours of the cauliflower and broccoli. In the background, there is a gradient of color, starting with a darker shade at the top, which could be a dark brown or black, and transitioning to a lighter, more muted blue or gray towards the bottom. This gradient provides a subtle backdrop that complements the main subjects without drawing attention away from them. Scattered around the base of the plate are a few small florets and pieces of broccoli and cauliflower, suggesting abundance and a casual, natural arrangement. These pieces rest directly on the countertop, with one small broccoli floret separated slightly from the rest, drawing the eye and adding to the composition's dynamic feel. The overall mood of the image is one of rustic elegance and natural beauty, with a focus on the textures and forms of the vegetables. The color palette is warm and earthy, with the green and white of the vegetables standing out against the muted background and countertop. The image is well-balanced, with the vegetables centrally placed and the space around them contributing to a sense of harmony and completeness.

$\mathbf{Q}^{(5)}$ The image features a still life composition of vegetables on a pedestal bowl against a dark, moody background. The bowl is centered on a surface with a warm, rusty hue, and the backdrop transitions from a dark brown at the top to a slightly lighter brown at the bottom, with a subtle gradient and a hint of a smoky effect. The pedestal bowl is made of a terracotta-like material with a matte finish and a wide, shallow basin atop a sturdy, cylindrical base. It is filled with an assortment of cauliflower and broccoli. The cauliflower is creamy white with densely packed florets and smooth, green leaves partially enveloping it. The broccoli has rich green florets with a tree-like structure and a slightly darker green stem. Several pieces of broccoli and cauliflower are scattered around the base of the bowl, resting on the surface. These pieces vary in size, from small florets to larger chunks, and some leaves from the vegetables are also present. The arrangement of the vegetables and the scattered pieces creates a natural, organic feel, as if they have been freshly harvested and casually placed there. The lighting in the image is soft and directional, coming from the upper left side, casting gentle shadows to the right and below the vegetables, enhancing the textures and contours of the broccoli and cauliflower. The light accentuates the natural patterns and creates a sense of depth and volume in the composition. The overall color palette is warm and earthy, with the greens of the broccoli and the creamy white of the cauliflower providing a natural contrast against the warm tones of the bowl and the surface. The image evokes a rustic and classical aesthetic, reminiscent of traditional still life paintings.

$\mathbf{Q}^{(1)}$ The image features a close-up of a dish containing a mix of broccoli and cauliflower. The broccoli pieces are scattered throughout the dish, with some located near the top, middle, and bottom of the image. The cauliflower pieces are also present, with some placed near the top and bottom of the dish. The vegetables are arranged in a way that creates a visually appealing and appetizing presentation.

$\mathbf{Q}^{(2)}$ The image features a large bowl filled with a variety of vegetables, including broccoli and cauliflower. The broccoli is scattered throughout the bowl, with some pieces placed closer to the center and others near the edges. The cauliflower is also present in the bowl, with some pieces located near the top and others near the bottom. In addition to the vegetables, there are several peas scattered around the bowl, adding a pop of green color to the scene. A spoon can be seen resting on the bowl, likely used for serving the vegetables. The overall scene is a healthy and colorful display of fresh vegetables, ready to be enjoyed.

$\mathbf{Q}^{(3)}$ The image features a large bowl filled with a variety of vegetables, including broccoli, peas, and cauliflower. The bowl is placed on a dining table, surrounded by several pieces of broccoli and peas. The broccoli is scattered around the bowl, with some pieces located near the top, middle, and bottom of the bowl. The peas are also scattered around the bowl, with some pieces near the top, middle, and bottom as well. The cauliflower is positioned at the bottom of the bowl, adding a contrasting color to the mix. The dining table is visible in the background, providing a context for the arrangement of the vegetables.

$\mathbf{Q}^{(4)}$ The image features a large bowl filled with a variety of vegetables, including broccoli and cauliflower. The bowl is placed on a dining table, surrounded by several other bowls containing different vegetables. The broccoli and cauliflower are arranged in a creative manner, with some pieces resembling faces or other shapes. The bowls are spread out across the table, with some placed closer to the viewer and others further away. The arrangement of vegetables and bowls creates a visually appealing and appetizing scene.

$\mathbf{Q}^{(5)}$ The image features a large bowl filled with a variety of vegetables, including broccoli, cauliflower, and carrots. The bowl is placed on a dining table surrounded by multiple plates of vegetables, each containing different types of vegetables. The table is set with multiple forks, knives, and spoons, indicating that it is ready for a meal. There are several people sitting around the table, with some of them reaching for the vegetables. The people are positioned at various angles, with some closer to the bowl and others further away. The overall atmosphere of the image is one of a shared meal, with everyone enjoying the fresh vegetables.

$\mathbf{Q}^{(1)}$ The image features a close-up view of a plate filled with various pieces of broccoli and cauliflower. The broccoli pieces are scattered across the plate, with some pieces being larger and more prominent than others. The cauliflower pieces are also spread out, with some being larger and more noticeable than others. The plate is placed on a dining table, and the vegetables are the main focus of the scene.

$\mathbf{Q}^{(2)}$ The image features a white plate filled with a variety of broccoli and cauliflower pieces. The broccoli pieces are scattered across the plate, with some pieces being larger and more prominent than others. The cauliflower pieces are also mixed in with the broccoli, creating a visually appealing and colorful display. The plate is placed on a wooden dining table, which serves as the background for the image. The combination of the vibrant colors and the contrasting textures of the vegetables make the dish look appetizing and inviting.

$\mathbf{Q}^{(3)}$ The image features a wooden dining table with a large, colorful plate of vegetables placed on it. The plate is filled with a variety of vegetables, including several pieces of broccoli, carrots, and cauliflower. The broccoli pieces are scattered across the plate, with some located closer to the center and others towards the edges. The carrots are also spread out, with some pieces positioned near the broccoli and others closer to the edge of the plate. The cauliflower is found in the middle of the plate, surrounded by the other vegetables. In addition to the plate, there is a fork and a knife placed on the table, ready for use. The fork is located on the left side of the plate, while the knife is positioned on the right side. The arrangement of the vegetables, the utensils, and the wooden table create a visually appealing and inviting dining scene.

$\mathbf{Q}^{(4)}$ The image features a wooden dining table with a large, colorful plate of food placed on top. The plate is filled with a variety of vegetables, including several pieces of broccoli, carrots, and cauliflower. The broccoli is spread across the plate, with some pieces located near the center and others towards the edges. The carrots are scattered throughout the plate, with some pieces closer to the center and others near the edges. The cauliflower is also spread across the plate, with some pieces near the center and others towards the edges. In addition to the vegetables, there is a bowl placed on the table, located towards the right side of the plate. A knife can be seen on the left side of the table, and a fork is positioned on the right side. The table setting creates an inviting atmosphere for a healthy and delicious meal.

$\mathbf{Q}^{(5)}$ The image features a wooden dining table with a large, colorful plate filled with a variety of vegetables. The plate is the main focus of the image, showcasing a diverse assortment of vegetables, including broccoli, carrots, and cauliflower. The vegetables are arranged in an appetizing manner, with some pieces overlapping and others placed next to each other. In addition to the plate, there are several bowls placed around the table. One bowl is located on the left side of the table, another is situated on the right side, and a third bowl is placed at the top of the table. A knife can be seen on the right side of the table, likely used for cutting and serving the vegetables. The overall scene is a vibrant and healthy display of vegetables, inviting viewers to enjoy a nutritious meal.

$\mathbf{Q}^{(1)}$ The image features a large bowl filled with a variety of vegetables, including several pieces of broccoli. The broccoli is spread throughout the bowl, with some pieces located closer to the front, while others are positioned further back. The bowl is placed on a dining table, which occupies the majority of the image. The table appears to be a part of a larger dining setup, with chairs visible in the background. The overall scene conveys a sense of abundance and freshness, as the vegetables are presented in a visually appealing manner.

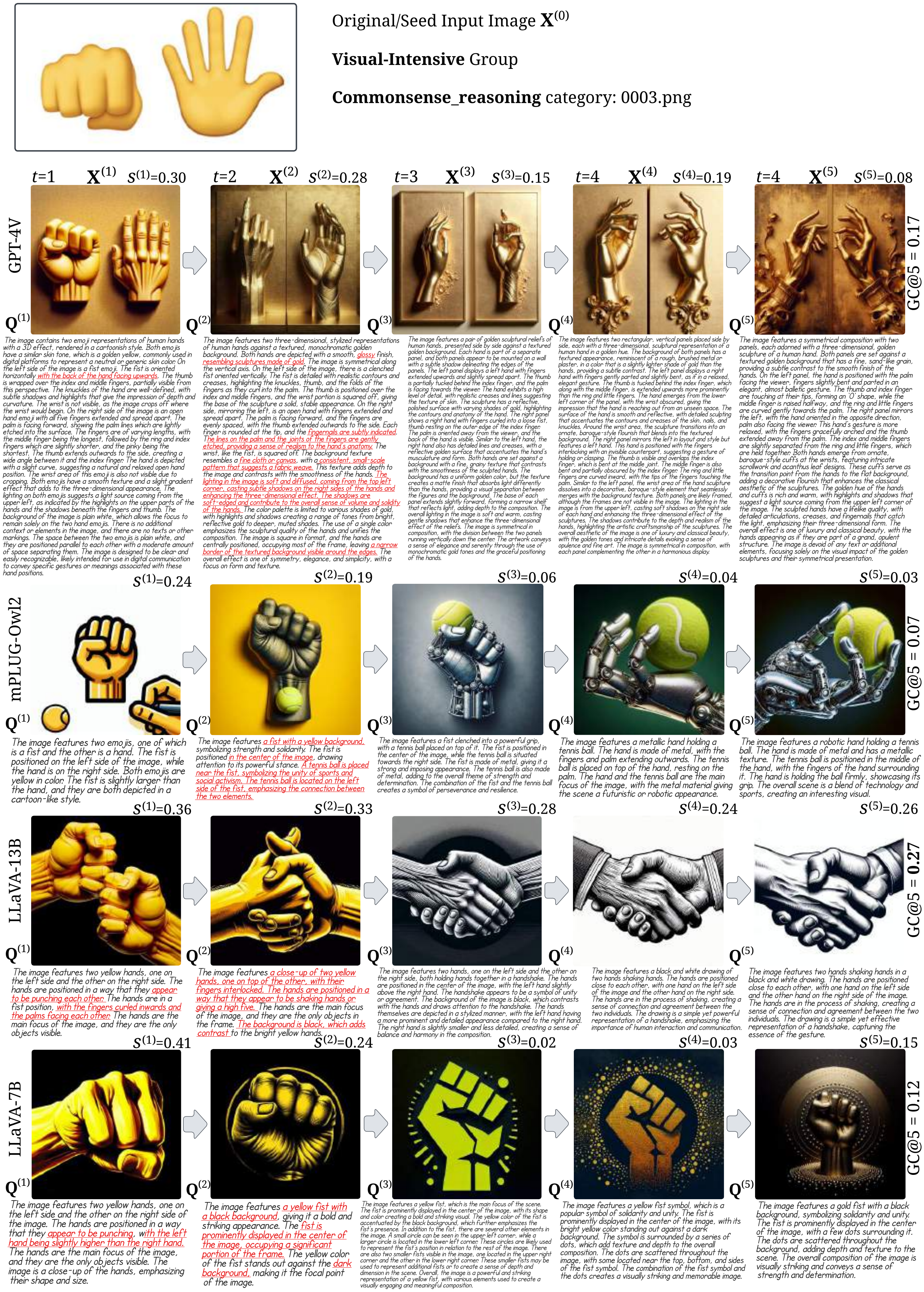
$\mathbf{Q}^{(2)}$ The image features a large wooden dining table with a variety of food items placed on it. The table is surrounded by numerous chairs, some of which are placed around the table, while others are positioned at the sides. The table is filled with a large assortment of broccoli, which is the main focus of the scene. In addition to the broccoli, there are several carrots scattered across the table, adding to the colorful and appetizing display. The table is set for a meal, with multiple wine glasses placed around the table, and a bowl is also visible on the table. The overall scene is inviting and ready for a group of people to gather and enjoy a meal together.

$\mathbf{Q}^{(3)}$ The image features a large dining table with a unique centerpiece: a pile of broccoli. The broccoli is arranged in a mountain-like formation, covering the entire table. The table is surrounded by numerous chairs, some of which are placed around the table, while others are positioned at the sides. In addition to the broccoli, there are several wine glasses scattered across the table, with some placed closer to the center and others near the edges. A bowl can also be seen on the table, likely containing additional food or condiments. The overall scene is a creative and unconventional take on a dining setup.

$\mathbf{Q}^{(4)}$ The image features a large dining table with a broccoli-shaped centerpiece, surrounded by several chairs. The table is set with various dishes, including multiple bowls, cups, and a knife. The broccoli centerpiece is the focal point of the table, and it is surrounded by an assortment of broccoli pieces, creating a visually appealing and healthy meal. The chairs are placed around the table, inviting guests to sit and enjoy the delicious meal.

$\mathbf{Q}^{(5)}$ The image features a large dining table set with numerous plates, bowls, and cups. The table is adorned with a variety of broccoli, which is the main focus of the scene. The broccoli is scattered across the table, with some pieces placed closer to the edges and others more centrally located. In addition to the broccoli, there are several cups and bowls placed on the table, with some cups positioned near the edges and others closer to the center. The table is also set with a fork and a knife, which are placed near the edge of the table. The overall scene appears to be a well-prepared dining table, ready for a meal with an abundance of broccoli as the main course.

Figure 3: Illustration of a 5-iteration GenCeption procedure run on color images to evaluate 4 VLLMs. Each iteration t shows the generated image $\mathbf{X}^{(t)}$, the description $\mathbf{Q}^{(t)}$ of the preceding image $\mathbf{X}^{(t-1)}$, and the similarity score $s^{(t)}$ relative to $\mathbf{X}^{(0)}$. The GC@5 metric for each VLLM is also presented. Hallucinated elements within descriptions $\mathbf{Q}^{(1)}$ and $\mathbf{Q}^{(2)}$ as compared to the seed image are indicated with red underlining.



The image contains two emoji representations of human hands with a 3D effect, rendered in a cartoonish style. Both emojis have a similar skin tone, which is a golden yellow, commonly used in digital platforms to represent a neutral or generic skin color. On the left side of the image is a fist emoji. The fist is oriented horizontally with the back of the hand facing upwards. The thumb is wrapped over the index and middle fingers, partially visible from this perspective. The knuckles of the hand are well-defined, with subtle shadows and highlights that give the impression of depth and curvature. The wrist is not visible, as the image crops off where the wrist would begin. On the right side of the image is an open hand emoji with all five fingers extended and spread apart. The palm is facing forward, showing the palm lines which are lightly etched into the surface. The fingers are of varying lengths, with the middle finger being the longest, followed by the ring and index fingers which are slightly shorter, and the pinky being the shortest. The thumb extends outwards to the side, creating a wide angle between it and the index finger. The hand is depicted with a slight curve, suggesting a natural and relaxed open hand position. The wrist area of this emoji is also not visible due to cropping. Both emojis have a smooth texture and a slight gradient effect that adds to the three-dimensional appearance. The lighting on both emojis suggests a light source coming from the upper left, as indicated by the highlights on the upper parts of the hands and the shadows beneath the fingers and thumb. The background of the image is plain white, which allows the focus to remain solely on the two hand emojis. There is no additional context or elements in the image, and there are no texts or other markings. The space between the two emojis is plain white, and they are positioned parallel to each other with a moderate amount of space separating them. The image is designed to be clear and easily recognizable, likely intended for use in digital communication to convey specific gestures or meanings associated with these hand positions.

The image features two three-dimensional, stylized representations of human hands against a textured, monochromatic golden background. Both hands are depicted with a smooth, glossy finish, resembling sculptures made of gold. The image is symmetrical along the vertical axis. On the left side of the image, there is a clenched fist oriented vertically. The fist is detailed with realistic contours and creases, highlighting the knuckles, thumb, and the folds of the fingers as they curl into the palm. The thumb is positioned over the index and middle fingers, and the wrist portion is squared off, giving the base of the sculpture a solid, stable appearance. On the right side, mirroring the left, is an open hand with fingers extended and spread apart. The palm is facing forward, and the fingers are evenly spaced, with the thumb extended outwards to the side. Each finger is rounded at the tip, and the fingernails are subtly indicated. The lines on the palm and the joints of the fingers are gently etched, providing a sense of realism to the hand's anatomy. The wrist, like the fist, is squared off. The background texture resembles a fine cloth or canvas, with a consistent, small-scale pattern that suggests a fabric weave. This texture adds depth to the image and contrasts with the smoothness of the hands. The lighting in the image is soft and diffused, coming from the top left corner, casting subtle shadows on the right sides of the hands and enhancing the three-dimensional effect. The shadows are soft-edged and contribute to the overall sense of volume and solidity of the hands. The color palette is limited to various shades of gold, with highlights and shadows creating a range of tones from bright, reflective gold to deeper, muted shades. The use of a single color emphasizes the sculptural quality of the hands and unifies the composition. The image is square in format, and the hands are centrally positioned, occupying most of the frame, leaving a narrow border of the textured background visible around the edges. The overall effect is one of symmetry, elegance, and simplicity, with a focus on form and texture.

The image features a pair of golden sculptural reliefs of human hands, presented side by side against a textured golden background. Each hand is part of a separate panel, and both panels appear to be mounted on a wall with a subtle shadow delineating the edges of the panels. The left panel displays a left hand with fingers extended upwards and slightly spread apart. The thumb is partially tucked behind the index finger, and the palm is facing towards the viewer. The hand exhibits a high level of detail, with realistic creases and lines suggesting the texture of skin. The sculpture has a reflective, polished surface with varying shades of gold, highlighting the contours and anatomy of the hand. The right panel shows a right hand with fingers curled into a loose fist, thumb resting on the outer edge of the index finger. The palm is oriented away from the viewer, and the back of the hand is visible. Similar to the left hand, the right hand also has detailed lines and creases, with a reflective golden surface that accentuates the hand's musculature and form. Both hands are set against a background with a fine, grainy texture that contrasts with the smoothness of the sculpted hands. The background has a uniform golden color, but the texture creates a matte finish that absorbs light differently than the hands, providing a visual separation between the figures and the background. The base of each panel extends slightly forward, forming a narrow shelf that reflects light, adding depth to the composition. The overall lighting in the image is soft and warm, casting gentle shadows that enhance the three-dimensional effect of the reliefs. The image is symmetrical in composition, with the division between the two panels running vertically down the center. The artwork conveys a sense of elegance and serenity through the use of monochromatic gold tones and the graceful positioning of the hands.

The image features two rectangular, vertical panels placed side by side, each with a three-dimensional, sculptural representation of a human hand in a golden hue. The background of both panels has a textured appearance, reminiscent of a rough, brushed metal or plaster, in a color that is a slightly lighter shade of gold than the hands, providing a subtle contrast. The left panel displays a right hand with fingers gently parted and slightly bent, as if in a relaxed, elegant gesture. The thumb is tucked behind the index finger, which along with the middle finger, is extended upwards more prominently than the ring and little fingers. The hand emerges from the lower left corner of the panel, with the wrist obscured, giving the impression that the hand is reaching out from an unseen space. The surface of the hand is smooth and reflective, with detailed sculpting that accentuates the contours and creases of the skin, nails, and knuckles. Around the wrist area, the sculpture transitions into an ornate, baroque-style flourish that blends into the textured background. The right panel mirrors the left in layout and style but features a left hand. This hand is positioned with the fingers interlocking with an invisible counterpart, suggesting a gesture of holding or clasping. The thumb is visible and overlaps the index finger, which is bent at the middle joint. The middle finger is also bent and partially obscured by the index finger. The ring and little fingers are curved inward, with the tips of the fingers touching the palm. Similar to the left panel, the wrist area of the hand sculpture dissolves into a decorative, baroque-style element that seamlessly merges with the background texture. Both panels are likely framed, although the frames are not visible in the image. The lighting in the image is from the upper left, casting soft shadows on the right side of each hand and enhancing the three-dimensional effect of the sculptures. The shadows contribute to the depth and realism of the hands, highlighting the artistic craftsmanship of the sculptures. The overall aesthetic of the image is one of luxury and classical beauty, with the golden tones and intricate details evoking a sense of opulence and fine art. The image is symmetrical in composition, with each panel complementing the other in a harmonious display.

The image features a symmetrical composition with two panels, each adorned with a three-dimensional, golden sculpture of a human hand. Both panels are set against a textured golden background that has a fine, sand-like grain, providing a subtle contrast to the smooth finish of the hands. On the left panel, the hand is positioned with the palm facing the viewer, fingers slightly bent and parted in an elegant, almost balletic gesture. The thumb and index finger are touching at their tips, forming an 'O' shape, while the middle finger is raised halfway, and the ring and little fingers are curved gently towards the palm. The right panel mirrors the left, with the hand oriented in the opposite direction, palm also facing the viewer. This hand's gesture is more relaxed, with the fingers gracefully arched and the thumb extended away from the palm. The index and middle fingers are slightly separated from the ring and little fingers, which are held together. Both hands emerge from ornate, baroque-style cuffs at the wrists, featuring intricate scrollwork and acanthus leaf designs. These cuffs serve as the transition point from the hands to the flat background, adding a decorative flourish that enhances the classical aesthetic of the sculptures. The golden hue of the hands and cuffs is rich and warm, with highlights and shadows that suggest a light source coming from the upper left corner of the image. The sculpted hands have a lifelike quality, with detailed articulations, creases, and fingernails that catch the light, emphasizing their three-dimensional form. The overall effect is one of luxury and classical beauty, with the hands appearing as if they are part of a grand, opulent structure. The image is devoid of any text or additional elements, focusing solely on the visual impact of the golden sculptures and their symmetrical presentation.

The image features two emojis, one of which is a fist and the other is a hand. The fist is positioned on the left side of the image, while the hand is on the right side. Both emojis are yellow in color. The fist is slightly larger than the hand, and they are both depicted in a cartoon-like style.

