A. Appendix

A.1. Comparison to Other Methods

We also investigate methods like ELLA [36] which replace the CLIP text encoder in SDXL with a LLM based text-encoder (*e.g.*, T5-XL) and use an adapter (470M params in the case of ELLA) to project these features to the original feature space. While both ELLA and `RankDPO` achieve similar performance on T2I-Compbench and DPG-bench as in Tab. 6, we must note that ELLA takes $18\times$ the training time and over $100\times$ images. Moreover, this imposes the additional cost of including the T5/LLaMa model and using the timestep based adapter (470M params) at every timestep, leading to increased inference time. We also compare ELLA and other preference optimization methods in Tab. 7. We see that `RankDPO` trained on `Syn-Pic` provides the best trade-off in terms of training data requirements, computational resources (training time) and downstream performance (as measured with the DPG-bench score). Finally, we do not have comparisons against methods that perform reward fine-tuning, since they need $\tilde{1}00$ A100 days to be applied at a large-scale for the smaller SD1.5 model at 512 resolution and have not been applied sucessfully to the larger SDXL model at 1024 resolution or show minimal benefits in enhancing text-image alignment [38].

A.2. Binary case of `RankDPO` Objective

The binary setting of `RankDPO` ends up with a fixed value for the discount function (since there are only two ranks $1, 2$) and as a result, the only addition is the gain function, which we discuss in Tab. 5.

A.3. `RankDPO` applied to other DPO objectives

A crucial benefit of `RankDPO` is that it can be applied independently on any pairwise objective. More formally, given weight between two image pairs $(\mathbf{x}^i, \mathbf{x}^j)$ as

$$\Delta_{i,j} = |G_i - G_j| \cdot \left| \frac{1}{D(\tau(i))} - \frac{1}{D(\tau(j))} \right|. \tag{8}$$

Then, the generalized objective is written as:

$$\mathcal{L}_{\text{RankGeneralized}}(\boldsymbol{\theta}) = -\mathbb{E}_{(\boldsymbol{c}, \mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^k) \sim \mathcal{D}, t \sim [0,T]} \left[\sum_{i>j} \Delta_{i,j} \cdot \mathcal{L}_{\text{pairwise}} \left(\mathbf{s}(\mathbf{x}^i, \boldsymbol{c}, t, \boldsymbol{\theta}), \mathbf{s}(\mathbf{x}^j, \boldsymbol{c}, t, \boldsymbol{\theta}) \right) \right]$$
$$\text{for } i, j \in \{1, \ldots, k\}, i \neq j \tag{9}$$

Such a formulation would let us extend other preference objectives [35] to include ranking based cues to improve the optimization.

A.4. Detailed Explanation of Evaluations

T2I-Compbench consists of 6000 compositional prompts from 6 different categories (color, shape, texture, spatial, non-spatial, complex). Following the trends of recent protocols [6, 54], the evaluation for these prompts are done using a combination of VQA models, object detectors and vision-language model scores (*e.g.*, CLIPScore [32]).

GenEval consists of 553 prompts comprising different challenges (single object, two objects, counting, position, color, color attribution). These are mostly evaluated using object detectors.

DPG-Bench aggregates prompts from several sources, and lengthens them using LLMs. These prompts on average have 67 words making it extremely challenging for prompt following. The generated images are mostly evaluated using VQA models under the Davidsonian Scene Graph [17] framework. We use the following evaluation metrics for different benchmarks:

- GeneEval. The evaluation for GenEval is performed using the Maskformer [16] object detection models. This is used to determine if the image contains objects specified in the prompts. For color, a CLIP model is used to identify the color of the objects.
- T2I-CompBench: Attribute Binding uses a BLIP-VQA model [47] to ask different (upto 8) questions about the generated images, and is used to validate if the answered questions match the details specified in the prompt.
- T2I-CompBench: Spatial uses a Unidet [92] model to perform object detection to see if the objects in the generated images follow the spatial orientation specified in the prompt.
- T2I-CompBench: Non-Spatial computes the CLIPScore for the prompt and the generated image.
- T2I-CompBench: Complex averages the score computed from Attribute Binding, Spatial, and Complex.
- DPG-Bench: DSG uses the Davidsonian Scene Graph [17] to compute question answer pairs and use a VQA model (mPLUG) [45] to answer the questions before computing the percentage of questions correctly answered.
- DPG-Bench: VQAScore [54] trains a multimodal LLM with a CLIP encoder and Flan-T5 decoder to predict the likelihood of the prompt being appropriate for the image.
- DPG-Bench: Q-Align Aesthetic Score [83] finetunes a multimodal LLM (*e.g.*, LLaVA [57] to predict the aesthetic score of an image from a scale of 0 to 1.

A.5. Cost Analysis.

We provide the estimates for the cost of labeling Pick-a-Picv2 as compared to `Syn-Pic` in Tab. 8. Even excluding the cost of generating 2M images, labeling $\tilde{1}$M pairwise preferences becomes expensive when following standard guidelines of [64] and paying \$0.05 per comparison.

Table 6. Comparison of T2I-Compbench Dataset with DPG-Bench, including model attributes, training time, and inference time increases.

Dataset	**Color**	**Shape**	**Texture**	**Spatial**	**Non-Spatial**	**DPG Score**	**Train Time (A100 Days)**	**Training Data**	**Same Inference Time**
SDXL	58.79	46.87	52.99	21.31	31.19	74.65	-	-	✓
ELLA (SDXL)	72.60	56.34	66.86	22.14	30.69	80.23	112	34M	✗
`RankDPO` (SDXL)	72.33	56.93	69.67	24.53	31.33	79.26	6	0.24M	✓

Table 7. Comparing features of our proposal against baselines that aim to improve T2I model quality post-training. ELLA* also replaces the CLIP text-encoders with T5-XL text-encoder and a 470M parameter adapter applied at each timestep, thereby