The image features a fist with a yellow background, symbolizing strength and solidarity. The fist is positioned in the center of the image, drawing attention to its powerful stance. A tennis ball is placed near the fist, symbolizing the unity of sports and social activism. The tennis ball is located on the left side of the fist, emphasizing the connection between the two elements.

The image features a fist clenched into a powerful grip, with a tennis ball placed on top of it. The fist is positioned in the center of the image, while the tennis ball is situated towards the right side. The fist is made of metal, giving it a strong and imposing appearance. The tennis ball is also made of metal, adding to the overall theme of strength and determination. The combination of the fist and the tennis ball creates a symbol of perseverance and resilience.

The image features a metallic hand holding a tennis ball. The hand is made of metal, with the fingers and palm extending outwards. The tennis ball is placed on top of the hand, resting on the palm. The hand and the tennis ball are the main focus of the image, with the metal material giving the scene a futuristic or robotic appearance.

The image features a robotic hand holding a tennis ball. The hand is made of metal and has a metallic texture. The tennis ball is positioned in the middle of the hand, with the fingers of the hand surrounding it. The hand is holding the ball firmly, showcasing its grip. The overall scene is a blend of technology and sports, creating an interesting visual.

The image features two yellow hands, one on the left side and the other on the right side. The hands are positioned in a way that they appear to be punching each other. The hands are in a fist position, with the fingers curled inwards and the palms facing each other. The hands are the main focus of the image, and they are the only objects visible.

The image features a close-up of two yellow hands, one on top of the other, with their fingers interlocked. The hands are positioned in a way that they appear to be shaking hands or giving a high five. The hands are the main focus of the image, and they are the only objects in the frame. The background is black, which adds contrast to the bright yellow hands.

The image features two hands, one on the left side and the other on the right side, both holding hands together in a handshake. The hands are positioned in the center of the image, with the left hand slightly above the right hand. The handshake appears to be a symbol of unity or agreement. The background of the image is black, which contrasts with the hands and draws attention to the handshake. The hands themselves are depicted in a stylized manner, with the left hand having a more prominent and detailed appearance compared to the right hand. The right hand is slightly smaller and less detailed, creating a sense of balance and harmony in the composition.

The image features a black and white drawing of two hands shaking hands. The hands are positioned close to each other, with one hand on the left side of the image and the other hand on the right side. The hands are in the process of shaking, creating a sense of connection and agreement between the two individuals. The drawing is a simple yet powerful representation of a handshake, emphasizing the importance of human interaction and communication.

The image features two hands shaking hands in a black and white drawing. The hands are positioned close to each other, with one hand on the left side and the other hand on the right side of the image. The hands are in the process of shaking, creating a sense of connection and agreement between the two individuals. The drawing is a simple yet effective representation of a handshake, capturing the essence of the gesture.

The image features two yellow hands, one on the left side and the other on the right side of the image. The hands are positioned in a way that they appear to be punching, with the left hand being slightly higher than the right hand. The hands are the main focus of the image, and they are the only objects visible. The image is a close-up of the hands, emphasizing their shape and size.

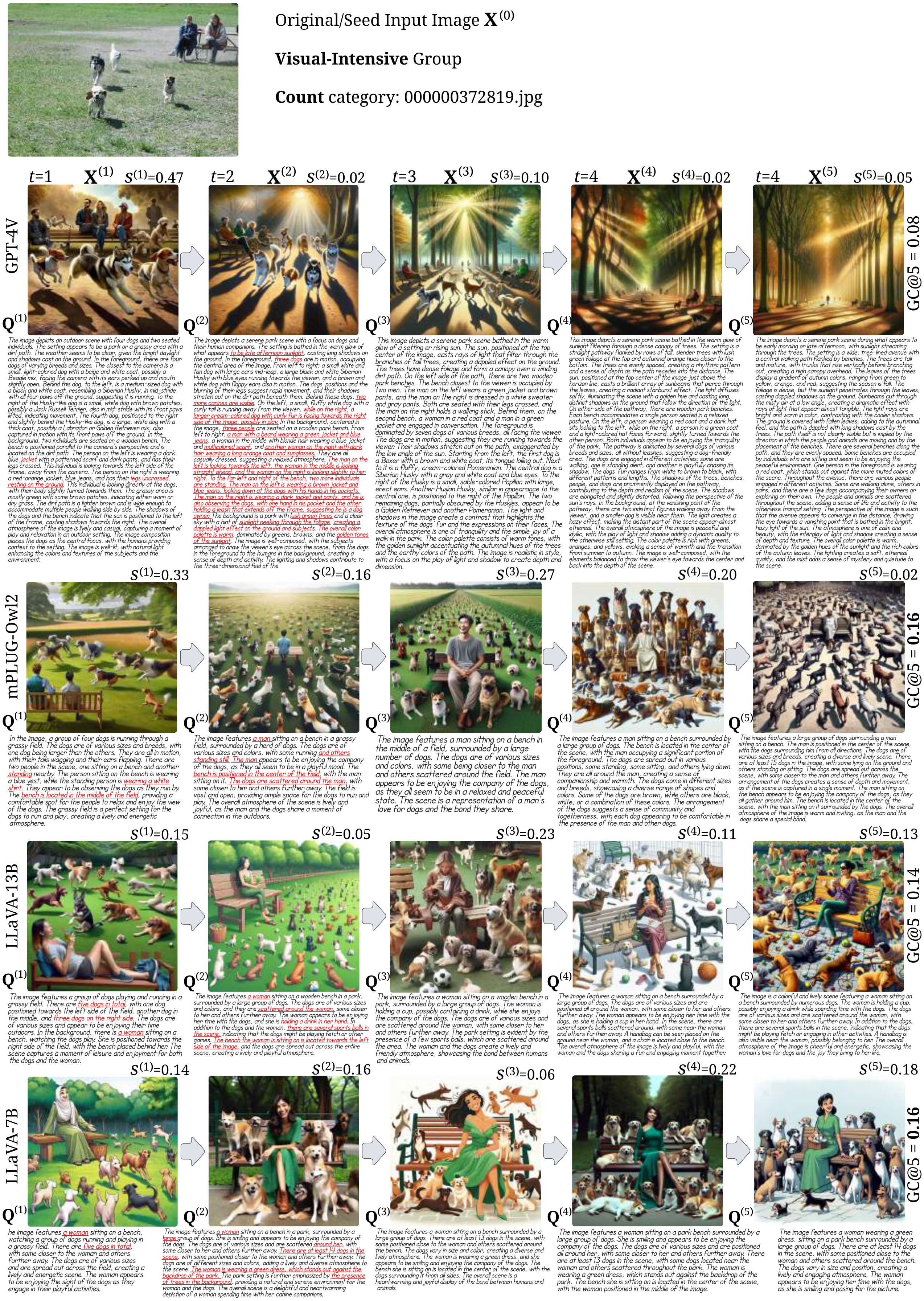
The image features a yellow fist with a black background, giving it a bold and striking appearance. The fist is prominently displayed in the center of the image, occupying a significant portion of the frame. The yellow color of the fist stands out against the dark background, making it the focal point of the image.

The image features a yellow fist, which is the main focus of the scene. The fist is prominently displayed in the center of the image, with its shape and color creating a bold and striking visual. The yellow color of the fist is accentuated by the black background, which further emphasizes the fist's presence. In addition to the fist, there are several other elements in the image. A small circle can be seen in the upper left corner, while a larger circle is located in the lower left corner. These circles are likely used to represent the fist's position in relation to the rest of the image. There are also two smaller fists visible in the image, one located in the upper right corner and the other in the lower right corner. These smaller fists may be used to represent additional fists or to create a sense of depth and dimension in the scene. Overall, the image is a powerful and striking representation of a yellow fist, with various elements used to create a visually engaging and meaningful composition.

The image features a yellow fist symbol, which is a popular symbol of solidarity and unity. The fist is prominently displayed in the center of the image, with its bright yellow color standing out against a dark background. The symbol is surrounded by a series of dots, which add texture and depth to the overall composition. The dots are scattered throughout the image, with some located near the top, bottom, and sides of the fist symbol. The combination of the fist symbol and the dots creates a visually striking and memorable image.

The image features a gold fist with a black background, symbolizing solidarity and unity. The fist is prominently displayed in the center of the image, with a few dots surrounding it. The dots are scattered throughout the background, adding depth and texture to the scene. The overall composition of the image is visually striking and conveys a sense of strength and determination.

Figure 4: Illustration of a 5-iteration GenCeption procedure run on commonsense_reasoning images to evaluate 4 VLLMs. Each iteration t shows the generated image $\mathbf{X}^{(t)}$, the description $\mathbf{Q}^{(t)}$ of the preceding image $\mathbf{X}^{(t-1)}$, and the similarity score $s^{(t)}$ relative to $\mathbf{X}^{(0)}$. The GC@5 metric for each VLLM is also presented. Hallucinated elements within descriptions $\mathbf{Q}^{(1)}$ and $\mathbf{Q}^{(2)}$ as compared to the seed image are indicated with red underlining.



The image depicts an outdoor scene with four dogs and two seated individuals. The setting appears to be a park or a grassy area with a dirt path. The weather seems to be clear, given the bright daylight and shadows cast on the ground. In the foreground, there are four dogs of varying breeds and sizes. The closest to the camera is a small, light-colored dog with a beige and white coat, possibly a Beagle mix, facing the camera with its ears perked up and mouth slightly open. Behind this dog, to the left, is a medium-sized dog with a black and white coat, resembling a Siberian Husky, in mid-stride with all four paws off the ground, suggesting it is running. To the right of the Husky-like dog is a small, white dog with brown patches, possibly a Jack Russell Terrier, also in mid-stride with its front paws lifted, indicating movement. The fourth dog, positioned to the right and slightly behind the Husky-like dog, is a large, white dog with a thick coat, possibly a Labrador or Golden Retriever mix, also captured in motion with its front paws off the ground. In the background, two individuals are seated on a wooden bench. The bench is positioned parallel to the camera's perspective and is located on the dirt path. The person on the left is wearing a dark blue jacket with a patterned scarf and dark pants, and has their legs crossed. This individual is looking towards the left side of the frame, away from the camera. The person on the right is wearing a red-orange jacket, blue jeans, and has their legs uncrossed, resting on the ground. This individual is looking directly at the dogs, with their body slightly turned towards them. The grassy area is mostly green with some brown patches, indicating either worn or dry grass. The dirt path is a lighter brown and is wide enough to accommodate multiple people walking side by side. The shadows of the dogs and the bench indicate that the sun is positioned to the left of the frame, casting shadows towards the right. The overall atmosphere of the image is lively and casual, capturing a moment of play and relaxation in an outdoor setting. The image composition places the dogs as the central focus, with the humans providing context to the setting. The image is well-lit, with natural light enhancing the colors and textures of the subjects and the environment.

The image depicts a serene park scene with a focus on dogs and their human companions. The setting is bathed in the warm glow of what appears to be late afternoon sunlight, casting long shadows on the ground. In the foreground, three dogs are in motion, occupying the central area of the image. From left to right: a small white and tan dog with large ears mid-leap, a large black and white Siberian Husky with blue eyes running towards the viewer, and a brown and white dog with floppy ears also in motion. The dogs' positions and the blurring of their legs suggest rapid movement, and their shadows stretch out on the dirt path beneath them. Behind these dogs, two more canines are visible. On the left, a small, fluffy white dog with a curly tail is running away from the viewer, while on the right, a larger cream-colored dog with curly fur is facing towards the right side of the image, possibly in play. In the background, centered in the image, three people are seated on a wooden park bench. From left to right: a man with a beard wearing a green jacket and blue jeans, a woman in the middle with blonde hair wearing a blue jacket and multicolored scarf, and another woman on the right with dark hair wearing a long orange coat and sunglasses. They are all casually dressed, suggesting a relaxed atmosphere. The man on the left is looking towards the left, the woman in the middle is looking straight ahead, and the woman on the right is looking slightly to her right. To the far left and right of the bench, two more individuals are standing. The man on the left is wearing a brown jacket and blue jeans, looking down at the dogs with his hands in his pockets. The man on the right is wearing a dark jacket and pants, and he is also observing the dogs, with one hand in his pocket and the other holding a leash that extends off the frame, suggesting he is a dog owner. The background is a park with lush green trees and a clear sky with a hint of sunlight peeking through the foliage, creating a dappled light effect on the ground and subjects. The overall color palette is warm, dominated by greens, browns, and the golden tones of the sunlight. The image is well-composed, with the subjects arranged to draw the viewer's eye across the scene, from the dogs in the foreground to the humans in the background, creating a sense of depth and activity. The lighting and shadows contribute to the three-dimensional feel of the

This image depicts a serene park scene bathed in the warm glow of a setting or rising sun. The sun, positioned at the top center of the image, casts rays of light that filter through the branches of tall trees, creating a dappled effect on the ground. The trees have dense foliage and form a canopy over a winding dirt path. On the left side of the path, there are two wooden park benches. The bench closest to the viewer is occupied by two men. The man on the left wears a green jacket and brown pants, and the man on the right is dressed in a white sweater and gray pants. Both are seated with their legs crossed, and the man on the right holds a walking stick. Behind them, on the second bench, a woman in a red coat and a man in a green jacket are engaged in conversation. The foreground is dominated by seven dogs of various breeds, all facing the viewer. The dogs are in motion, suggesting they are running towards the viewer. Their shadows stretch out on the path, exaggerated by the low angle of the sun. Starting from the left, the first dog is a Boxer with a brown and white coat, its tongue lolling out. Next to it is a fluffy, cream-colored Pomeranian. The central dog is a Siberian Husky with a gray and white coat and blue eyes. To the right of the Husky is a small, sable-colored Papillon with large, erect ears. Another Husian Husky, similar in appearance to the central one, is positioned to the right of the Papillon. The two remaining dogs, partially obscured by the Huskies, appear to be a Golden Retriever and another Pomeranian. The light and shadows in the image create a contrast that highlights the texture of the dogs' fur and the expressions on their faces. The overall atmosphere is one of tranquility and the simple joy of a walk in the park. The color palette consists of warm tones, with the golden sunlight accentuating the autumnal hues of the trees and the earthy colors of the path. The image is realistic in style, with a focus on the play of light and shadow to create depth and dimension.

This image depicts a serene park scene bathed in the warm glow of sunlight filtering through a dense canopy of trees. The setting is a straight pathway flanked by rows of tall, slender trees with lush green foliage at the top and autumnal orange hues closer to the bottom. The trees are evenly spaced, creating a rhythmic pattern and a sense of depth as the path recedes into the distance. The sun, positioned at the top center of the image just above the horizon line, casts a brilliant array of sunbeams that pierce through the leaves, creating a radiant starburst effect. The light diffuses softly, illuminating the scene with a golden hue and casting long, distinct shadows on the ground that follow the direction of the light. On either side of the pathway, there are wooden park benches. Each bench accommodates a single person seated in a relaxed posture. On the left, a person wearing a red coat and a dark hat sits looking to the left, while on the right, a person in a green coat and a light-colored hat faces forward, slightly turned towards the other person. Both individuals appear to be enjoying the tranquility of the park. The pathway is animated by several dogs of various breeds and sizes, all without leashes, suggesting a dog-friendly area. The dogs are engaged in different activities: some are walking, one is standing alert, and another is playfully chasing its shadow. The dogs' fur ranges from white to brown to black, with different patterns and lengths. The shadows of the trees, benches, people, and dogs are prominently displayed on the pathway, contributing to the depth and realism of the scene. The shadows are elongated and slightly distorted, following the perspective of the sun's rays. In the background, at the vanishing point of the pathway, there are two indistinct figures walking away from the viewer, and a smaller dog is visible near them. The light creates a hazy effect, making the distant part of the scene appear almost ethereal. The overall atmosphere of the image is peaceful and idyllic, with the play of light and shadow adding a dynamic quality to the otherwise still setting. The color palette is rich with greens, oranges, and yellows, evoking a sense of warmth and the transition from summer to autumn. The image is well-composed, with the elements balanced to draw the viewer's eye towards the center and back into the depth of the scene.

The image depicts a serene park scene during what appears to be early morning or late afternoon, with sunlight streaming through the trees. The setting is a wide, tree-lined avenue with a central walking path flanked by benches. The trees are tall and mature, with trunks that rise vertically before branching out, creating a high canopy overhead. The leaves of the trees display a gradient of autumn colors, ranging from green to yellow, orange, and red, suggesting the season is fall. The foliage is dense, but the sunlight penetrates through the leaves, casting dappled shadows on the ground. Sunbeams cut through the misty air at a low angle, creating a dramatic effect with rays of light that appear almost tangible. The light rays are bright and warm in color, contrasting with the cooler shadows. The ground is covered with fallen leaves, adding to the autumnal feel, and the path is dappled with long shadows cast by the trees. The path itself is not clearly visible but is implied by the direction in which the people and animals are moving and by the placement of the benches. There are several benches along the path, and they are evenly spaced. Some benches are occupied by individuals who are sitting and seem to be enjoying the peaceful environment. One person in the foreground is wearing a red coat, which stands out against the more muted colors of the scene. Throughout the avenue, there are various people engaged in different activities. Some are walking alone, others in pairs, and there are a few dogs accompanying their owners or exploring on their own. The people and animals are scattered throughout the scene, adding a sense of life and activity to the otherwise tranquil setting. The perspective of the image is such that the avenue appears to converge in the distance, drawing the eye towards a vanishing point that is bathed in the bright, hazy light of the sun. The atmosphere is one of calm and beauty, with the interplay of light and shadow creating a sense of depth and texture. The overall color palette is warm, dominated by the golden hues of the sunlight and the rich colors of the autumn leaves. The lighting creates a soft, ethereal quality, and the mist adds a sense of mystery and quietude to the scene.

In the image, a group of four dogs is running through a grassy field. The dogs are of various sizes and breeds, with one dog being larger than the others. They are all in motion, with their tails wagging and their ears flapping. There are two people in the scene, one sitting on a bench and another standing nearby. The person sitting on the bench is wearing a blue vest, while the standing person is wearing a white shirt. They appear to be observing the dogs as they run by. The bench is located in the middle of the field, providing a comfortable spot for the people to relax and enjoy the view of the dogs. The grassy field is a perfect setting for the dogs to run and play, creating a lively and energetic atmosphere.

The image features a man sitting on a bench in a grassy field, surrounded by a herd of dogs. The dogs are of various sizes and colors, with some running and others standing still. The man appears to be enjoying the company of the dogs, as they all seem to be in a playful mood. The bench is positioned in the center of the field, with the man sitting on it. The dogs are scattered around the man, with some closer to him and others further away. The field is vast and open, providing ample space for the dogs to run and play. The overall atmosphere of the scene is lively and joyful, as the man and the dogs share a moment of connection in the outdoors.

The image features a man sitting on a bench in the middle of a field, surrounded by a large number of dogs. The dogs are of various sizes and colors, with some being closer to the man and others scattered around the field. The man appears to be enjoying the company of the dogs, as they all seem to be in a relaxed and peaceful state. The scene is a representation of a man's love for dogs and the bond they share.

The image features a man sitting on a bench surrounded by a large group of dogs. The bench is located in the center of the scene, with the man occupying a significant portion of the foreground. The dogs are spread out in various positions, some standing, some sitting, and others lying down. They are all around the man, creating a sense of companionship and warmth. The dogs come in different sizes and breeds, showcasing a diverse range of shapes and colors. Some of the dogs are brown, while others are black, white, or a combination of these colors. The arrangement of the dogs suggests a sense of community and togetherness, with each dog appearing to be comfortable in the presence of the man and other dogs.

The image features a large group of dogs surrounding a man sitting on a bench. The man is positioned in the center of the scene, with the dogs surrounding him from all directions. The dogs are of various sizes and breeds, creating a diverse and lively scene. There are at least 15 dogs in the image, with some lying on the ground and others standing or sitting. The dogs are spread out across the scene, with some closer to the man and others further away. The arrangement of the dogs creates a sense of depth and movement, as if the scene is captured in a single moment. The man sitting on the bench appears to be enjoying the company of the dogs, as they all gather around him. The bench is located in the center of the scene, with the man sitting on it surrounded by the dogs. The overall atmosphere of the image is warm and inviting, as the man and the dogs share a special bond.

The image features a group of dogs playing and running in a grassy field. There are five dogs in total, with one dog positioned towards the left side of the field, another dog in the middle, and three dogs on the right side. The dogs are of various sizes and appear to be enjoying their time outdoors. In the background, there is a woman sitting on a bench, watching the dogs play. She is positioned towards the right side of the field, with the bench placed behind her. The scene captures a moment of leisure and enjoyment for both the dogs and the woman.

The image features a woman sitting on a wooden bench in a park, surrounded by a large group of dogs. The dogs are of various sizes and colors, and they are scattered around the woman, some closer to her and others further away. The woman appears to be enjoying her time with the dogs, and she is holding a drink in her hand. In addition to the dogs and the woman, there are several sports balls in the scene, indicating that the dogs might be playing fetch or other games. The bench the woman is sitting on is located towards the left side of the image, and the dogs are spread out across the entire scene, creating a lively and playful atmosphere.

The image features a woman sitting on a wooden bench in a park, surrounded by a large group of dogs. The woman is holding a cup, possibly containing a drink, while she enjoys the company of the dogs. The dogs are of various sizes and are scattered around the woman, with some closer to her and others further away. The park setting is evident by the presence of a few sports balls, which are scattered around the area. The woman and the dogs create a lively and friendly atmosphere, showcasing the bond between humans and animals.

The image features a woman sitting on a bench surrounded by a large group of dogs. The dogs are of various sizes and are positioned all around the woman, with some closer to her and others further away. The woman appears to be enjoying her time with the dogs, as she is holding a cup in her hand. In the scene, there are several sports balls scattered around, with some near the woman and others further away. A handbag can be seen placed on the ground near the woman, and a chair is located close to the bench. The overall atmosphere of the image is lively and playful, with the woman and the dogs sharing a fun and engaging moment together.

The image is a colorful and lively scene featuring a woman sitting on a bench surrounded by numerous dogs. The woman is holding a cup, possibly enjoying a drink while spending time with the dogs. The dogs are of various sizes and are scattered around the woman, with some closer to her and others further away. In addition to the dogs, there are several sports balls in the scene, indicating that the dogs might be playing fetch or engaging in other activities. A handbag is also visible near the woman, possibly belonging to her. The overall atmosphere of the image is cheerful and energetic, showcasing the woman's love for dogs and the joy they bring to her life.

he image features a woman sitting on a bench, watching a group of dogs running and playing in a grassy field. There are five dogs in total, with some closer to the woman and others further away. The dogs are of various sizes and are spread out across the field, creating a lively and energetic scene. The woman appears to be enjoying the sight of the dogs as they engage in their playful activities.

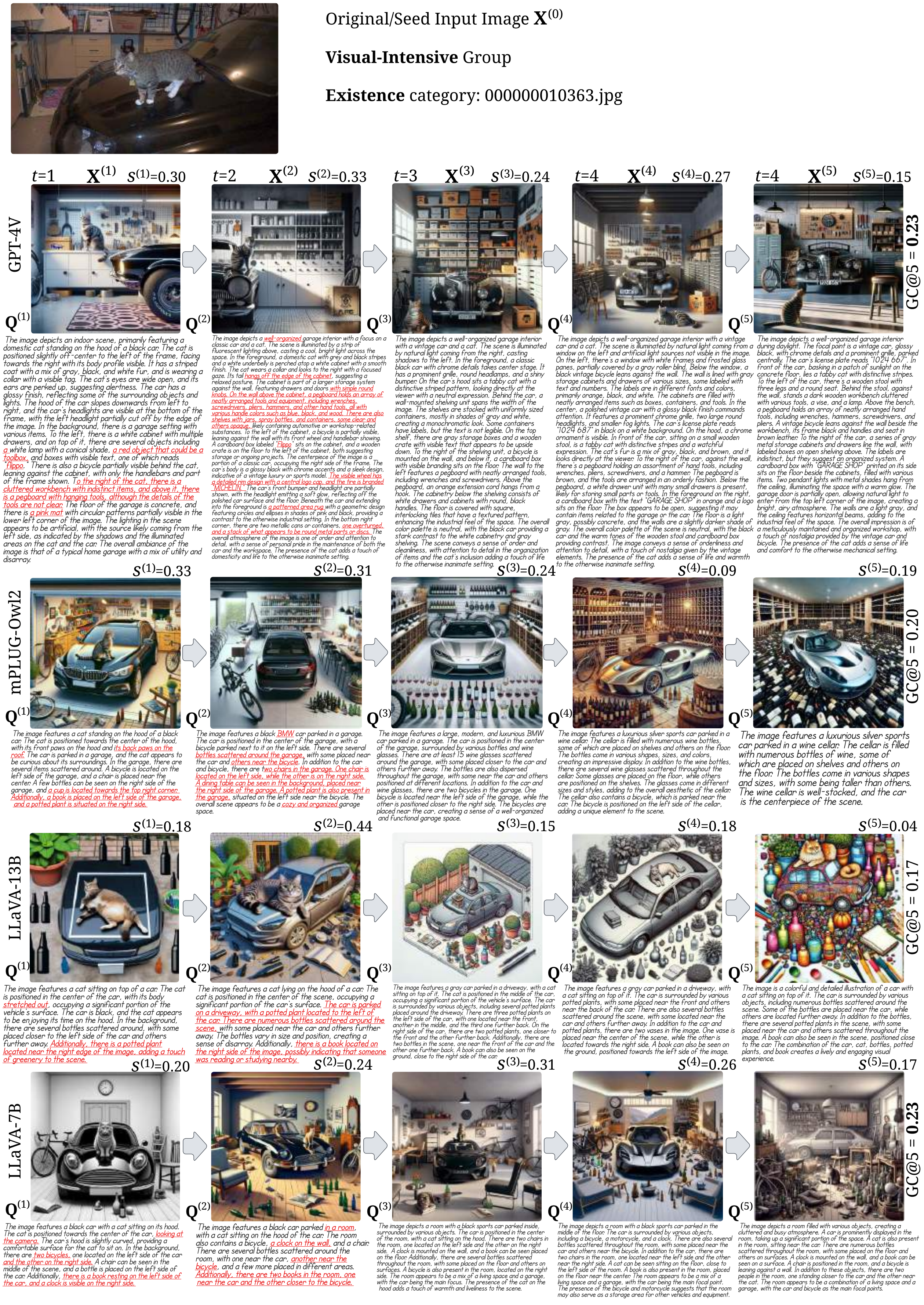
The image features a woman sitting on a bench in a park, surrounded by a large group of dogs. She is smiling and appears to be enjoying the company of the dogs. The dogs are of various sizes and are scattered around her, with some closer to her and others further away. There are at least 14 dogs in the scene, with some positioned closer to the woman and others further away. The dogs are of different sizes and colors, adding a lively and diverse atmosphere to the scene. The woman is wearing a green dress, which stands out against the backdrop of the park. The park setting is further emphasized by the presence of trees in the background, providing a natural and serene environment for the woman and the dogs. The overall scene is a delightful and heartwarming depiction of a woman spending time with her canine companions.

The image features a woman sitting on a bench surrounded by a large group of dogs. There are at least 13 dogs in the scene, with some positioned close to the woman and others scattered around the bench. The dogs vary in size and color, creating a diverse and lively atmosphere. The woman is wearing a green dress, and she appears to be smiling and enjoying the company of the dogs. The bench she is sitting on is located in the center of the scene, with the dogs surrounding it from all sides. The overall scene is a heartwarming and joyful display of the bond between humans and animals.

The image features a woman sitting on a park bench surrounded by a large group of dogs. She is smiling and appears to be enjoying the company of the dogs. The dogs are of various sizes and are positioned all around her, with some closer to her and others further away. There are at least 13 dogs in the scene, with some dogs located near the woman and others scattered throughout the park. The woman is wearing a green dress, which stands out against the backdrop of the park. The bench she is sitting on is located in the center of the scene, with the woman positioned in the middle of the image.

The image features a woman wearing a green dress, sitting on a park bench surrounded by a large group of dogs. There are at least 14 dogs in the scene, with some positioned close to the woman and others scattered around the bench. The dogs vary in size and position, creating a lively and engaging atmosphere. The woman appears to be enjoying her time with the dogs, as she is smiling and posing for the picture.

Figure 5: Illustration of a 5-iteration GenCeption procedure run on count images to evaluate 4 VLLMs. Each iteration t shows the generated image $\mathbf{X}^{(t)}$, the description $\mathbf{Q}^{(t)}$ of the preceding image $\mathbf{X}^{(t-1)}$, and the similarity score $s^{(t)}$ relative to $\mathbf{X}^{(0)}$. The GC@5 metric for each VLLM is also presented. Hallucinated elements within descriptions $\mathbf{Q}^{(1)}$ and $\mathbf{Q}^{(2)}$ as compared to the seed image are indicated with red underlining.



The image depicts an indoor scene, primarily featuring a domestic cat standing on the hood of a black car. The cat is positioned slightly off-center to the left of the frame, facing towards the right with its body profile visible. It has a striped coat with a mix of gray, black, and white fur, and is wearing a collar with a visible tag. The cat's eyes are wide open, and its ears are perked up, suggesting alertness. The car has a glossy finish, reflecting some of the surrounding objects and lights. The hood of the car slopes downwards from left to right, and the car's headlights are visible at the bottom of the frame, with the left headlight partially cut off by the edge of the image. In the background, there is a garage setting with various items. To the left, there is a white cabinet with multiple drawers, and on top of it, there are several objects including a white lamp with a conical shade, a red object that could be a toolbox, and boxes with visible text, one of which reads "flippo." There is also a bicycle partially visible behind the cat, leaning against the cabinet, with only the handlebars and part of the frame shown. To the right of the cat, there is a cluttered workbench with indistinct items, and above it, there is a pegboard with hanging tools, although the details of the tools are not clear. The floor of the garage is concrete, and there is a pink mat with circular patterns partially visible in the lower left corner of the image. The lighting in the scene appears to be artificial, with the source likely coming from the left side, as indicated by the shadows and the illuminated areas on the cat and the car. The overall ambiance of the image is that of a typical home garage with a mix of utility and disarray.

The image depicts a well-organized garage interior with a focus on a classic car and a cat. The scene is illuminated by a strip of fluorescent lighting above, casting a cool, bright light across the space. In the foreground, a domestic cat with grey and black stripes and a white underbelly is perched atop a white cabinet with a smooth finish. The cat wears a collar and looks to the right with a focused gaze. Its tail hangs off the edge of the cabinet, suggesting a relaxed posture. The cabinet is part of a larger storage system against the wall, featuring drawers and doors with simple round knobs. On the wall above the cabinet, a pegboard holds an array of neatly arranged tools and equipment, including wrenches, screwdrivers, pliers, hammers, and other hand tools, all with various handle colors such as blue, black, and wood. There are also shelves with jars, spray bottles, and containers, some clear and others opaque, likely containing automotive or workshop-related substances. To the left of the cabinet, a bicycle is partially visible, leaning against the wall with its front wheel and handlebar showing. A cardboard box labeled "Flippo" sits on the cabinet, and a wooden crate is on the floor to the left of the cabinet, both suggesting storage or ongoing projects. The centerpiece of the image is a portion of a classic car, occupying the right side of the frame. The car's body is a glossy black with chrome accents and a sleek design, indicative of a vintage luxury or sports model. The visible wheel has a detailed rim design with a central logo cap, and the tire is branded "MICHELIN." The car's front bumper and headlight are partially shown, with the headlight emitting a soft glow, reflecting off the polished car surface and the floor. Beneath the car and extending into the foreground is a patterned area rug with a geometric design featuring circles and ellipses in shades of pink and black, providing a contrast to the otherwise industrial setting. In the bottom right corner, there are two metallic cans or containers, one overturned, and a stack of what appears to be round metal parts or discs. The overall atmosphere of the image is one of order and attention to detail, with a sense of personal pride in the maintenance of both the car and the workspace. The presence of the cat adds a touch of domesticity and life to the otherwise inanimate setting.

The image depicts a well-organized garage interior with a vintage car and a cat. The scene is illuminated by natural light coming from the right, casting shadows to the left. In the foreground, a classic black car with chrome details takes center stage. It has a prominent grille, round headlamps, and a shiny bumper. On the car's hood sits a tabby cat with a distinctive striped pattern, looking directly at the viewer with a neutral expression. Behind the car, a wall-mounted shelving unit spans the width of the image. The shelves are stocked with uniformly sized containers, mostly in shades of gray and white, creating a monochromatic look. Some containers have labels, but the text is not legible. On the top shelf, there are gray storage boxes and a wooden crate with visible text that appears to be upside down. To the right of the shelving unit, a bicycle is mounted on the wall, and below it, a cardboard box with visible branding sits on the floor. The wall to the left features a pegboard with neatly arranged tools, including wrenches and screwdrivers. Above the pegboard, an orange extension cord hangs from a hook. The cabinetry below the shelving consists of white drawers and cabinets with round, black handles. The floor is covered with square, interlocking tiles that have a textured pattern, enhancing the industrial feel of the space. The overall color palette is neutral, with the black car providing a stark contrast to the white cabinetry and gray shelving. The scene conveys a sense of order and cleanliness, with attention to detail in the organization of items and the cat's inclusion adding a touch of life to the otherwise inanimate setting.

The image depicts a well-organized garage interior with a vintage car and a cat. The scene is illuminated by natural light coming from a window on the left and artificial light sources not visible in the image. On the left, there's a window with white frames and frosted glass panes, partially covered by a gray roller blind. Below the window, a black vintage bicycle leans against the wall. The wall is lined with gray storage cabinets and drawers of various sizes, some labeled with text and numbers. The labels are in different fonts and colors, primarily orange, black, and white. The cabinets are filled with neatly arranged items such as boxes, containers, and tools. In the center, a polished vintage car with a glossy black finish commands attention. It features a prominent chrome grille, two large round headlights, and smaller fog lights. The car's license plate reads "1024 687" in black on a white background. On the hood, a chrome ornament is visible. In front of the car, sitting on a small wooden stool, is a tabby cat with distinctive stripes and a watchful expression. The cat's fur is a mix of gray, black, and brown, and it looks directly at the viewer. To the right of the car, against the wall, there's a pegboard holding an assortment of hand tools, including wrenches, pliers, screwdrivers, and a hammer. The pegboard is brown, and the tools are arranged in an orderly fashion. Below the pegboard, a white drawer unit with many small drawers is present, likely for storing small parts or tools. In the foreground on the right, a cardboard box with the text "GARAGE SHOP" in orange and a logo sits on the floor. The box appears to be open, suggesting it may contain items related to the garage or the car. The floor is a light gray, possibly concrete, and the walls are a slightly darker shade of gray. The overall color palette of the scene is neutral, with the black car and the warm tones of the wooden stool and cardboard box providing contrast. The image conveys a sense of orderliness and attention to detail, with a touch of nostalgia given by the vintage elements. The presence of the cat adds a sense of life and warmth to the otherwise inanimate setting.

The image depicts a well-organized garage interior during daylight. The focal point is a vintage car, glossy black, with chrome details and a prominent grille, parked centrally. The car's license plate reads "1024 667". In front of the car, basking in a patch of sunlight on the concrete floor, lies a tabby cat with distinctive stripes. To the left of the car, there's a wooden stool with three legs and a round seat. Behind the stool, against the wall, stands a dark wooden workbench cluttered with various tools, a vise, and a lamp. Above the bench, a pegboard holds an array of neatly arranged hand tools, including wrenches, hammers, screwdrivers, and pliers. A vintage bicycle leans against the wall beside the workbench, its frame black and handles and seat in brown leather. To the right of the car, a series of gray metal storage cabinets and drawers line the wall, with labeled boxes on open shelving above. The labels are indistinct, but they suggest an organized system. A cardboard box with "GARAGE SHOP" printed on its side sits on the floor beside the cabinets, filled with various items. Two pendant lights with metal shades hang from the ceiling, illuminating the space with a warm glow. The garage door is partially open, allowing natural light to enter from the top left corner of the image, creating a bright, airy atmosphere. The walls are a light gray, and the ceiling features horizontal beams, adding to the industrial feel of the space. The overall impression is of a meticulously maintained and organized workshop, with a touch of nostalgia provided by the vintage car and bicycle. The presence of the cat adds a sense of life and comfort to the otherwise mechanical setting.

The image features a cat standing on the hood of a black car. The cat is positioned towards the center of the hood, with its front paws on the hood and its back paws on the roof. The car is parked in a garage, and the cat appears to be curious about its surroundings. In the garage, there are several items scattered around. A bicycle is located on the left side of the garage, and a chair is placed near the center. A few bottles can be seen on the right side of the garage, and a cup is located towards the top right corner. Additionally, a book is placed on the left side of the garage, and a potted plant is situated on the right side.

The image features a black BMW car parked in a garage. The car is positioned in the center of the garage, with a bicycle parked next to it on the left side. There are several bottles scattered around the garage, with some placed near the car and others near the bicycle. In addition to the car and bicycle, there are two chairs in the garage. One chair is located on the left side, while the other is on the right side. A dining table can be seen in the background, placed near the right side of the garage. A potted plant is also present in the garage, situated on the left side near the bicycle. The overall scene appears to be a cozy and organized garage space.

The image features a large, modern, and luxurious BMW car parked in a garage. The car is positioned in the center of the garage, surrounded by various bottles and wine glasses. There are at least 15 wine glasses scattered around the garage, with some placed closer to the car and others further away. The bottles are also dispersed throughout the garage, with some near the car and others positioned at different locations. In addition to the car and wine glasses, there are two bicycles in the garage. One bicycle is located near the left side of the garage, while the other is positioned closer to the right side. The bicycles are placed near the car, creating a sense of a well-organized and functional garage space.

The image features a luxurious silver sports car parked in a wine cellar. The cellar is filled with numerous wine bottles, some of which are placed on shelves and others on the floor. The bottles come in various shapes, sizes, and colors, creating an impressive display. In addition to the wine bottles, there are several wine glasses scattered throughout the cellar. Some glasses are placed on the floor, while others are positioned on the shelves. The glasses come in different sizes and styles, adding to the overall aesthetic of the cellar. The cellar also contains a bicycle, which is parked near the car. The bicycle is positioned on the left side of the cellar, adding a unique element to the scene.

The image features a luxurious silver sports car parked in a wine cellar. The cellar is filled with numerous bottles of wine, some of which are placed on shelves and others on the floor. The bottles come in various shapes and sizes, with some being taller than others. The wine cellar is well-stocked, and the car is the centerpiece of the scene.

The image features a cat sitting on top of a car. The cat is positioned in the center of the car, with its body stretched out, occupying a significant portion of the vehicle's surface. The car is black, and the cat appears to be enjoying its time on the hood. In the background, there are several bottles scattered around, with some placed closer to the left side of the car and others further away. Additionally, there is a potted plant located near the right edge of the image, adding a touch of greenery to the scene.

The image features a cat lying on the hood of a car. The cat is positioned in the center of the scene, occupying a significant portion of the car's surface. The car is parked on a driveway, with a potted plant located to the left of the car. There are numerous bottles scattered around the scene, with some placed near the car and others further away. The bottles vary in size and position, creating a sense of disarray. Additionally, there is a book located on the right side of the image, possibly indicating that someone was reading or studying nearby.

The image features a gray car parked in a driveway, with a cat sitting on top of it. The cat is positioned in the middle of the car, occupying a significant portion of the vehicle's surface. The car is surrounded by various objects, including several potted plants placed around the driveway. There are three potted plants on the left side of the car, with one located near the front, another in the middle, and the third one further back. On the right side of the car, there are two potted plants, one closer to the front and the other further back. Additionally, there are two bottles in the scene, one near the front of the car and the other one further back. A book can also be seen on the ground, close to the right side of the car.

The image features a gray car parked in a driveway, with a cat sitting on top of it. The car is surrounded by various potted plants, with some placed near the front and others near the back of the car. There are also several bottles scattered around the scene, with some located near the car and others further away. In addition to the car and potted plants, there are two vases in the image. One vase is placed near the center of the scene, while the other is located towards the right side. A book can also be seen on the ground, positioned towards the left side of the image.

The image is a colorful and detailed illustration of a car with a cat sitting on top of it. The car is surrounded by various objects, including numerous bottles scattered around the scene. Some of the bottles are placed near the car, while others are located further away. In addition to the bottles, there are several potted plants in the scene, with some placed near the car and others scattered throughout the image. A book can also be seen in the scene, positioned close to the car. The combination of the car, cat, bottles, potted plants, and book creates a lively and engaging visual experience.

The image features a black car with a cat sitting on its hood. The cat is positioned towards the center of the car, looking at the camera. The car's hood is slightly curved, providing a comfortable surface for the cat to sit on. In the background, there are two bicycles, one located on the left side of the car and the other on the right side. A chair can be seen in the middle of the scene, and a bottle is placed on the left side of the car. Additionally, there is a book resting on the left side of the car, and a clock is visible on the right side.

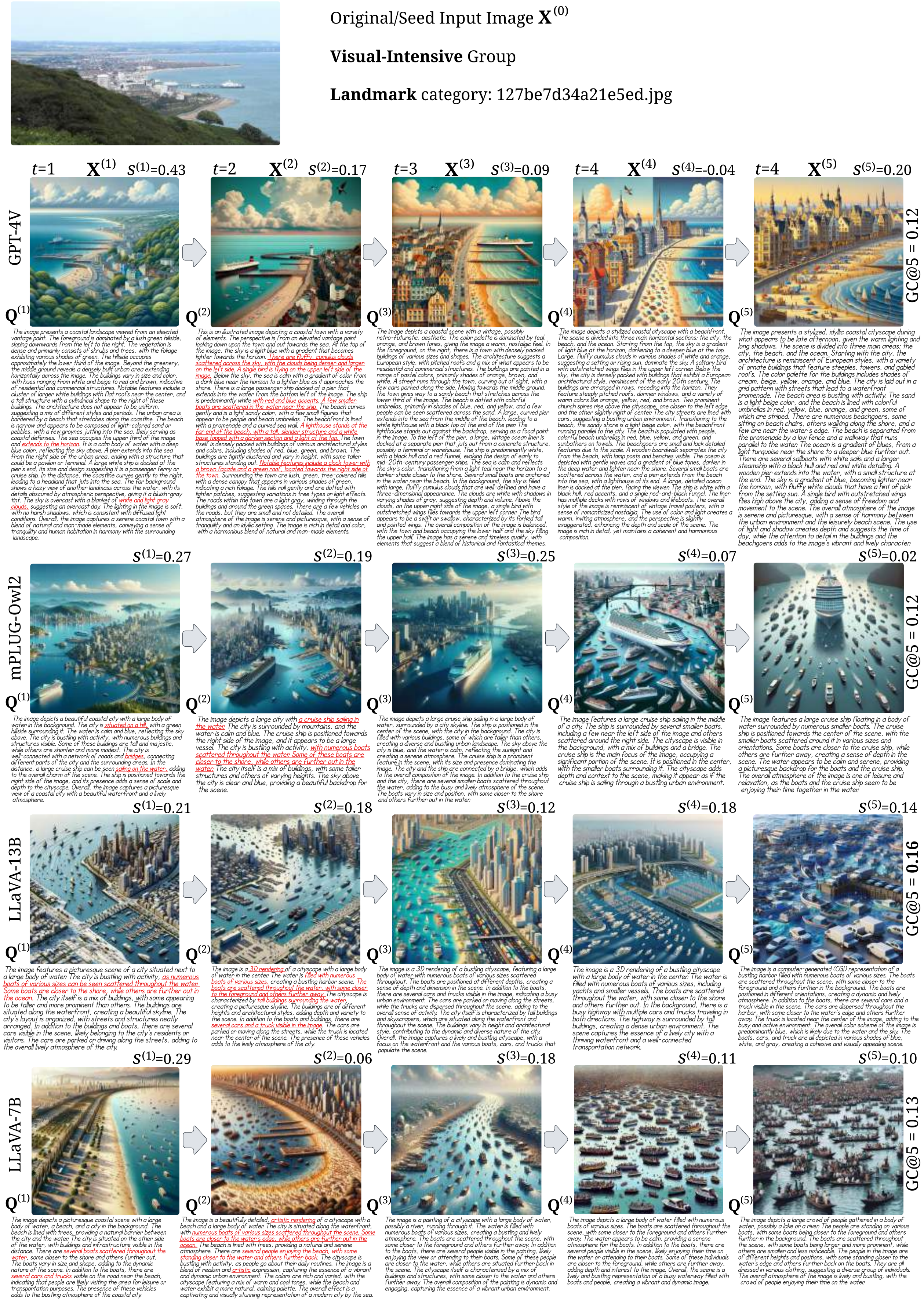
The image features a black car parked in a room, with a cat sitting on the hood of the car. The room also contains a bicycle, a clock on the wall, and a chair. There are several bottles scattered around the room, with one near the car, another near the bicycle, and a few more placed in different areas. Additionally, there are two books in the room, one near the car and the other closer to the bicycle.

The image depicts a room with a black sports car parked inside, surrounded by various objects. The car is positioned in the center of the room, with a cat sitting on the hood. There are two chairs in the room, one located on the left side and the other on the right side. A clock is mounted on the wall, and a book can be seen placed on the floor. Additionally, there are several bottles scattered throughout the room, with some placed on the floor and others on surfaces. A bicycle is also present in the room, located on the right side. The room appears to be a mix of a living space and a garage, with the car being the main focus. The presence of the cat on the hood adds a touch of warmth and liveliness to the scene.

The image depicts a room with a black sports car parked in the middle of the floor. The car is surrounded by various objects, including a bicycle, a motorcycle, and a clock. There are also several bottles scattered throughout the room, with some placed near the car and others near the bicycle. In addition to the car, there are two chairs in the room, one located near the left side and the other near the right side. A cat can be seen sitting on the floor, close to the left side of the room. A book is also present in the room, placed on the floor near the center. The room appears to be a mix of a living space and a garage, with the car being the main focal point. The presence of the bicycle and motorcycle suggests that the room may also serve as a storage area for other vehicles and equipment.

The image depicts a room filled with various objects, creating a cluttered and busy atmosphere. A car is prominently displayed in the room, taking up a significant portion of the space. A cat is also present in the room, sitting near the car. There are numerous bottles scattered throughout the room, with some placed on the floor and others on surfaces. A clock is mounted on the wall, and a book can be seen on a surface. A chair is positioned in the room, and a bicycle is leaning against a wall. In addition to these objects, there are two people in the room, one standing closer to the car and the other near the cat. The room appears to be a combination of a living space and a garage, with the car and bicycle as the main focal points.

Figure 6: Illustration of a 5-iteration GenCeption procedure run on existence images to evaluate 4 VLLMs. Each iteration t shows the generated image $\mathbf{X}^{(t)}$, the description $\mathbf{Q}^{(t)}$ of the preceding image $\mathbf{X}^{(t-1)}$, and the similarity score $s^{(t)}$ relative to $\mathbf{X}^{(0)}$. The GC@5 metric for each VLLM is also presented. Hallucinated elements within descriptions $\mathbf{Q}^{(1)}$ and $\mathbf{Q}^{(2)}$ as compared to the seed image are indicated with red underlining.


Original/Seed Input Image $\mathbf{X}^{(0)}$
Visual-Intensive Group
Landmark category: 127be7d34a21e5ed.jpg
t=1 $\mathbf{X}^{(1)}$ $s^{(1)}$=0.43
t=2 $\mathbf{X}^{(2)}$ $s^{(2)}$=0.17
t=3 $\mathbf{X}^{(3)}$ $s^{(3)}$=0.09
t=4 $\mathbf{X}^{(4)}$ $s^{(4)}$=-0.04
t=4 $\mathbf{X}^{(5)}$ $s^{(5)}$=0.20
GPT-4V
GC@5 = 0.12
mPLUG-Owl2
$s^{(1)}$=0.27 $s^{(2)}$=0.19 $s^{(3)}$=0.25 $s^{(4)}$=0.07 $s^{(5)}$=0.02
GC@5 = 0.12
LLaVA-13B
$s^{(1)}$=0.21 $s^{(2)}$=0.18 $s^{(3)}$=0.12 $s^{(4)}$=0.18 $s^{(5)}$=0.14
GC@5 = **0.16**
LLaVA-7B
$s^{(1)}$=0.29 $s^{(2)}$=0.06 $s^{(3)}$=0.18 $s^{(4)}$=0.11 $s^{(5)}$=0.10
GC@5 = 0.13

GPT-4V

$\mathbf{Q}^{(1)}$ The image presents a coastal landscape viewed from an elevated vantage point. The foreground is dominated by a lush green hillside, sloping downwards from the left to the right. The vegetation is dense and primarily consists of shrubs and trees, with the foliage exhibiting various shades of green. The hillside occupies approximately the lower third of the image. Beyond the greenery, the middle ground reveals a densely built urban area extending horizontally across the image. The buildings vary in size and color, with hues ranging from white and beige to red and brown, indicative of residential and commercial structures. Notable features include a cluster of larger white buildings with flat roofs near the center, and a tall structure with a cylindrical shape to the right of these buildings. The architecture does not appear to be uniform, suggesting a mix of different styles and periods. The urban area is bordered by a beach that stretches along the coastline. The beach is narrow and appears to be composed of light-colored sand or pebbles, with a few groynes jutting into the sea, likely serving as coastal defenses. The sea occupies the upper third of the image and extends to the horizon. It is a calm body of water with a deep blue color, reflecting the sky above. A pier extends into the sea from the right side of the urban area, ending with a structure that could be a pavilion or terminal. A large white ship is docked at the pier's end, its size and design suggesting it is a passenger ferry or cruise ship. In the distance, the coastline curves gently to the right, leading to a headland that juts into the sea. The far background shows a hazy view of another landmass across the water, with its details obscured by atmospheric perspective, giving it a bluish-gray tint. The sky is overcast with a blanket of white and light gray clouds, suggesting an overcast day. The lighting in the image is soft, with no harsh shadows, which is consistent with diffused light conditions. Overall, the image captures a serene coastal town with a blend of natural and man-made elements, conveying a sense of tranquility and human habitation in harmony with the surrounding landscape.

$\mathbf{Q}^{(2)}$ This is an illustrated image depicting a coastal town with a variety of elements. The perspective is from an elevated vantage point looking down upon the town and out towards the sea. At the top of the image, the sky is a light blue with a gradient that becomes lighter towards the horizon. There are fluffy, cumulus clouds scattered across the sky, with the clouds being denser and larger on the left side. A single bird is flying on the upper left side of the image. Below the sky, the sea is calm with a gradient of color from a dark blue near the horizon to a lighter blue as it approaches the shore. There is a large passenger ship docked at a pier that extends into the water from the bottom left of the image. The ship is predominantly white with red and blue accents. A few smaller boats are scattered in the water near the ship. The beach curves gently and is a light sandy color, with a few small figures that appear to be people and beach umbrellas. The beachfront is lined with a promenade and a curved sea wall. A lighthouse stands at the far end of the beach, with a tall, slender structure and a white base topped with a darker section and a light at the top. The town itself is densely packed with buildings of various architectural styles and colors, including shades of red, blue, green, and brown. The buildings are tightly clustered and vary in height, with some taller structures standing out. Notable features include a clock tower with a brown facade and a green roof, located towards the right side of the town. Surrounding the town are lush, green, tree-covered hills with a dense canopy that appears in various shades of green, indicating a rich foliage. The hills roll gently and are dotted with lighter patches, suggesting variations in tree types or light effects. The roads within the town are a light gray, winding through the buildings and around the green spaces. There are a few vehicles on the roads, but they are small and not detailed. The overall atmosphere of the image is serene and picturesque, with a sense of tranquility and an idyllic setting. The image is rich in detail and color, with a harmonious blend of natural and man-made elements.

$\mathbf{Q}^{(3)}$ The image depicts a coastal scene with a vintage, possibly retro-futuristic, aesthetic. The color palette is dominated by teal, orange, and brown tones, giving the image a warm, nostalgic feel. In the foreground, on the right, there is a town with densely packed buildings of various sizes and shapes. The architecture suggests a European style, with pitched roofs and a mix of what appears to be residential and commercial structures. The buildings are painted in a range of pastel colors, primarily shades of orange, brown, and white. A street runs through the town, curving out of sight, with a few cars parked along the side. Moving towards the middle ground, the town gives way to a sandy beach that stretches across the lower third of the image. The beach is dotted with colorful umbrellas, primarily in shades of blue, red, and yellow, and a few people can be seen scattered across the sand. A large, curved pier extends into the sea from the middle of the beach, leading to a white lighthouse with a black top at the end of the pier. The lighthouse stands out against the backdrop, serving as a focal point in the image. To the left of the pier, a large, vintage ocean liner is docked at a separate pier that juts out from a concrete structure, possibly a terminal or warehouse. The ship is predominantly white, with a black hull and a red funnel, evoking the design of early to mid-20th-century passenger ships. The sea is calm and reflects the sky's color, transitioning from a light teal near the horizon to a darker shade closer to the shore. Several small boats are anchored in the water near the beach. In the background, the sky is filled with large, fluffy cumulus clouds that are well-defined and have a three-dimensional appearance. The clouds are white with shadows in varying shades of gray, suggesting depth and volume. Above the clouds, on the upper right side of the image, a single bird with outstretched wings flies towards the upper left corner. The bird appears to be a swift or swallow, characterized by its forked tail and pointed wings. The overall composition of the image is balanced, with the town and beach occupying the lower half and the sky filling the upper half. The image has a serene and timeless quality, with elements that suggest a blend of historical and fantastical themes.

$\mathbf{Q}^{(4)}$ The image depicts a stylized coastal cityscape with a beachfront. The scene is divided into three main horizontal sections: the city, the beach, and the ocean. Starting from the top, the sky is a gradient of light blue at the horizon, darkening to a deeper blue at the top. Large, fluffy cumulus clouds in various shades of white and orange, suggesting a setting or rising sun, dominate the sky. A solitary bird with outstretched wings flies in the upper left corner. Below the sky, the city is densely packed with buildings that exhibit a European architectural style, reminiscent of the early 20th century. The buildings are arranged in rows, receding into the horizon. They feature steeply pitched roofs, dormer windows, and a variety of warm colors like orange, yellow, red, and brown. Two prominent church spires rise above the cityscape, one closer to the left edge and the other slightly right of center. The city streets are lined with cars, suggesting a bustling urban environment. Transitioning to the beach, the sandy shore is a light beige color, with the beachfront running parallel to the city. The beach is populated with people, colorful beach umbrellas in red, blue, yellow, and green, and sunbathers on towels. The beachgoers are small and lack detailed features due to the scale. A wooden boardwalk separates the city from the beach, with lamp posts and benches visible. The ocean is depicted with gentle waves and a gradient of blue tones, darker in the deep water and lighter near the shore. Several small boats are scattered across the water, and a pier extends from the beach into the sea, with a lighthouse at its end. A large, detailed ocean liner is docked at the pier, facing the viewer. The ship is white with a black hull, red accents, and a single red-and-black funnel. The liner has multiple decks with rows of windows and lifeboats. The overall style of the image is reminiscent of vintage travel posters, with a sense of romanticized nostalgia. The use of color and light creates a warm, inviting atmosphere, and the perspective is slightly exaggerated, enhancing the depth and scale of the scene. The image is rich in detail, yet maintains a coherent and harmonious composition.

$\mathbf{Q}^{(5)}$ The image presents a stylized, idyllic coastal cityscape during what appears to be late afternoon, given the warm lighting and long shadows. The scene is divided into three main areas: the city, the beach, and the ocean. Starting with the city, the architecture is reminiscent of European styles, with a variety of ornate buildings that feature steeples, towers, and gabled roofs. The color palette for the buildings includes shades of cream, beige, yellow, orange, and blue. The city is laid out in a grid pattern with streets that lead to a waterfront promenade. The beach area is bustling with activity. The sand is a light beige color, and the beach is lined with colorful umbrellas in red, yellow, blue, orange, and green, some of which are striped. There are numerous beachgoers, some sitting on beach chairs, others walking along the shore, and a few are near the water's edge. The beach is separated from the promenade by a low fence and a walkway that runs parallel to the water. The ocean is a gradient of blues, from a light turquoise near the shore to a deeper blue further out. There are several sailboats with white sails and a larger steamship with a black hull and red and white detailing. A wooden pier extends into the water, with a small structure at the end. The sky is a gradient of blue, becoming lighter near the horizon, with fluffy white clouds that have a hint of pink from the setting sun. A single bird with outstretched wings flies high above the city, adding a sense of freedom and movement to the scene. The overall atmosphere of the image is serene and picturesque, with a sense of harmony between the urban environment and the leisurely beach scene. The use of light and shadow creates depth and suggests the time of day, while the attention to detail in the buildings and the beachgoers adds to the image's vibrant and lively character.

mPLUG-Owl2

$\mathbf{Q}^{(1)}$ The image depicts a beautiful coastal city with a large body of water in the background. The city is situated on a hill, with a green hillside surrounding it. The water is calm and blue, reflecting the sky above. The city is bustling with activity, with numerous buildings and structures visible. Some of these buildings are tall and majestic, while others are shorter and more modest. The city is well-connected with a network of roads and bridges, connecting different parts of the city and the surrounding areas. In the distance, a large cruise ship can be seen sailing on the water, adding to the overall charm of the scene. The ship is positioned towards the right side of the image, and its presence adds a sense of scale and depth to the cityscape. Overall, the image captures a picturesque view of a coastal city with a beautiful waterfront and a lively atmosphere.

$\mathbf{Q}^{(2)}$ The image depicts a large city with a cruise ship sailing in the water. The city is surrounded by mountains, and the water is calm and blue. The cruise ship is positioned towards the right side of the image, and it appears to be a large vessel. The city is bustling with activity, with numerous boats scattered throughout the water. Some of these boats are closer to the shore, while others are further out in the water. The city itself is a mix of buildings, with some taller structures and others of varying heights. The sky above the city is clear and blue, providing a beautiful backdrop for the scene.

$\mathbf{Q}^{(3)}$ The image depicts a large cruise ship sailing in a large body of water, surrounded by a city skyline. The ship is positioned in the center of the scene, with the city in the background. The city is filled with various buildings, some of which are taller than others, creating a diverse and bustling urban landscape. The sky above the city is blue, and the water is calm, reflecting the sunlight and creating a serene atmosphere. The cruise ship is a prominent feature in the scene, with its size and presence dominating the image. The city and the ship are connected by a bridge, which adds to the overall composition of the image. In addition to the cruise ship and the city, there are several smaller boats scattered throughout the water, adding to the busy and lively atmosphere of the scene. The boats vary in size and position, with some closer to the shore and others further out in the water.

$\mathbf{Q}^{(4)}$ The image features a large cruise ship sailing in the middle of a city. The ship is surrounded by several smaller boats, including a few near the left side of the image and others scattered around the right side. The cityscape is visible in the background, with a mix of buildings and a bridge. The cruise ship is the main focus of the image, occupying a significant portion of the scene. It is positioned in the center, with the smaller boats surrounding it. The cityscape adds depth and context to the scene, making it appear as if the cruise ship is sailing through a bustling urban environment.

$\mathbf{Q}^{(5)}$ The image features a large cruise ship floating in a body of water surrounded by numerous smaller boats. The cruise ship is positioned towards the center of the scene, with the smaller boats scattered around it in various sizes and orientations. Some boats are closer to the cruise ship, while others are further away, creating a sense of depth in the scene. The water appears to be calm and serene, providing a picturesque backdrop for the boats and the cruise ship. The overall atmosphere of the image is one of leisure and relaxation, as the boats and the cruise ship seem to be enjoying their time together in the water.

LLaVA-13B

$\mathbf{Q}^{(1)}$ The image features a picturesque scene of a city situated next to a large body of water. The city is bustling with activity, as numerous boats of various sizes can be seen scattered throughout the water. Some boats are closer to the shore, while others are further out in the ocean. The city itself is a mix of buildings, with some appearing to be taller and more prominent than others. The buildings are situated along the waterfront, creating a beautiful skyline. The city's layout is organized, with streets and structures neatly arranged. In addition to the buildings and boats, there are several cars visible in the scene, likely belonging to the city's residents or visitors. The cars are parked or driving along the streets, adding to the overall lively atmosphere of the city.

$\mathbf{Q}^{(2)}$ The image is a 3D rendering of a cityscape with a large body of water in the center. The water is filled with numerous boats of various sizes, creating a bustling harbor scene. The boats are scattered throughout the water, with some closer to the foreground and others further away. The cityscape is characterized by tall buildings surrounding the water, creating a picturesque skyline. The buildings are of different heights and architectural styles, adding depth and variety to the scene. In addition to the boats and buildings, there are several cars and a truck visible in the image. The cars are parked or moving along the streets, while the truck is located near the center of the scene. The presence of these vehicles adds to the lively atmosphere of the city.

$\mathbf{Q}^{(3)}$ The image is a 3D rendering of a bustling cityscape, featuring a large body of water with numerous boats of various sizes scattered throughout. The boats are positioned at different depths, creating a sense of depth and dimension in the scene. In addition to the boats, there are several cars and trucks visible in the image, indicating a busy urban environment. The cars are parked or moving along the streets, while the trucks are dispersed throughout the scene, adding to the overall sense of activity. The city itself is characterized by tall buildings and skyscrapers, which are situated along the waterfront and throughout the scene. The buildings vary in height and architectural style, contributing to the dynamic and diverse nature of the city. Overall, the image captures a lively and bustling cityscape, with a focus on the waterfront and the various boats, cars, and trucks that populate the scene.

$\mathbf{Q}^{(4)}$ The image is a 3D rendering of a bustling cityscape with a large body of water in the center. The water is filled with numerous boats of various sizes, including yachts and smaller vessels. The boats are scattered throughout the water, with some closer to the shore and others further out. In the background, there is a busy highway with multiple cars and trucks traveling in both directions. The highway is surrounded by tall buildings, creating a dense urban environment. The scene captures the essence of a lively city with a thriving waterfront and a well-connected transportation network.

$\mathbf{Q}^{(5)}$ The image is a computer-generated (CGI) representation of a bustling harbor filled with numerous boats of various sizes. The boats are scattered throughout the scene, with some closer to the foreground and others further in the background. The boats are positioned in different orientations, creating a dynamic and lively atmosphere. In addition to the boats, there are several cars and a truck visible in the scene. The cars are dispersed throughout the harbor, with some closer to the water's edge and others further away. The truck is located near the center of the image, adding to the busy and active environment. The overall color scheme of the image is predominantly blue, which is likely due to the water and the sky. The boats, cars, and truck are all depicted in various shades of blue, white, and gray, creating a cohesive and visually appealing scene.

LLaVA-7B

$\mathbf{Q}^{(1)}$ The image depicts a picturesque coastal scene with a large body of water, a beach, and a city in the background. The beach is lined with trees, providing a natural barrier between the city and the water. The city is situated on the other side of the water, with buildings and infrastructure visible in the distance. There are several boats scattered throughout the water, some closer to the shore and others further out. The boats vary in size and shape, adding to the dynamic nature of the scene. In addition to the boats, there are several cars and trucks visible on the road near the beach, indicating that people are likely visiting the area for leisure or transportation purposes. The presence of these vehicles adds to the bustling atmosphere of the coastal city.

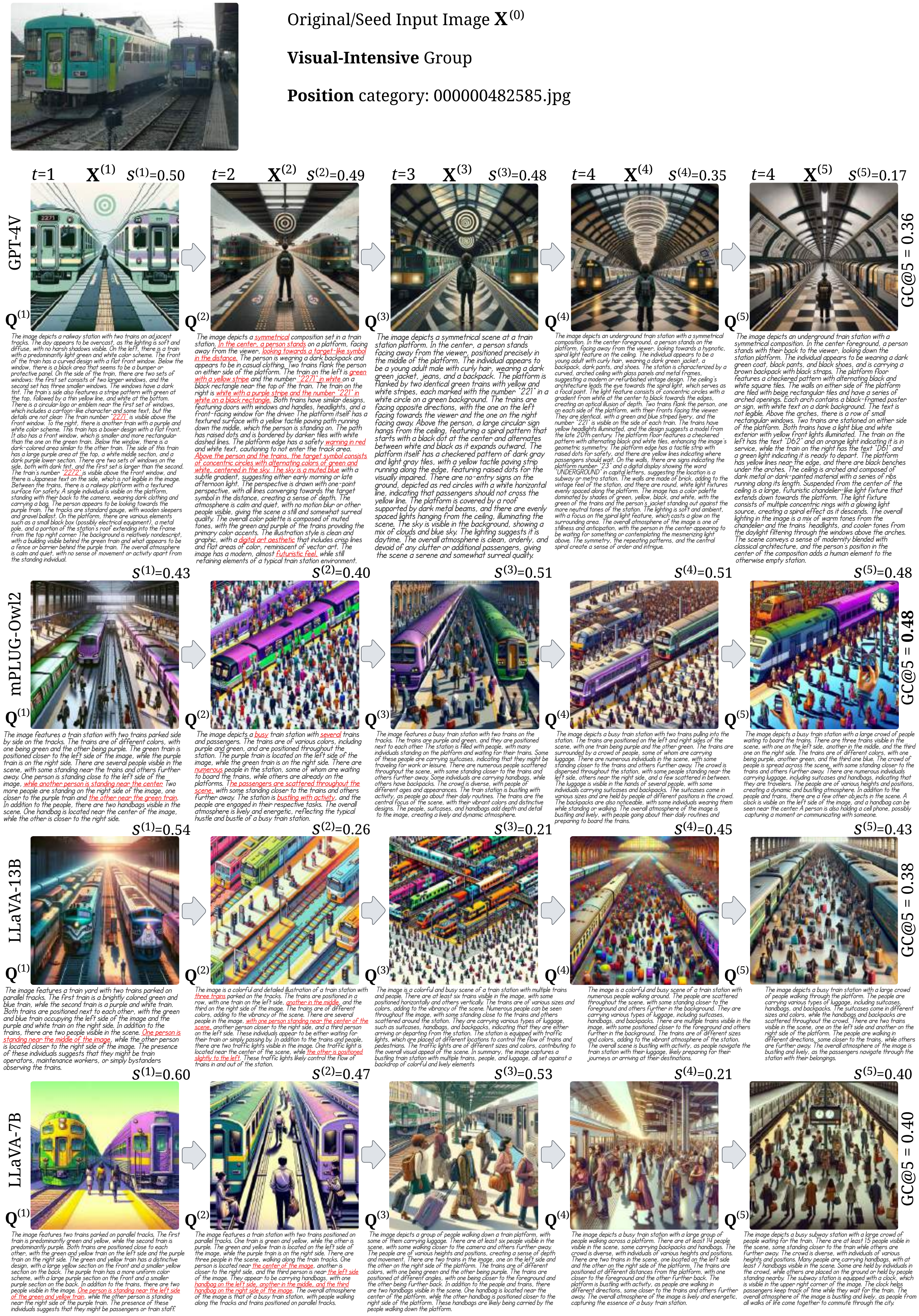
$\mathbf{Q}^{(2)}$ The image is a beautifully detailed, artistic rendering of a cityscape with a beach and a large body of water. The city is situated along the waterfront, with numerous boats of various sizes scattered throughout the scene. Some boats are closer to the water's edge, while others are further out in the ocean. The beach is lined with trees, providing a natural and serene atmosphere. There are several people enjoying the beach, with some standing closer to the water and others further back. The cityscape is bustling with activity, as people go about their daily routines. The image is a blend of realism and artistic expression, capturing the essence of a vibrant and dynamic urban environment. The colors are rich and varied, with the cityscape featuring a mix of warm and cool tones, while the beach and water exhibit a more natural, calming palette. The overall effect is a captivating and visually stunning representation of a modern city by the sea.

$\mathbf{Q}^{(3)}$ The image is a painting of a cityscape with a large body of water, possibly a river, running through it. The water is filled with numerous boats of various sizes, creating a bustling and lively atmosphere. The boats are scattered throughout the scene, with some closer to the foreground and others further away. In addition to the boats, there are several people visible in the painting, likely enjoying the view or attending to their boats. Some of these people are closer to the water, while others are situated further back in the scene. The cityscape itself is characterized by a mix of buildings and structures, with some closer to the water and others further away. The overall composition of the painting is dynamic and engaging, capturing the essence of a vibrant urban environment.

$\mathbf{Q}^{(4)}$ The image depicts a large body of water filled with numerous boats of various sizes. The boats are scattered throughout the scene, with some closer to the foreground and others further away. The water appears to be calm, providing a serene atmosphere for the boats. In addition to the boats, there are several people visible in the scene, likely enjoying their time on the water or attending to their boats. Some of these individuals are closer to the foreground, while others are further away, adding depth and interest to the image. Overall, the scene is a lively and bustling representation of a busy waterway filled with boats and people, creating a vibrant and dynamic image.

$\mathbf{Q}^{(5)}$ The image depicts a large crowd of people gathered in a body of water, possibly a lake or a river. The people are standing on various boats, with some boats being closer to the foreground and others further in the background. The boats are scattered throughout the scene, with some boats being larger and more prominent, while others are smaller and less noticeable. The people in the image are of different heights and positions, with some standing closer to the water's edge and others further back on the boats. They are all dressed in various clothing, suggesting a diverse group of individuals. The overall atmosphere of the image is lively and bustling, with the crowd of people enjoying their time on the water.

Figure 7: Illustration of a 5-iteration GenCeption procedure run on landmark images to evaluate 4 VLLMs. Each iteration t shows the generated image $\mathbf{X}^{(t)}$, the description $\mathbf{Q}^{(t)}$ of the preceding image $\mathbf{X}^{(t-1)}$, and the similarity score $s^{(t)}$ relative to $\mathbf{X}^{(0)}$. The GC@5 metric for each VLLM is also presented. Hallucinated elements within descriptions $\mathbf{Q}^{(1)}$ and $\mathbf{Q}^{(2)}$ as compared to the seed image are indicated with red underlining.



The image depicts a railway station with two trains on adjacent tracks. The day appears to be overcast, as the lighting is soft and diffuse, with no harsh shadows visible. On the left, there is a train with a predominantly light green and white color scheme. The front of the train has a curved design with a flat front window. Below the window, there is a black area that seems to be a bumper or protective panel. On the side of the train, there are two sets of windows: the first set consists of two larger windows, and the second set has three smaller windows. The windows have a dark tint. The train's side also features a stripe pattern with green at the top, followed by a thin yellow line, and white at the bottom. There is a circular logo or emblem near the first set of windows, which includes a cartoon-like character and some text, but the details are not clear. The train number "2271" is visible above the front window. To the right, there is another train with a purple and white color scheme. This train has a boxier design with a flat front. It also has a front window, which is smaller and more rectangular than the one on the green train. Below the window, there is a dark-colored area similar to the other train. The side of this train has a large purple area at the top, a white middle section, and a dark purple lower section. There are two sets of windows on the side, both with dark tint, and the first set is larger than the second. The train's number "2272" is visible above the front window, and there is Japanese text on the side, which is not legible in the image. Between the trains, there is a railway platform with a textured surface for safety. A single individual is visible on the platform, standing with their back to the camera, wearing dark clothing and carrying a bag. The person appears to be looking towards the purple train. The tracks are standard gauge, with wooden sleepers and gravel ballast. On the platform, there are various elements such as a small black box (possibly electrical equipment), a metal pole, and a portion of the station's roof extending into the frame from the top right corner. The background is relatively nondescript, with a building visible behind the green train and what appears to be a fence or barrier behind the purple train. The overall atmosphere is calm and quiet, with no sense of movement or activity apart from the standing individual.

The image depicts a symmetrical composition set in a train station. In the center, a person stands on a platform, facing away from the viewer, looking towards a target-like symbol in the distance. The person is wearing a dark backpack and appears to be in casual clothing. Two trains flank the person on either side of the platform. The train on the left is green with a yellow stripe and the number "2271" in white on a black rectangle near the top of the train. The train on the right is white with a purple stripe and the number "221" in white on a black rectangle. Both trains have similar designs, featuring doors with windows and handles, headlights, and a front-facing window for the driver. The platform itself has a textured surface with a yellow tactile paving path running down the middle, which the person is standing on. The path has raised dots and is bordered by darker tiles with white dashed lines. The platform edge has a safety warning in red and white text, cautioning to not enter the track area. Above the person and the trains, the target symbol consists of concentric circles with alternating colors of green and white, centered in the sky. The sky is a muted blue with a subtle gradient, suggesting either early morning or late afternoon light. The perspective is drawn with one-point perspective, with all lines converging towards the target symbol in the distance, creating a sense of depth. The atmosphere is calm and quiet, with no motion blur or other people visible, giving the scene a still and somewhat surreal quality. The overall color palette is composed of muted tones, with the green and purple of the trains providing the primary color accents. The illustration style is clean and graphic, with a digital art aesthetic that includes crisp lines and flat areas of color, reminiscent of vector art. The image has a modern, almost futuristic feel, while still retaining elements of a typical train station environment.

The image depicts a symmetrical scene at a train station platform. In the center, a person stands facing away from the viewer, positioned precisely in the middle of the platform. The individual appears to be a young adult male with curly hair, wearing a dark green jacket, jeans, and a backpack. The platform is flanked by two identical green trains with yellow and white stripes, each marked with the number "221" in a white circle on a green background. The trains are facing opposite directions, with the one on the left facing towards the viewer and the one on the right facing away. Above the person, a large circular sign hangs from the ceiling, featuring a spiral pattern that starts with a black dot at the center and alternates between white and black as it expands outward. The platform itself has a checkered pattern of dark gray and light gray tiles, with a yellow tactile paving strip running along the edge, featuring raised dots for the visually impaired. There are no-entry signs on the ground, depicted as red circles with a white horizontal line, indicating that passengers should not cross the yellow line. The platform is covered by a roof supported by dark metal beams, and there are evenly spaced lights hanging from the ceiling, illuminating the scene. The sky is visible in the background, showing a mix of clouds and blue sky. The lighting suggests it is daytime. The overall atmosphere is clean, orderly, and devoid of any clutter or additional passengers, giving the scene a serene and somewhat surreal quality.

The image depicts an underground train station with a symmetrical composition. In the center foreground, a person stands on the platform, facing away from the viewer, looking towards a hypnotic, spiral light feature on the ceiling. The individual appears to be a young adult with curly hair, wearing a dark green jacket, a backpack, dark pants, and shoes. The station is characterized by a curved, arched ceiling with glass panels and metal frames, suggesting a modern or refurbished vintage design. The ceiling's architecture leads the eye towards the spiral light, which serves as a focal point. The light feature consists of concentric circles with a gradient from white at the center to black towards the edges, creating an optical illusion of depth. Two trains flank the person, one on each side of the platform, with their fronts facing the viewer. They are identical, with a green and blue striped livery, and the number "221" is visible on the side of each train. The trains have yellow headlights illuminated, and the design suggests a model from the late 20th century. The platform floor features a checkered pattern with alternating black and white tiles, enhancing the image's geometric symmetry. The platform edge has a tactile strip with raised dots for safety, and there are yellow lines indicating where passengers should wait. On the walls, there are signs indicating the platform number "23" and a digital display showing the word "UNDERGROUND" in capital letters, suggesting the location is a subway or metro station. The walls are made of brick, adding to the vintage feel of the station, and there are round, white light fixtures evenly spaced along the platform. The image has a color palette dominated by shades of green, yellow, black, and white, with the green of the trains and the person's jacket standing out against the more neutral tones of the station. The lighting is soft and ambient, with a focus on the spiral light feature, which casts a glow on the surrounding area. The overall atmosphere of the image is one of stillness and anticipation, with the person in the center appearing to be waiting for something or contemplating the mesmerizing light above. The symmetry, the repeating patterns, and the central spiral create a sense of order and intrigue.

The image depicts an underground train station with a symmetrical composition. In the center foreground, a person stands with their back to the viewer, looking down the station platform. The individual appears to be wearing a dark green coat, black pants, and black shoes, and is carrying a brown backpack with black straps. The platform floor features a checkered pattern with alternating black and white square tiles. The walls on either side of the platform are tiled with beige rectangular tiles and have a series of arched openings. Each arch contains a black-framed poster or sign, with white text on a dark background. The text is not legible. Above the arches, there is a row of small rectangular windows. Two trains are stationed on either side of the platform. Both trains have a light blue and white exterior with yellow front lights illuminated. The train on the left has the text "D62" and an orange light indicating it is in service, while the train on the right has the text "D61" and a green light indicating it is ready to depart. The platform has yellow lines near the edge, and there are black benches under the arches. The ceiling is arched and composed of dark metal or dark-painted material with a series of ribs running along its length. Suspended from the center of the ceiling is a large, futuristic chandelier-like light fixture that extends down towards the platform. The light fixture consists of multiple concentric rings with a glowing light source, creating a spiral effect as it descends. The overall lighting in the image is a mix of warm tones from the chandelier and the trains' headlights, and cooler tones from the daylight filtering through the windows above the arches. The scene conveys a sense of modernity blended with classical architecture, and the person's position in the center of the composition adds a human element to the otherwise empty station.

The image features a train station with two trains parked side by side on the tracks. The trains are of different colors, with one being green and the other being purple. The green train is positioned closer to the left side of the image, while the purple train is on the right side. There are several people visible in the scene, with some standing near the trains and others further away. One person is standing close to the left side of the image, while another person is standing near the center. Two more people are standing on the right side of the image, one closer to the purple train and the other near the green train. In addition to the people, there are two handbags visible in the scene. One handbag is located near the center of the image, while the other is closer to the right side.

The image depicts a busy train station with several trains and passengers. The trains are of various colors, including purple and green, and are positioned throughout the station. The purple train is located on the left side of the image, while the green train is on the right side. There are numerous people in the station, some of whom are waiting to board the trains, while others are already on the platforms. The passengers are scattered throughout the scene, with some standing closer to the trains and others further away. The station is bustling with activity, and the people are engaged in their respective tasks. The overall atmosphere is lively and energetic, reflecting the typical hustle and bustle of a busy train station.

The image features a busy train station with two trains on the tracks. The trains are purple and green, and they are positioned next to each other. The station is filled with people, with many individuals standing on the platform and waiting for their trains. Some of these people are carrying suitcases, indicating that they might be traveling for work or leisure. There are numerous people scattered throughout the scene, with some standing closer to the trains and others further away. Some individuals are carrying handbags, while others have backpacks. The crowd is diverse, with people of different ages and appearances. The train station is bustling with activity, as people go about their daily routines. The trains are the central focus of the scene, with their vibrant colors and distinctive designs. The people, suitcases, and handbags add depth and detail to the image, creating a lively and dynamic atmosphere.

The image depicts a busy train station with two trains pulling into the station. The trains are positioned on the left and right sides of the scene, with one train being purple and the other green. The trains are surrounded by a crowd of people, some of whom are carrying luggage. There are numerous individuals in the scene, with some standing closer to the trains and others further away. The crowd is dispersed throughout the station, with some people standing near the left side, others near the right side, and a few scattered in between. The luggage is visible in the hands of several people, with some individuals carrying suitcases and backpacks. The suitcases come in various sizes and are held by people at different positions in the crowd. The backpacks are also noticeable, with some individuals wearing them while standing or walking. The overall atmosphere of the image is bustling and lively, with people going about their daily routines and preparing to board the trains.

The image depicts a busy train station with a large crowd of people waiting to board the trains. There are three trains visible in the scene, with one on the left side, another in the middle, and the third one on the right side. The trains are of different colors, with one being purple, another green, and the third one blue. The crowd of people is spread across the scene, with some standing closer to the trains and others further away. There are numerous individuals carrying luggage, including suitcases and handbags, indicating that they are travelers. The people are of various heights and positions, creating a dynamic and bustling atmosphere. In addition to the people and trains, there are a few other objects in the scene. A clock is visible on the left side of the image, and a handbag can be seen near the center. A person is also holding a cell phone, possibly capturing a moment or communicating with someone.

The image features a train yard with two trains parked on parallel tracks. The first train is a brightly colored green and blue train, while the second train is a purple and white train. Both trains are positioned next to each other, with the green and blue train occupying the left side of the image and the purple and white train on the right side. In addition to the trains, there are two people visible in the scene. One person is standing near the middle of the image, while the other person is located closer to the right side of the image. The presence of these individuals suggests that they might be train operators, maintenance workers, or simply bystanders observing the trains.

The image is a colorful and detailed illustration of a train station with three trains parked on the tracks. The trains are positioned in a row, with one train on the left side, another in the middle, and the third on the right side of the image. The trains are of different colors, adding to the vibrancy of the scene. There are several people in the image, with one person standing near the center of the scene, another person closer to the right side, and a third person on the left side. These individuals appear to be either waiting for their train or simply passing by. In addition to the trains and people, there are two traffic lights visible in the image. One traffic light is located near the center of the scene, while the other is positioned slightly to the left. These traffic lights likely control the flow of trains in and out of the station.

The image is a colorful and busy scene of a train station with multiple trains and people. There are at least six trains visible in the image, with some positioned horizontally and others vertically. The trains are of various sizes and colors, adding to the vibrancy of the scene. Numerous people can be seen throughout the image, with some standing close to the trains and others scattered around the station. They are carrying various types of luggage, such as suitcases, handbags, and backpacks, indicating that they are either arriving or departing from the station. The station is equipped with traffic lights, which are placed at different locations to control the flow of trains and pedestrians. The traffic lights are of different sizes and colors, contributing to the overall visual appeal of the scene. In summary, the image captures a bustling train station with multiple trains, people, and luggage, all set against a backdrop of colorful and lively elements

The image is a colorful and busy scene of a train station with numerous people walking around. The people are scattered throughout the scene, with some standing closer to the foreground and others further in the background. They are carrying various types of luggage, including suitcases, handbags, and backpacks. There are multiple trains visible in the image, with some positioned closer to the foreground and others further in the background. The trains are of different sizes and colors, adding to the vibrant atmosphere of the station. The overall scene is bustling with activity, as people navigate the train station with their luggage, likely preparing for their journeys or arriving at their destinations.

The image depicts a busy train station with a large crowd of people walking through the platform. The people are carrying various types of luggage, including suitcases, handbags, and backpacks. The suitcases come in different sizes and colors, while the handbags and backpacks are scattered throughout the crowd. There are two trains visible in the scene, one on the left side and another on the right side of the platform. The people are walking in different directions, some closer to the trains, while others are further away. The overall atmosphere of the image is bustling and lively, as the passengers navigate through the station with their belongings.

The image features two trains parked on parallel tracks. The first train is predominantly green and yellow, while the second train is predominantly purple. Both trains are positioned close to each other, with the green and yellow train on the left side and the purple train on the right side. The green and yellow train has a distinctive design, with a large yellow section on the front and a smaller yellow section on the back. The purple train has a more uniform color scheme, with a large purple section on the front and a smaller purple section on the back. In addition to the trains, there are two people visible in the image. One person is standing near the left side of the green and yellow train, while the other person is standing near the right side of the purple train. The presence of these individuals suggests that they might be passengers or train staff.

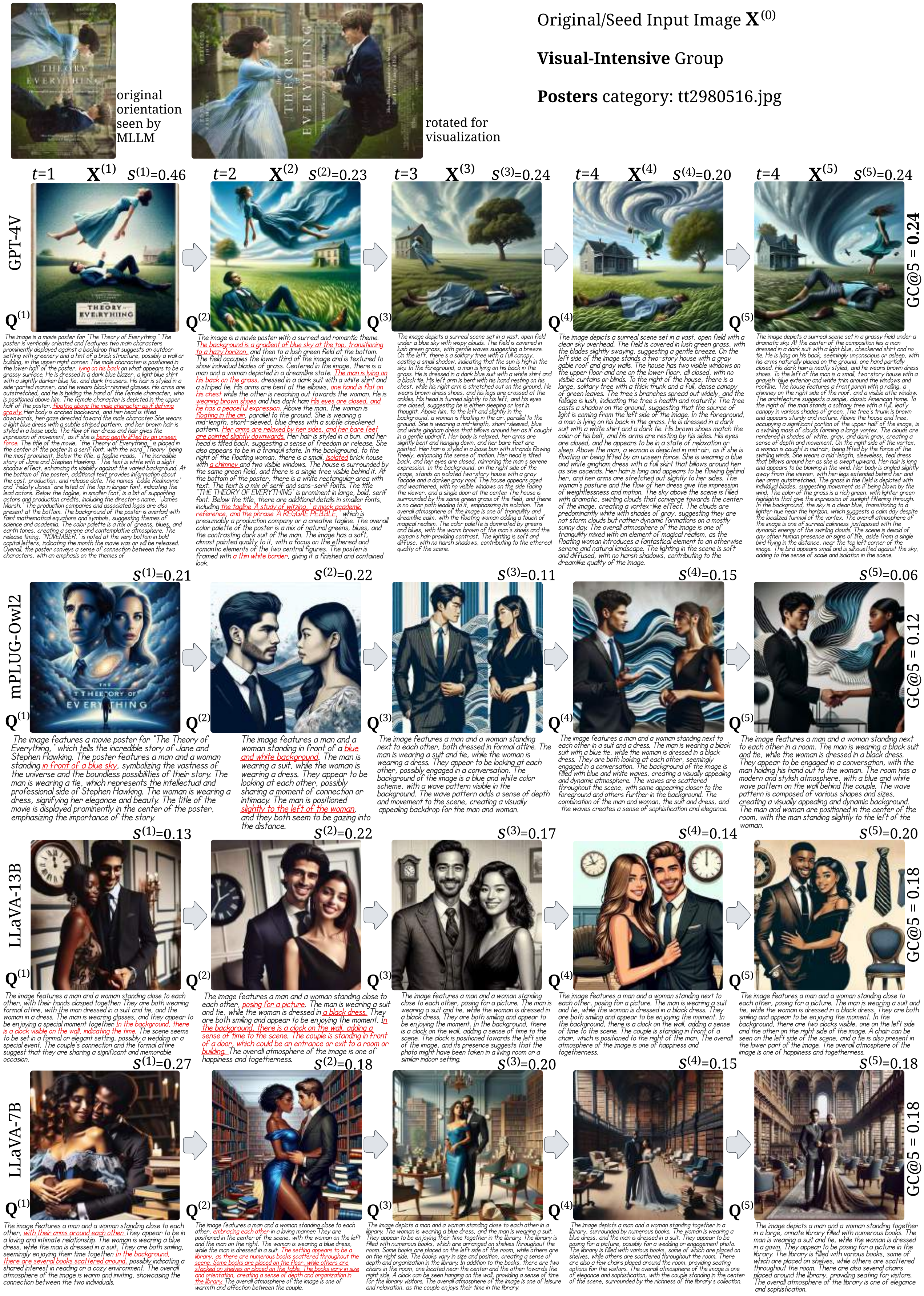
The image features a train station with two trains positioned on parallel tracks. One train is green and yellow, while the other is purple. The green and yellow train is located on the left side of the image, while the purple train is on the right side. There are three people in the scene, walking along the train tracks. One person is located near the center of the image, another is closer to the right side, and the third person is near the left side of the image. They appear to be carrying handbags, with one handbag on the left side, another in the middle, and the third handbag on the right side of the image. The overall atmosphere of the image is that of a busy train station, with people walking along the tracks and trains positioned on parallel tracks.

The image depicts a group of people walking down a train platform, with some of them carrying luggage. There are at least six people visible in the scene, with some walking closer to the camera and others further away. The people are of various heights and positions, creating a sense of depth and movement. There are two trains in the image, one on the left side and the other on the right side of the platform. The trains are of different colors, with one being green and the other being purple. The trains are positioned at different angles, with one being closer to the foreground and the other being further back. In addition to the people and trains, there are two handbags visible in the scene. One handbag is located near the center of the platform, while the other handbag is positioned closer to the right side of the platform. These handbags are likely being carried by the people walking down the platform.

The image depicts a busy train station with a large group of people walking across a platform. There are at least 14 people visible in the scene, some carrying backpacks and handbags. The crowd is diverse, with individuals of various heights and positions. There are two trains in the scene, one located on the left side and the other on the right side of the platform. The trains are positioned at different distances from the platform, with one closer to the foreground and the other further back. The platform is bustling with activity, as people are walking in different directions, some closer to the trains and others further away. The overall atmosphere of the image is lively and energetic, capturing the essence of a busy train station.

The image depicts a busy subway station with a large crowd of people waiting for the train. There are at least 15 people visible in the scene, some standing closer to the train while others are further away. The crowd is diverse, with individuals of various heights and positions. Many people are carrying handbags, with at least 7 handbags visible in the scene. Some are held by individuals in the crowd, while others are placed on the ground or held by people standing nearby. The subway station is equipped with a clock, which is visible in the upper right corner of the image. The clock helps passengers keep track of time while they wait for the train. The overall atmosphere of the image is bustling and lively, as people from all walks of life come together to commute through the city.

Figure 8: Illustration of a 5-iteration GenCeption procedure run on position images to evaluate 4 VLLMs. Each iteration t shows the generated image $\mathbf{X}^{(t)}$, the description $\mathbf{Q}^{(t)}$ of the preceding image $\mathbf{X}^{(t-1)}$, and the similarity score $s^{(t)}$ relative to $\mathbf{X}^{(0)}$. The GC@5 metric for each VLLM is also presented. Hallucinated elements within descriptions $\mathbf{Q}^{(1)}$ and $\mathbf{Q}^{(2)}$ as compared to the seed image are indicated with red underlining.



GPT-4V $\mathbf{Q}^{(1)}$: The image is a movie poster for "The Theory of Everything." The poster is vertically oriented and features two main characters prominently displayed against a backdrop that suggests an outdoor setting with greenery and a hint of a brick structure, possibly a wall or building, in the upper right corner. The male character is positioned in the lower half of the poster, lying on his back on what appears to be a grassy surface. He is dressed in a dark blue blazer, a light blue shirt with a slightly darker blue tie, and dark trousers. His hair is styled in a side-parted manner, and he wears black-rimmed glasses. His arms are outstretched, and he is holding the hand of the female character, who is positioned above him. The female character is depicted in the upper half of the poster, floating above the male character as if defying gravity. Her body is arched backward, and her head is tilted downwards, her gaze directed toward the male character. She wears a light blue dress with a subtle striped pattern, and her brown hair is styled in a loose updo. The flow of her dress and hair gives the impression of movement, as if she is being gently lifted by an unseen force. The title of the movie, "the Theory of Everything," is placed in the center of the poster in a serif font, with the word "Theory" being the most prominent. Below the title, a tagline reads, "The incredible story of Jane and Stephen Hawking." The text is white with a slight shadow effect, enhancing its visibility against the varied background. At the bottom of the poster, additional text provides information about the cast, production, and release date. The names "Eddie Redmayne" and "Felicity Jones" are listed at the top in larger font, indicating the lead actors. Below the tagline, in smaller font, is a list of supporting actors and production credits, including the director's name, "James Marsh." The production companies and associated logos are also present at the bottom. The background of the poster is overlaid with faint mathematical equations and symbols, suggesting themes of science and academia. The color palette is a mix of greens, blues, and earth tones, creating a serene and contemplative atmosphere. The release timing, "NOVEMBER," is noted at the very bottom in bold capital letters, indicating the month the movie was or will be released. Overall, the poster conveys a sense of connection between the two characters, with an emphasis on the themes of

GPT-4V $\mathbf{Q}^{(2)}$: The image is a movie poster with a surreal and romantic theme. The background is a gradient of blue sky at the top, transitioning to a hazy horizon, and then to a lush green field at the bottom. The field occupies the lower third of the image and is textured to show individual blades of grass. Centered in the image, there is a man and a woman depicted in a dreamlike state. The man is lying on his back on the grass, dressed in a dark suit with a white shirt and a striped tie. His arms are bent at the elbows, one hand is flat on his chest while the other is reaching out towards the woman. He is wearing brown shoes and has dark hair. His eyes are closed, and he has a peaceful expression. Above the man, the woman is floating in the air, parallel to the ground. She is wearing a mid-length, short-sleeved, blue dress with a subtle checkered pattern. Her arms are relaxed by her sides, and her bare feet are pointed slightly downwards. Her hair is styled in a bun, and her head is tilted back, suggesting a sense of freedom or release. She also appears to be in a tranquil state. In the background, to the right of the floating woman, there is a small, isolated brick house with a chimney and two visible windows. The house is surrounded by the same green field, and there is a single tree visible behind it. At the bottom of the poster, there is a white rectangular area with text. The text is a mix of serif and sans-serif fonts. The title "THE THEORY OF EVERYTHING" is prominent in large, bold, serif font. Below the title, there are additional details in smaller fonts, including the tagline "A study of witzing," a mock academic reference, and the phrase "A REGGAE PEBBLE," which is presumably a production company or a creative tagline. The overall color palette of the poster is a mix of natural greens, blues, and the contrasting dark suit of the man. The image has a soft, almost painted quality to it, with a focus on the ethereal and romantic elements of the two central figures. The poster is framed with a thin white border, giving it a finished and contained look.

GPT-4V $\mathbf{Q}^{(3)}$: The image depicts a surreal scene set in a vast, open field under a blue sky with wispy clouds. The field is covered in lush green grass, with gentle waves suggesting a breeze. On the left, there's a solitary tree with a full canopy, casting a small shadow, indicating that the sun is high in the sky. In the foreground, a man is lying on his back in the grass. He is dressed in a dark blue suit with a white shirt and a black tie. His left arm is bent with his hand resting on his chest, while his right arm is stretched out on the ground. He wears brown dress shoes, and his legs are crossed at the ankles. His head is turned slightly to his left, and his eyes are closed, suggesting he is either sleeping or lost in thought. Above him, to the left and slightly in the background, a woman is floating in the air, parallel to the ground. She is wearing a mid-length, short-sleeved, blue and white gingham dress that billows around her as if caught in a gentle updraft. Her body is relaxed, her arms are slightly bent and hanging down, and her bare feet are pointed. Her hair is styled in a loose bun with strands flowing freely, enhancing the sense of motion. Her head is tilted back, and her eyes are closed, mirroring the man's serene expression. In the background, on the right side of the image, stands an isolated two-story house with a gray facade and a darker gray roof. The house appears aged and weathered, with no visible windows on the side facing the viewer, and a single door at the center. The house is surrounded by the same green grass of the field, and there is no clear path leading to it, emphasizing its isolation. The overall atmosphere of the image is one of tranquility and dreamlike calm, with the floating woman adding a touch of magical realism. The color palette is dominated by greens and blues, with the warm brown of the man's shoes and the woman's hair providing contrast. The lighting is soft and diffuse, with no harsh shadows, contributing to the ethereal quality of the scene.

GPT-4V $\mathbf{Q}^{(4)}$: The image depicts a surreal scene set in a vast, open field with a clear sky overhead. The field is covered in lush green grass, with the blades slightly swaying, suggesting a gentle breeze. On the left side of the image stands a two-story house with a gray gable roof and gray walls. The house has two visible windows on the upper floor and one on the lower floor, all closed, with no visible curtains or blinds. To the right of the house, there is a large, solitary tree with a thick trunk and a full, dense canopy of green leaves. The tree's branches spread out widely, and the foliage is lush, indicating the tree's health and maturity. The tree casts a shadow on the ground, suggesting that the source of light is coming from the left side of the image. In the foreground, a man is lying on his back in the grass. He is dressed in a dark suit with a white shirt and a dark tie. His brown shoes match the color of his belt, and his arms are resting by his sides. His eyes are closed, and he appears to be in a state of relaxation or sleep. Above the man, a woman is depicted in mid-air, as if she is floating or being lifted by an unseen force. She is wearing a blue and white gingham dress with a full skirt that billows around her as she ascends. Her hair is long and appears to be flowing behind her, and her arms are stretched out slightly to her sides. The woman's posture and the flow of her dress give the impression of weightlessness and motion. The sky above the scene is filled with dramatic, swirling clouds that converge towards the center of the image, creating a vortex-like effect. The clouds are predominantly white with shades of gray, suggesting they are not storm clouds but rather dynamic formations on a mostly sunny day. The overall atmosphere of the image is one of tranquility mixed with an element of magical realism, as the floating woman introduces a fantastical element to an otherwise serene and natural landscape. The lighting in the scene is soft and diffused, with no harsh shadows, contributing to the dreamlike quality of the image.

GPT-4V $\mathbf{Q}^{(5)}$: The image depicts a surreal scene set in a grassy field under a dramatic sky. At the center of the composition lies a man dressed in a dark suit with a light blue, checkered shirt and no tie. He is lying on his back, seemingly unconscious or asleep, with his arms naturally placed on the ground, one hand partially closed. His dark hair is neatly styled, and he wears brown dress shoes. To the left of the man is a small, two-story house with a grayish-blue exterior and white trim around the windows and roofline. The house features a front porch with a railing, a chimney on the right side of the roof, and a visible attic window. The architecture suggests a simple, classic American home. To the right of the man stands a solitary tree with a full, leafy canopy in various shades of green. The tree's trunk is brown and appears sturdy and mature. Above the house and tree, occupying a significant portion of the upper half of the image, is a swirling mass of clouds forming a large vortex. The clouds are rendered in shades of white, gray, and dark gray, creating a sense of depth and movement. On the right side of the vortex, a woman is caught in mid-air, being lifted by the force of the swirling winds. She wears a mid-length, sleeveless, teal dress that billows around her as she is swept upward. Her hair is long and appears to be blowing in the wind. Her body is angled slightly away from the viewer, with her legs extended behind her and her arms outstretched. The grass in the field is depicted with individual blades, suggesting movement as if being blown by the wind. The color of the grass is a rich green, with lighter green highlights that give the impression of sunlight filtering through. In the background, the sky is a clear blue, transitioning to a lighter hue near the horizon, which suggests a calm day despite the localized turmoil of the vortex. The overall atmosphere of the image is one of surreal calmness juxtaposed with the dynamic energy of the swirling clouds. The scene is devoid of any other human presence or signs of life, aside from a single bird flying in the distance, near the top left corner of the image. The bird appears small and is silhouetted against the sky, adding to the sense of scale and isolation in the scene.

mPLUG-Owl2 $\mathbf{Q}^{(1)}$: The image features a movie poster for "The Theory of Everything," which tells the incredible story of Jane and Stephen Hawking. The poster features a man and a woman standing in front of a blue sky, symbolizing the vastness of the universe and the boundless possibilities of their story. The man is wearing a tie, which represents the intellectual and professional side of Stephen Hawking. The woman is wearing a dress, signifying her elegance and beauty. The title of the movie is displayed prominently in the center of the poster, emphasizing the importance of the story.

mPLUG-Owl2 $\mathbf{Q}^{(2)}$: The image features a man and a woman standing in front of a blue and white background. The man is wearing a suit, while the woman is wearing a dress. They appear to be looking at each other, possibly sharing a moment of connection or intimacy. The man is positioned slightly to the left of the woman, and they both seem to be gazing into the distance.

mPLUG-Owl2 $\mathbf{Q}^{(3)}$: The image features a man and a woman standing next to each other, both dressed in formal attire. The man is wearing a suit and tie, while the woman is wearing a dress. They appear to be looking at each other, possibly engaged in a conversation. The background of the image is a blue and white color scheme, with a wave pattern visible in the background. The wave pattern adds a sense of depth and movement to the scene, creating a visually appealing backdrop for the man and woman.

mPLUG-Owl2 $\mathbf{Q}^{(4)}$: The image features a man and a woman standing next to each other in a suit and a dress. The man is wearing a black suit with a blue tie, while the woman is dressed in a black dress. They are both looking at each other, seemingly engaged in a conversation. The background of the image is filled with blue and white waves, creating a visually appealing and dynamic atmosphere. The waves are scattered throughout the scene, with some appearing closer to the foreground and others further in the background. The combination of the man and woman, the suit and dress, and the waves creates a sense of sophistication and elegance.

mPLUG-Owl2 $\mathbf{Q}^{(5)}$: The image features a man and a woman standing next to each other in a room. The man is wearing a black suit and tie, while the woman is dressed in a black dress. They appear to be engaged in a conversation, with the man holding his hand out to the woman. The room has a modern and stylish atmosphere, with a blue and white wave pattern on the wall behind the couple. The wave pattern is composed of various shapes and sizes, creating a visually appealing and dynamic background. The man and woman are positioned in the center of the room, with the man standing slightly to the left of the woman.

LLaVA-13B $\mathbf{Q}^{(1)}$: The image features a man and a woman standing close to each other, with their hands clasped together. They are both wearing formal attire, with the man dressed in a suit and tie, and the woman in a dress. The man is wearing glasses, and they appear to be enjoying a special moment together. In the background, there is a clock visible on the wall, indicating the time. The scene seems to be set in a formal or elegant setting, possibly a wedding or a special event. The couple's connection and the formal attire suggest that they are sharing a significant and memorable occasion.

LLaVA-13B $\mathbf{Q}^{(2)}$: The image features a man and a woman standing close to each other, posing for a picture. The man is wearing a suit and tie, while the woman is dressed in a black dress. They are both smiling and appear to be enjoying the moment. In the background, there is a clock on the wall, adding a sense of time to the scene. The couple is standing in front of a door, which could be an entrance or exit to a room or building. The overall atmosphere of the image is one of happiness and togetherness.

LLaVA-13B $\mathbf{Q}^{(3)}$: The image features a man and a woman standing close to each other, posing for a picture. The man is wearing a suit and tie, while the woman is dressed in a black dress. They are both smiling and appear to be enjoying the moment. In the background, there is a clock on the wall, adding a sense of time to the scene. The clock is positioned towards the left side of the image, and its presence suggests that the photo might have been taken in a living room or a similar indoor setting.

LLaVA-13B $\mathbf{Q}^{(4)}$: The image features a man and a woman standing next to each other, posing for a picture. The man is wearing a suit and tie, while the woman is dressed in a black dress. They are both smiling and appear to be enjoying the moment. In the background, there is a clock on the wall, adding a sense of time to the scene. The couple is standing in front of a chair, which is positioned to the right of the man. The overall atmosphere of the image is one of happiness and togetherness.

LLaVA-13B $\mathbf{Q}^{(5)}$: The image features a man and a woman standing close to each other, posing for a picture. The man is wearing a suit and tie, while the woman is dressed in a black dress. They are both smiling and appear to be enjoying the moment. In the background, there are two clocks visible, one on the left side and the other on the right side of the image. A chair can be seen on the left side of the scene, and a tie is also present in the lower part of the image. The overall atmosphere of the image is one of happiness and togetherness.

LLaVA-7B $\mathbf{Q}^{(1)}$: The image features a man and a woman standing close to each other, with their arms around each other. They appear to be in a loving and intimate relationship. The woman is wearing a blue dress, while the man is dressed in a suit. They are both smiling, seemingly enjoying their time together. In the background, there are several books scattered around, possibly indicating a shared interest in reading or a cozy environment. The overall atmosphere of the image is warm and inviting, showcasing the connection between the two individuals.

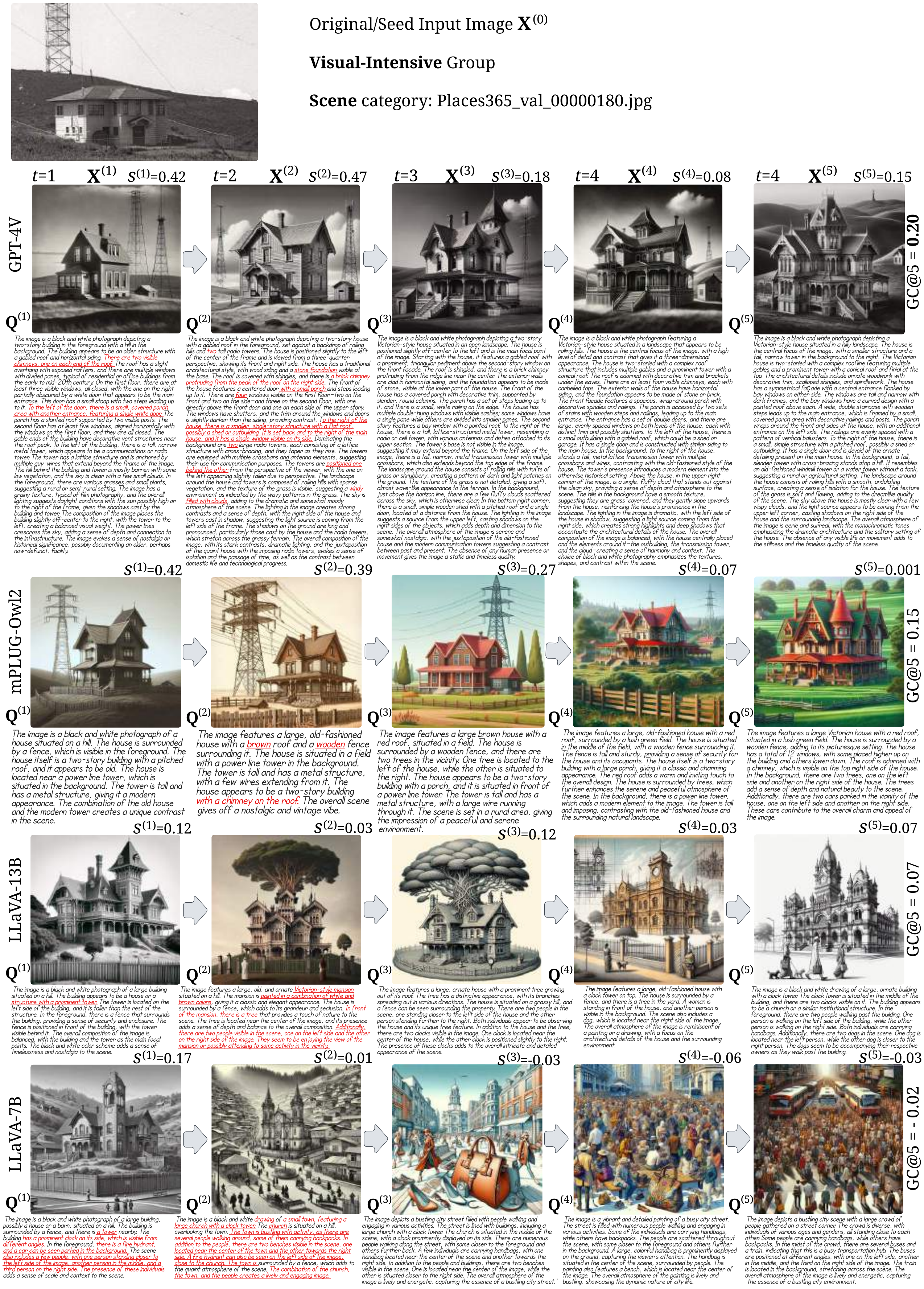
LLaVA-7B $\mathbf{Q}^{(2)}$: The image features a man and a woman standing close to each other, embracing each other in a loving manner. They are positioned in the center of the scene, with the woman on the left and the man on the right. The woman is wearing a blue dress, while the man is dressed in a suit. The setting appears to be a library, as there are numerous books scattered throughout the scene. Some books are placed on the floor, while others are stacked on shelves or placed on the table. The books vary in size and orientation, creating a sense of depth and organization in the library. The overall atmosphere of the image is one of warmth and affection between the couple.

LLaVA-7B $\mathbf{Q}^{(3)}$: The image depicts a man and a woman standing close to each other in a library. The woman is wearing a blue dress, and the man is wearing a suit. They appear to be enjoying their time together in the library. The library is filled with numerous books, which are arranged on shelves throughout the room. Some books are placed on the left side of the room, while others are on the right side. The books vary in size and position, creating a sense of depth and organization in the library. In addition to the books, there are two chairs in the room, one located near the center and the other towards the right side. A clock can be seen hanging on the wall, providing a sense of time for the library visitors. The overall atmosphere of the image is one of leisure and relaxation, as the couple enjoys their time in the library.

LLaVA-7B $\mathbf{Q}^{(4)}$: The image depicts a man and a woman standing together in a library, surrounded by numerous books. The woman is wearing a blue dress, and the man is dressed in a suit. They appear to be posing for a picture, possibly for a wedding or engagement photo. The library is filled with various books, some of which are placed on shelves, while others are scattered throughout the room. There are also a few chairs placed around the room, providing seating options for the visitors. The overall atmosphere of the image is one of elegance and sophistication, with the couple standing in the center of the scene, surrounded by the richness of the library's collection.

LLaVA-7B $\mathbf{Q}^{(5)}$: The image depicts a man and a woman standing together in a large, ornate library filled with numerous books. The man is wearing a suit and tie, while the woman is dressed in a gown. They appear to be posing for a picture in the library. The library is filled with various books, some of which are placed on shelves, while others are scattered throughout the room. There are also several chairs placed around the library, providing seating for visitors. The overall atmosphere of the library is one of elegance and sophistication.

Figure 9: Illustration of a 5-iteration GenCeption procedure run on poster images to evaluate 4 VLLMs. Each iteration t shows the generated image $\mathbf{X}^{(t)}$, the description $\mathbf{Q}^{(t)}$ of the preceding image $\mathbf{X}^{(t-1)}$, and the similarity score $s^{(t)}$ relative to $\mathbf{X}^{(0)}$. The GC@5 metric for each VLLM is also presented. Hallucinated elements within descriptions $\mathbf{Q}^{(1)}$ and $\mathbf{Q}^{(2)}$ as compared to the seed image are indicated with red underlining.


Original/Seed Input Image $\mathbf{X}^{(0)}$
Visual-Intensive Group
Scene category: Places365_val_00000180.jpg
t=1 $\mathbf{X}^{(1)}$ $s^{(1)}$=0.42
t=2 $\mathbf{X}^{(2)}$ $s^{(2)}$=0.47
t=3 $\mathbf{X}^{(3)}$ $s^{(3)}$=0.18
t=4 $\mathbf{X}^{(4)}$ $s^{(4)}$=0.08
t=4 $\mathbf{X}^{(5)}$ $s^{(5)}$=0.15
GPT-4V
GC@5 = 0.20
$s^{(1)}$=0.42
$s^{(2)}$=0.39
$s^{(3)}$=0.27
$s^{(4)}$=0.07
$s^{(5)}$=0.001
mPLUG-Owl2
GC@5 = 0.15
$s^{(1)}$=0.12
$s^{(2)}$=0.03
$s^{(3)}$=0.12
$s^{(4)}$=0.03
$s^{(5)}$=0.07
LLaVA-13B
GC@5 = 0.07
$s^{(1)}$=0.17
$s^{(2)}$=0.01
$s^{(3)}$=-0.03
$s^{(4)}$=-0.06
$s^{(5)}$=-0.03
LLaVA-7B
GC@5 = - 0.02

$\mathbf{Q}^{(1)}$ The image is a black and white photograph depicting a two-story building in the foreground with a hill in the background. The building appears to be an older structure with a gabled roof and horizontal siding. There are two visible chimneys, one on each end of the roof. The roof has a slight overhang with exposed rafters, and there are multiple windows with divided panes, typical of residential or office buildings from the early to mid-20th century. On the first floor, there are at least three visible windows, all closed, with the one on the right partially obscured by a white door that appears to be the main entrance. This door has a small stoop with two steps leading up to it. To the left of the door, there is a small, covered porch area with another entrance, featuring a single white door. The porch has a slanted roof supported by two visible posts. The second floor has at least five windows, aligned horizontally with the windows on the first floor, and they are all closed. The gable ends of the building have decorative vent structures near the roof peak. To the left of the building, there is a tall, narrow metal tower, which appears to be a communications or radio tower. The tower has a lattice structure and is anchored by multiple guy-wires that extend beyond the frame of the image. The hill behind the building and tower is mostly barren with some low vegetation, and the sky is clear with a few small clouds. In the foreground, there are various grasses and small plants, suggesting a rural or semi-rural setting. The image has a grainy texture, typical of film photography, and the overall lighting suggests daylight conditions with the sun possibly high or to the right of the frame, given the shadows cast by the building and tower. The composition of the image places the building slightly off-center to the right, with the tower to the left, creating a balanced visual weight. The power lines crisscross the sky, adding a sense of depth and connection to the infrastructure. The image evokes a sense of nostalgia or historical significance, possibly documenting an older, perhaps now-defunct, facility.

$\mathbf{Q}^{(2)}$ The image is a black and white photograph depicting a two-story house with a gabled roof in the foreground, set against a backdrop of rolling hills and two tall radio towers. The house is positioned slightly to the left of the center of the frame and is viewed from a three-quarter perspective, showing its front and right side. The house has a traditional architectural style, with wood siding and a stone foundation visible at the base. The roof is covered with shingles, and there is a brick chimney protruding from the peak of the roof on the right side. The front of the house features a centered door with a small porch and steps leading up to it. There are four windows visible on the first floor—two on the front and two on the side—and three on the second floor, with one directly above the front door and one on each side of the upper story. The windows have shutters, and the trim around the windows and doors is slightly darker than the siding, providing contrast. To the right of the house, there is a smaller, single-story structure with a flat roof, possibly a shed or outbuilding. It is set back and to the right of the main house, and it has a single window visible on its side. Dominating the background are two large radio towers, each consisting of a lattice structure with cross-bracing, and they taper as they rise. The towers are equipped with multiple crossbars and antenna elements, suggesting their use for communication purposes. The towers are positioned one behind the other from the perspective of the viewer, with the one on the left appearing slightly taller due to perspective. The landscape around the house and towers is composed of rolling hills with sparse vegetation, and the texture of the grass is visible, suggesting a windy environment as indicated by the wavy patterns in the grass. The sky is filled with clouds, adding to the dramatic and somewhat moody atmosphere of the scene. The lighting in the image creates strong contrasts and a sense of depth, with the right side of the house and towers cast in shadow, suggesting the light source is coming from the left side of the frame. The shadows on the ground are long and pronounced, particularly those cast by the house and the radio towers, which stretch across the grassy terrain. The overall composition of the image, with its stark contrasts, dramatic lighting, and the juxtaposition of the quaint house with the imposing radio towers, evokes a sense of isolation and the passage of time, as well as the contrast between domestic life and technological progress.

$\mathbf{Q}^{(3)}$ The image is a black and white photograph depicting a two-story Victorian-style house situated in an open landscape. The house is positioned slightly off-center to the left and is the main focal point of the image. Starting with the house, it features a gabled roof with a prominent, triangular pediment above the second-story window on the front facade. The roof is shingled, and there is a brick chimney protruding from the ridge line near the center. The exterior walls are clad in horizontal siding, and the foundation appears to be made of stone, visible at the lower part of the house. The front of the house has a covered porch with decorative trim, supported by slender, round columns. The porch has a set of steps leading up to it, and there is a small, white railing on the edge. The house has multiple double-hung windows with visible sashes; some windows have a single pane while others are divided into smaller panes. The second story features a bay window with a pointed roof. To the right of the house, there is a tall, lattice-structured metal tower, resembling a radio or cell tower, with various antennas and dishes attached to its upper section. The tower's base is not visible in the image, suggesting it may extend beyond the frame. On the left side of the image, there is a tall, narrow, metal transmission tower with multiple crossbars, which also extends beyond the top edge of the frame. The landscape around the house consists of rolling hills with tufts of grass or shrubbery, creating a pattern of dark and light patches on the ground. The texture of the grass is not detailed, giving a soft, almost wave-like appearance to the terrain. In the background, just above the horizon line, there are a few fluffy clouds scattered across the sky, which is otherwise clear. In the bottom right corner, there is a small, simple wooden shed with a pitched roof and a single door, located at a distance from the house. The lighting in the image suggests a source from the upper left, casting shadows on the right sides of the objects, which adds depth and dimension to the scene. The overall atmosphere of the image is serene and somewhat nostalgic, with the juxtaposition of the old-fashioned house and the modern communication towers suggesting a contrast between past and present. The absence of any human presence or movement gives the image a static and timeless quality.

$\mathbf{Q}^{(4)}$ The image is a black and white photograph featuring a Victorian-style house situated in a landscape that appears to be rolling hills. The house is the central focus of the image, with a high level of detail and contrast that gives it a three-dimensional appearance. The house is two-storied with a complex roof structure that includes multiple gables and a prominent tower with a conical roof. The roof is adorned with decorative trim and brackets under the eaves. There are at least four visible chimneys, each with corbelled tops. The exterior walls of the house have horizontal siding, and the foundation appears to be made of stone or brick. The front facade features a spacious, wrap-around porch with decorative spindles and railings. The porch is accessed by two sets of stairs with wooden steps and railings, leading up to the main entrance. The entrance has a set of double doors, and there are large, evenly spaced windows on both levels of the house, each with distinct trim and possibly shutters. To the left of the house, there is a small outbuilding with a gabled roof, which could be a shed or garage. It has a single door and is constructed with similar siding to the main house. In the background, to the right of the house, stands a tall, metal lattice transmission tower with multiple crossbars and wires, contrasting with the old-fashioned style of the house. The tower's presence introduces a modern element into the otherwise historical setting. Above the house, in the upper right corner of the image, is a single, fluffy cloud that stands out against the clear sky, providing a sense of depth and atmosphere to the scene. The hills in the background have a smooth texture, suggesting they are grass-covered, and they gently slope upwards from the house, reinforcing the house's prominence in the landscape. The lighting in the image is dramatic, with the left side of the house in shadow, suggesting a light source coming from the right side, which creates strong highlights and deep shadows that accentuate the architectural details of the house. The overall composition of the image is balanced, with the house centrally placed and the elements around it—the outbuilding, the transmission tower, and the cloud—creating a sense of harmony and context. The choice of black and white photography emphasizes the textures, shapes, and contrast within the scene.

$\mathbf{Q}^{(5)}$ The image is a black and white photograph depicting a Victorian-style house situated in a hilly landscape. The house is the central focus of the image, with a smaller structure and a tall, narrow tower in the background to the right. The Victorian house is two-storied with a complex roofline featuring multiple gables and a prominent tower with a conical roof and finial at the top. The architectural details include ornate woodwork with decorative trim, scalloped shingles, and spindlework. The house has a symmetrical façade with a central entrance flanked by bay windows on either side. The windows are tall and narrow with dark frames, and the bay windows have a curved design with a pointed roof above each. A wide, double staircase with wooden steps leads up to the main entrance, which is framed by a small, covered porch area with decorative railings and posts. The porch wraps around the front and sides of the house, with an additional entrance on the left side. The railings are evenly spaced with a pattern of vertical balusters. To the right of the house, there is a small, simple structure with a pitched roof, possibly a shed or outbuilding. It has a single door and is devoid of the ornate detailing present on the main house. In the background, a tall, slender tower with cross-bracing stands atop a hill. It resembles an old-fashioned windmill tower or a water tower without a tank, suggesting a rural or agricultural setting. The landscape around the house consists of rolling hills with a smooth, undulating surface, creating a sense of isolation for the house. The texture of the grass is soft and flowing, adding to the dreamlike quality of the scene. The sky above the house is mostly clear with a few wispy clouds, and the light source appears to be coming from the upper left corner, casting shadows on the right side of the house and the surrounding landscape. The overall atmosphere of the image is eerie and surreal, with the monochromatic tones emphasizing the dramatic architecture and the solitary setting of the house. The absence of any visible life or movement adds to the stillness and the timeless quality of the scene.

$\mathbf{Q}^{(1)}$ The image is a black and white photograph of a house situated on a hill. The house is surrounded by a fence, which is visible in the foreground. The house itself is a two-story building with a pitched roof, and it appears to be old. The house is located near a power line tower, which is situated in the background. The tower is tall and has a metal structure, giving it a modern appearance. The combination of the old house and the modern tower creates a unique contrast in the scene.

$\mathbf{Q}^{(2)}$ The image features a large, old-fashioned house with a brown roof and a wooden fence surrounding it. The house is situated in a field with a power line tower in the background. The tower is tall and has a metal structure, with a few wires extending from it. The house appears to be a two-story building with a chimney on the roof. The overall scene gives off a nostalgic and vintage vibe.

$\mathbf{Q}^{(3)}$ The image features a large brown house with a red roof, situated in a field. The house is surrounded by a wooden fence, and there are two trees in the vicinity. One tree is located to the left of the house, while the other is situated to the right. The house appears to be a two-story building with a porch, and it is situated in front of a power line tower. The tower is tall and has a metal structure, with a large wire running through it. The scene is set in a rural area, giving the impression of a peaceful and serene environment.

$\mathbf{Q}^{(4)}$ The image features a large, old-fashioned house with a red roof, surrounded by a lush green field. The house is situated in the middle of the field, with a wooden fence surrounding it. The fence is tall and sturdy, providing a sense of security for the house and its occupants. The house itself is a two-story building with a large porch, giving it a classic and charming appearance. The red roof adds a warm and inviting touch to the overall design. The house is surrounded by trees, which further enhances the serene and peaceful atmosphere of the scene. In the background, there is a power line tower, which adds a modern element to the image. The tower is tall and imposing, contrasting with the old-fashioned house and the surrounding natural landscape.

$\mathbf{Q}^{(5)}$ The image features a large Victorian house with a red roof, situated in a lush green field. The house is surrounded by a wooden fence, adding to its picturesque setting. The house has a total of 12 windows, with some placed higher up on the building and others lower down. The roof is adorned with a chimney, which is visible on the top right side of the house. In the background, there are two trees, one on the left side and another on the right side of the house. The trees add a sense of depth and natural beauty to the scene. Additionally, there are two cars parked in the vicinity of the house, one on the left side and another on the right side. These cars contribute to the overall charm and appeal of the image.

$\mathbf{Q}^{(1)}$ The image is a black and white photograph of a large building situated on a hill. The building appears to be a house or a structure with a prominent tower. The tower is located on the left side of the building, and it is taller than the rest of the structure. In the foreground, there is a fence that surrounds the building, providing a sense of security and enclosure. The fence is positioned in front of the building, with the tower visible behind it. The overall composition of the image is balanced, with the building and the tower as the main focal points. The black and white color scheme adds a sense of timelessness and nostalgia to the scene.

$\mathbf{Q}^{(2)}$ The image features a large, old, and ornate Victorian-style mansion situated on a hill. The mansion is painted in a combination of white and brown colors, giving it a classic and elegant appearance. The house is surrounded by a fence, which adds to its grandeur and seclusion. In front of the mansion, there is a tree that provides a touch of nature to the scene. The tree is located near the center of the image, and its presence adds a sense of depth and balance to the overall composition. Additionally, there are two people visible in the scene, one on the left side and the other on the right side of the image. They seem to be enjoying the view of the mansion or possibly attending to some activity in the vicinity.

$\mathbf{Q}^{(3)}$ The image features a large, ornate house with a prominent tree growing out of its roof. The tree has a distinctive appearance, with its branches spreading out in various directions. The house is situated on a grassy hill, and a fence can be seen surrounding the property. There are two people in the scene, one standing closer to the left side of the house and the other person standing further to the right. Both individuals appear to be observing the house and its unique tree feature. In addition to the house and the tree, there are two clocks visible in the image. One clock is located near the center of the house, while the other clock is positioned slightly to the right. The presence of these clocks adds to the overall intricate and detailed appearance of the scene.

$\mathbf{Q}^{(4)}$ The image features a large, old-fashioned house with a clock tower on top. The house is surrounded by a fence, and there is a tree in the yard. A woman is standing in front of the house, and another person is visible in the background. The scene also includes a dog, which is located near the right side of the image. The overall atmosphere of the image is reminiscent of a painting or a drawing, with a focus on the architectural details of the house and the surrounding environment.

$\mathbf{Q}^{(5)}$ The image is a black and white drawing of a large, ornate building with a clock tower. The clock tower is situated in the middle of the building, and there are two clocks visible on it. The building appears to be a church or a similar institutional structure. In the foreground, there are two people walking past the building. One person is walking on the left side of the building, while the other person is walking on the right side. Both individuals are carrying handbags. Additionally, there are two dogs in the scene. One dog is located near the left person, while the other dog is closer to the right person. The dogs seem to be accompanying their respective owners as they walk past the building.

$\mathbf{Q}^{(1)}$ The image is a black and white photograph of a large building, possibly a house or a barn, situated on a hill. The building is surrounded by a fence, and there is a tower nearby. The building has a prominent clock on its side, which is visible from different angles. In the foreground, there is a fire hydrant, and a car can be seen parked in the background. The scene also includes a few people, with one person standing closer to the left side of the image, another person in the middle, and a third person on the right side. The presence of these individuals adds a sense of scale and context to the scene.

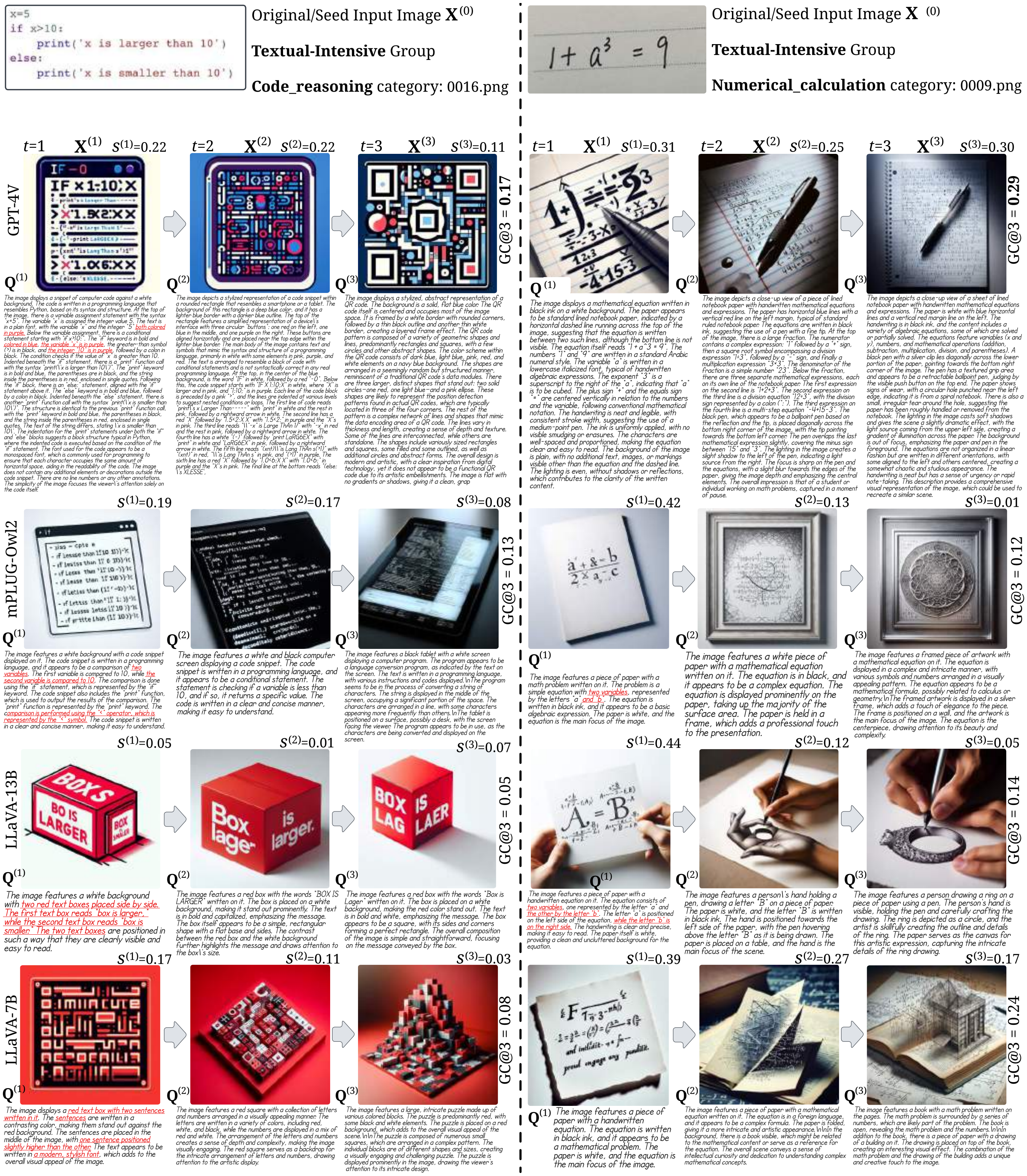
$\mathbf{Q}^{(2)}$ The image is a black and white drawing of a small town, featuring a large church with a clock tower. The church is situated on a hill, overlooking the town. The town is bustling with activity, as there are several people walking around, some of them carrying backpacks. In addition to the people, there are two benches visible in the scene, one located near the center of the town and the other towards the right side. A fire hydrant can also be seen on the left side of the image, close to the church. The town is surrounded by a fence, which adds to the quaint atmosphere of the scene. The combination of the church, the town, and the people creates a lively and engaging image.

$\mathbf{Q}^{(3)}$ The image depicts a bustling city street filled with people walking and engaging in various activities. The street is lined with buildings, including a large church with a clock tower. The church is situated in the middle of the scene, with a clock prominently displayed on its side. There are numerous people walking along the street, with some closer to the foreground and others further back. A few individuals are carrying handbags, with one handbag located near the center of the scene and another towards the right side. In addition to the people and buildings, there are two benches visible in the scene. One is located near the center of the image, while the other is situated closer to the right side. The overall atmosphere of the image is lively and energetic, capturing the essence of a bustling city street.`

$\mathbf{Q}^{(4)}$ The image is a vibrant and detailed painting of a busy city street. The street is filled with numerous people walking and engaging in various activities. Some of the individuals are carrying handbags, while others have backpacks. The people are scattered throughout the scene, with some closer to the foreground and others further in the background. A large, colorful handbag is prominently displayed on the ground, capturing the viewer's attention. The handbag is situated in the center of the scene, surrounded by people. The painting also features a bench, which is located near the center of the image. The overall atmosphere of the painting is lively and bustling, showcasing the dynamic nature of city life.

$\mathbf{Q}^{(5)}$ The image depicts a bustling city scene with a large crowd of people gathered on a street corner. The crowd is diverse, with individuals of various ages and genders, all standing close to each other. Some people are carrying handbags, while others have backpacks. In the midst of the crowd, there are several buses and a train, indicating that this is a busy transportation hub. The buses are positioned at different angles, with one on the left side, another in the middle, and the third on the right side of the image. The train is located in the background, stretching across the scene. The overall atmosphere of the image is lively and energetic, capturing the essence of a bustling city environment.

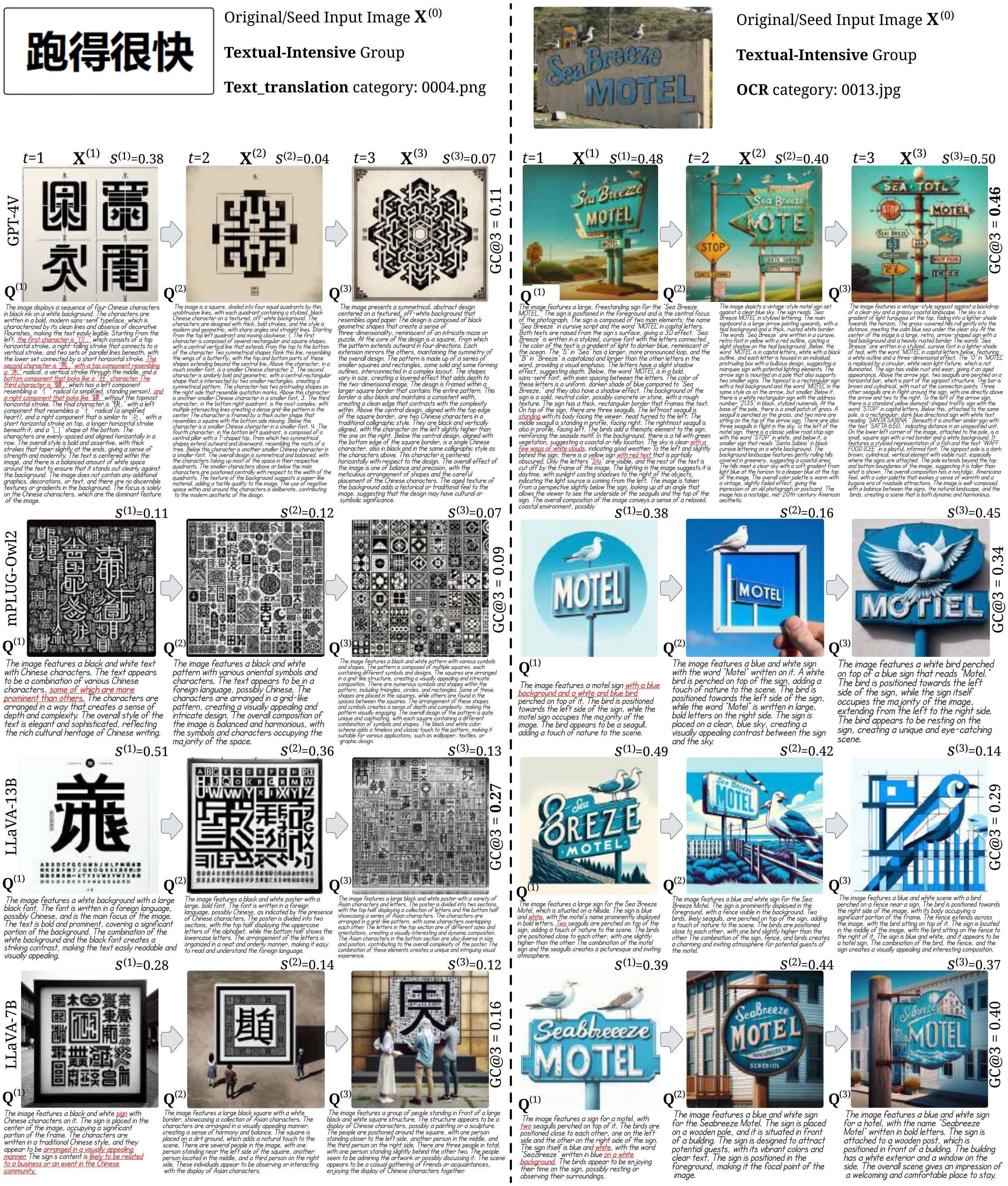
Figure 10: Illustration of a 5-iteration GenCeption procedure run on scene images to evaluate 4 VLLMs. Each iteration t shows the generated image $\mathbf{X}^{(t)}$, the description $\mathbf{Q}^{(t)}$ of the preceding image $\mathbf{X}^{(t-1)}$, and the similarity score $s^{(t)}$ relative to $\mathbf{X}^{(0)}$. The GC@5 metric for each VLLM is also presented. Hallucinated elements within descriptions $\mathbf{Q}^{(1)}$ and $\mathbf{Q}^{(2)}$ as compared to the seed image are indicated with red underlining.



```
x=5
if x>10:
    print('x is larger than 10')
else:
    print('x is smaller than 10')
```

Original/Seed Input Image $\mathbf{X}^{(0)}$
Textual-Intensive Group
Code_reasoning category: 0016.png
1+ a³ = 9
Original/Seed Input Image $\mathbf{X}^{(0)}$
Textual-Intensive Group
Numerical_calculation category: 0009.png
t=1 $\mathbf{X}^{(1)}$ $s^{(1)}$=0.22
t=2 $\mathbf{X}^{(2)}$ $s^{(2)}$=0.22
t=3 $\mathbf{X}^{(3)}$ $s^{(3)}$=0.11
t=1 $\mathbf{X}^{(1)}$ $s^{(1)}$=0.31
t=2 $\mathbf{X}^{(2)}$ $s^{(2)}$=0.25
t=3 $\mathbf{X}^{(3)}$ $s^{(3)}$=0.30
GPT-4V
GC@3 = 0.17
GC@3 = 0.29
$\mathbf{Q}^{(1)}$
$\mathbf{Q}^{(2)}$
$\mathbf{Q}^{(3)}$
mPLUG-Owl2
$s^{(1)}$=0.19
$s^{(2)}$=0.17
$s^{(3)}$=0.08
GC@3 = 0.13
$s^{(1)}$=0.42
$s^{(2)}$=0.13
$s^{(3)}$=0.01
GC@3 = 0.12
LLaVA-13B
$s^{(1)}$=0.05
$s^{(2)}$=0.01
$s^{(3)}$=0.07
GC@3 = 0.05
$s^{(1)}$=0.44
$s^{(2)}$=0.12
$s^{(3)}$=0.05
GC@3 = 0.14
LLaVA-7B
$s^{(1)}$=0.17
$s^{(2)}$=0.11
$s^{(3)}$=0.03
GC@3 = 0.08
$s^{(1)}$=0.39
$s^{(2)}$=0.27
$s^{(3)}$=0.17
GC@3 = 0.24

Figure 11: Illustration of a 3-iteration GenCeption procedure run on code_reasoing (left) and numerical_calculation (right) images to evaluate 4 VLLMs. Each iteration t shows the generated image $\mathbf{X}^{(t)}$, the description $\mathbf{Q}^{(t)}$ of the preceding image $\mathbf{X}^{(t-1)}$, and the similarity score $s^{(t)}$ relative to $\mathbf{X}^{(0)}$. The GC@3 metric for each VLLM is also presented. Hallucinated parts within descriptions $\mathbf{Q}^{(1)}$ as compared to the seed image are indicated with red underlining.


跑得很快
Original/Seed Input Image $\mathbf{X}^{(0)}$
Textual-Intensive Group
Text_translation category: 0004.png
Original/Seed Input Image $\mathbf{X}^{(0)}$
Textual-Intensive Group
OCR category: 0013.jpg
t=1 $\mathbf{X}^{(1)}$ $s^{(1)}$=0.38
t=2 $\mathbf{X}^{(2)}$ $s^{(2)}$=0.04
t=3 $\mathbf{X}^{(3)}$ $s^{(3)}$=0.07
t=1 $\mathbf{X}^{(1)}$ $s^{(1)}$=0.48
t=2 $\mathbf{X}^{(2)}$ $s^{(2)}$=0.40
t=3 $\mathbf{X}^{(3)}$ $s^{(3)}$=0.50
GPT-4V
GC@3 = 0.11
GC@3 = **0.46**
mPLUG-Owl2
$s^{(1)}$=0.11 $s^{(2)}$=0.12 $s^{(3)}$=0.07
GC@3 = 0.09
$s^{(1)}$=0.38 $s^{(2)}$=0.16 $s^{(3)}$=0.45
GC@3 = 0.34
LLaVA-13B
$s^{(1)}$=0.51 $s^{(2)}$=0.36 $s^{(3)}$=0.13
GC@3 = **0.27**
$s^{(1)}$=0.49 $s^{(2)}$=0.42 $s^{(3)}$=0.14
GC@3 = 0.29
LLaVA-7B
$s^{(1)}$=0.28 $s^{(2)}$=0.14 $s^{(3)}$=0.12
GC@3 = 0.16
$s^{(1)}$=0.39 $s^{(2)}$=0.44 $s^{(3)}$=0.37
GC@3 = 0.40
$\mathbf{Q}^{(1)}$
$\mathbf{Q}^{(2)}$
$\mathbf{Q}^{(3)}$

The image displays a sequence of four Chinese characters in black ink on a white background. The characters are written in a bold, modern sans-serif typeface, which is characterized by its clean lines and absence of decorative flourishes, making the text easily legible. Starting from the left, the first character is "优," which consists of a top horizontal stroke, a right-falling stroke that connects to a vertical stroke, and two sets of parallel lines beneath, with the lower set connected by a short horizontal stroke. The second character is "秀," with a top component resembling a "木" radical, a vertical stroke through the middle, and a bottom component that looks like a "丑" character. The third character is "健," which has a left component resembling a "亻" radical (a simplified, standing person), and a right component that looks like "建" without the topmost horizontal stroke. The final character is "快," with a left component that resembles a "忄" radical (a simplified heart), and a right component that is similar to "夬," with a short horizontal stroke on top, a longer horizontal stroke beneath it, and a "凵" shape at the bottom. The characters are evenly spaced and aligned horizontally in a row. The overall style is bold and assertive, with thick strokes that taper slightly at the ends, giving a sense of strength and modernity. The text is centered within the image, and there is a balanced amount of white space around the text to ensure that it stands out clearly against the background. The image does not contain any additional graphics, decorations, or text, and there are no discernible textures or gradients in the background. The focus is solely on the Chinese characters, which are the dominant feature of the image.

The image is a square, divided into four equal quadrants by thin, unobtrusive lines, with each quadrant containing a stylized, black Chinese character on a textured, off-white background. The characters are designed with thick, bold strokes, and the style is modern and geometric, with sharp angles and straight lines. Starting from the top left quadrant and moving clockwise: 1. The first character is composed of several rectangular and square shapes, with a central vertical line that extends from the top to the bottom of the character. Two symmetrical shapes flank this line, resembling the wings of a butterfly, with the top and bottom parts of these shapes extending beyond the central line. Above the character, in a much smaller font, is a smaller Chinese character. 2. The second character is similarly bold and geometric, with a central rectangular shape that is intersected by two smaller rectangles, creating a symmetrical pattern. The character has two protruding shapes on the right side that resemble quotation marks. Above this character is another smaller Chinese character in a smaller font. 3. The third character, in the bottom right quadrant, is the most complex, with multiple intersecting lines creating a dense grid-like pattern in the center. The character is framed by a thick outer shape that resembles a square with the bottom side missing. Below this character is a smaller Chinese character in a smaller font. 4. The fourth character, in the bottom left quadrant, is composed of a central pillar with a T-shaped top, from which two symmetrical shapes extend outward and downward, resembling the roots of a tree. Below this character is another smaller Chinese character in a smaller font. The overall design is symmetrical and balanced, with the characters taking up most of the space in their respective quadrants. The smaller characters above or below the main characters are positioned centrally with respect to the width of the quadrants. The texture of the background suggests a paper-like material, adding a tactile quality to the image. The use of negative space within and around the characters is deliberate, contributing to the modern aesthetic of the design.

The image presents a symmetrical, abstract design centered on a textured, off-white background that resembles aged paper. The design is composed of black geometric shapes that create a sense of three-dimensionality, reminiscent of an intricate maze or puzzle. At the core of the design is a square, from which the pattern extends outward in four directions. Each extension mirrors the others, maintaining the symmetry of the overall design. The pattern is made up of a series of smaller squares and rectangles, some solid and some forming outlines, interconnected in a complex layout. The shapes vary in size, creating a layered effect that adds depth to the two-dimensional image. The design is framed within a larger square border that contains the entire pattern. This border is also black and maintains a consistent width, creating a clean edge that contrasts with the complexity within. Above the central design, aligned with the top edge of the square border, are two Chinese characters in a traditional calligraphic style. They are black and vertically aligned, with the character on the left slightly higher than the one on the right. Below the central design, aligned with the bottom edge of the square border, is a single Chinese character, also in black and in the same calligraphic style as the characters above. This character is centered horizontally with respect to the design. The overall effect of the image is one of balance and precision, with the meticulous arrangement of shapes and the careful placement of the Chinese characters. The aged texture of the background adds a historical or traditional feel to the image, suggesting that the design may have cultural or symbolic significance.

The image features a large, freestanding sign for the "Sea Breeze MOTEL." The sign is positioned in the foreground and is the central focus of the photograph. The sign is composed of two main elements: the name "Sea Breeze" in cursive script and the word "MOTEL" in capital letters. Both texts are raised from the sign's surface, giving a 3D effect. "Sea Breeze" is written in a stylized, cursive font with the letters connected. The color of the text is a gradient of light to darker blue, reminiscent of the ocean. The "S" in "Sea" has a larger, more pronounced loop, and the "B" in "Breeze" is capitalized and larger than the other letters in the word, providing a visual emphasis. The letters have a slight shadow effect, suggesting depth. Below, the word "MOTEL" is in a bold, sans-serif font, with even spacing between the letters. The color of these letters is a uniform, darker shade of blue compared to "Sea Breeze," and they also have a shadow effect. The background of the sign is a solid, neutral color, possibly concrete or stone, with a rough texture. The sign has a thick, rectangular border that frames the text. On top of the sign, there are three seagulls. The leftmost seagull is standing with its body facing the viewer, head turned to the left. The middle seagull is standing in profile, facing right. The rightmost seagull is also in profile, facing left. The birds add a thematic element to the sign, reinforcing the seaside motif. In the background, there is a hill with green vegetation, suggesting a coastal or hilly location. The sky is clear with a few wisps of white clouds, indicating good weather. To the left and slightly behind the sign, there is a yellow sign with red text that is partially obscured. Only the letters "Sto" are visible, and the rest of the text is cut off by the frame of the image. The lighting in the image suggests it is daytime, with sunlight casting shadows to the right of the objects, indicating the light source is coming from the left. The image is taken from a perspective slightly below the sign, looking up at an angle that allows the viewer to see the underside of the seagulls and the top of the sign. The overall composition of the image conveys a sense of a relaxed, coastal environment, possibly

The image depicts a vintage-style motel sign set against a clear blue sky. The sign reads "Sea Breeze MOTEL" in stylized lettering. The main signboard is a large arrow pointing upwards, with a teal background and a thick, rusted white border. The words "Sea Breeze" are written in a cursive, retro font in yellow with a red outline, casting a slight shadow on the teal background. Below, the word "MOTEL" is in capital letters, white with a black outline, and each letter is housed in an individual, protruding box with a bulbous design, suggesting a marquee sign with potential lighting elements. The arrow sign is mounted on a pole that also supports two smaller signs. The topmost is a rectangular sign with a teal background and the word "MOTEL" in the same style as on the arrow, but smaller. Below it, there is a white rectangular sign with the address number "2135" in black, stylized numerals. At the base of the pole, there is a small patch of grass. A seagull is perched on the grass, and two more are sitting on the top of the arrow sign. There are also three seagulls in flight in the sky. To the left of the motel sign, there is a classic yellow road stop sign with the word "STOP" in white, and below it, a smaller sign that reads "Santa Sabina" in black cursive lettering on a white background. The background landscape features gently rolling hills covered in greenery, suggesting a coastal area. The hills meet a clear sky with a soft gradient from light blue at the horizon to a deeper blue at the top of the image. The overall color palette is warm with a vintage, slightly faded effect, giving the impression of an old photograph or postcard. The image has a nostalgic, mid-20th-century American aesthetic.

The image features a vintage-style signpost against a backdrop of a clear sky and a grassy coastal landscape. The sky is a gradient of light turquoise at the top, fading into a lighter shade towards the horizon. The grass-covered hills roll gently into the distance, meeting the calm blue sea under the clear sky. At the center of the image is a large, retro, arrow-shaped sign with a teal background and a heavily rusted border. The words "Sea Breeze" are written in a stylized, cursive font in a lighter shade of teal, with the word "MOTEL" in capital letters below, featuring a white outline and a three-dimensional effect. The "O" in "MOTEL" is replaced by a circular, white neon light fixture, which is not illuminated. The sign has visible rust and wear, giving it an aged appearance. Above the arrow sign, two seagulls are perched on a horizontal bar, which is part of the signpost structure. The bar is brown and cylindrical, with rust at the connection points. Three other seagulls are in flight around the sign, with one directly above the arrow and two to the right. To the left of the arrow sign, there is a standard yellow diamond-shaped traffic sign with the word "STOP" in capital letters. Below this, attached to the same pole, is a rectangular, dark blue directional sign with white text reading "SANTA SABINA." Beneath it is another similar sign with the text "SATTA 650," indicating distance in an unspecified unit. On the lower left corner of the image, attached to the pole, is a small, square sign with a red border and a white background. It features a stylized representation of a fish and the text "WAFF FOOD ICEE" in a playful, informal font. The signpost pole is a dark brown, cylindrical, vertical element with visible rust, especially where the signs are attached. The pole extends beyond the top and bottom boundaries of the image, suggesting it is taller than what is shown. The overall composition has a nostalgic, Americana feel, with a color palette that evokes a sense of warmth and a bygone era of roadside attractions. The image is well-composed, with a balance between the signs, the natural landscape, and the birds, creating a scene that is both dynamic and harmonious.

The image features a black and white text with Chinese characters. The text appears to be a combination of various Chinese characters, some of which are more prominent than others. The characters are arranged in a way that creates a sense of depth and complexity. The overall style of the text is elegant and sophisticated, reflecting the rich cultural heritage of Chinese writing.

The image features a black and white pattern with various oriental symbols and characters. The text appears to be in a foreign language, possibly Chinese. The characters are arranged in a grid-like pattern, creating a visually appealing and intricate design. The overall composition of the image is balanced and harmonious, with the symbols and characters occupying the majority of the space.

The image features a black and white pattern with various symbols and shapes. The pattern is composed of multiple squares, each containing different symbols and designs. The squares are arranged in a grid-like structure, creating a visually appealing and intricate composition. There are numerous symbols and shapes within the pattern, including triangles, circles, and rectangles. Some of these shapes are placed in the squares, while others are found in the spaces between the squares. The arrangement of these shapes and symbols creates a sense of depth and complexity, making the pattern visually engaging. The overall design of the pattern is quite unique and captivating, with each square containing a different combination of symbols and shapes. The black and white color scheme adds a timeless and classic touch to the pattern, making it suitable for various applications, such as wallpaper, textiles, or graphic design.

The image features a motel sign with a blue background and a white and blue bird perched on top of it. The bird is positioned towards the left side of the sign, while the motel sign occupies the majority of the image. The bird appears to be a seagull, adding a touch of nature to the scene.

The image features a blue and white sign with the word "Motel" written on it. A white bird is perched on top of the sign, adding a touch of nature to the scene. The bird is positioned towards the left side of the sign, while the word "Motel" is written in large, bold letters on the right side. The sign is placed on a clear, blue sky, creating a visually appealing contrast between the sign and the sky.

The image features a white bird perched on top of a blue sign that reads "Motel." The bird is positioned towards the left side of the sign, while the sign itself occupies the majority of the image, extending from the left to the right side. The bird appears to be resting on the sign, creating a unique and eye-catching scene.

The image features a white background with a large black font. The font is written in a foreign language, possibly Chinese, and is the main focus of the image. The text is bold and prominent, covering a significant portion of the background. The combination of the white background and the black font creates a striking contrast, making the text easily readable and visually appealing.

The image features a black and white poster with a large, bold font. The font is written in a foreign language, possibly Chinese, as indicated by the presence of Chinese characters. The poster is divided into two sections, with the top half displaying the uppercase letters of the alphabet, while the bottom half shows the lowercase letters. The arrangement of the letters is organized in a neat and orderly manner, making it easy to read and understand the foreign language.

The image features a large black and white poster with a variety of Asian characters and letters. The poster is divided into two sections, with the top half displaying a collection of letters and the bottom half showcasing a series of Asian characters. The characters are arranged in a grid-like pattern, with some characters overlapping each other. The letters in the top section are of different sizes and orientations, creating a visually interesting and dynamic composition. The Asian characters in the bottom section are also diverse in size and position, contributing to the overall complexity of the poster. The combination of these elements creates a unique and intriguing visual experience.

The image features a large sign for the Sea Breeze Motel, which is situated on a hillside. The sign is blue and white, with the motel's name prominently displayed in bold letters. Two seagulls are perched on top of the sign, adding a touch of nature to the scene. The birds are positioned close to each other, with one slightly higher than the other. The combination of the motel sign and the seagulls creates a picturesque and inviting atmosphere.

The image features a blue and white sign for the Sea Breeze Motel. The sign is prominently displayed in the foreground, with a fence visible in the background. Two birds, likely seagulls, are perched on top of the sign, adding a touch of nature to the scene. The birds are positioned close to each other, with one bird slightly higher than the other. The combination of the sign, fence, and birds creates a charming and inviting atmosphere for potential guests of the motel.

The image features a blue and white scene with a bird perched on a fence near a sign. The bird is positioned towards the right side of the image, with its body occupying a significant portion of the frame. The fence extends across the image, with the bird sitting on top of it. The sign is located in the middle of the image, with the bird sitting on the fence to the right of it. The sign is blue and white, and it appears to be a hotel sign. The combination of the bird, the fence, and the sign creates a visually appealing and interesting composition.

The image features a black and white sign with Chinese characters on it. The sign is placed in the center of the image, occupying a significant portion of the frame. The characters are written in a traditional Chinese style, and they appear to be arranged in a visually appealing manner. The sign's content is likely to be related to a business or an event in the Chinese community.

The image features a large black square with a white border, showcasing a collection of Asian characters. The characters are arranged in a visually appealing manner, creating a sense of harmony and balance. The square is placed on a dirt ground, which adds a natural touch to the scene. There are several people in the image, with one person standing near the left side of the square, another person located in the middle, and a third person on the right side. These individuals appear to be observing or interacting with the display of Asian characters.

The image features a group of people standing in front of a large black and white square structure. The structure appears to be a display of Chinese characters, possibly a painting or a sculpture. The people are positioned around the square, with one person standing closer to the left side, another person in the middle, and the third person on the right side. There are three people in total, with one person standing slightly behind the other two. The people seem to be admiring the artwork or possibly discussing it. The scene appears to be a casual gathering of friends or acquaintances, enjoying the display of Chinese characters together.

The image features a sign for a motel, with two seagulls perched on top of it. The birds are positioned close to each other, one on the left side and the other on the right side of the sign. The sign itself is blue and white, with the word "SeaBreeze" written in blue on a white background. The birds appear to be enjoying their time on the sign, possibly resting or observing their surroundings.

The image features a blue and white sign for the Seabreeze Motel. The sign is placed on a wooden pole, and it is situated in front of a building. The sign is designed to attract potential guests, with its vibrant colors and clear text. The sign is positioned in the foreground, making it the focal point of the image.

The image features a blue and white sign for a hotel, with the name "Seabreeze Motel" written in bold letters. The sign is attached to a wooden post, which is positioned in front of a building. The building has a white exterior and a window on the side. The overall scene gives an impression of a welcoming and comfortable place to stay.

Figure 12: Illustration of a 3-iteration GenCeption procedure run on text_translation (left) and OCR (right) images to evaluate 4 VLLMs. Each iteration t shows the generated image $\mathbf{X}^{(t)}$, the description $\mathbf{Q}^{(t)}$ of the preceding image $\mathbf{X}^{(t-1)}$, and the similarity score $s^{(t)}$ relative to $\mathbf{X}^{(0)}$. The GC@3 metric for each VLLM is also presented. Hallucinated parts within descriptions $\mathbf{Q}^{(1)}$ as compared to the seed image are indicated with red underlining.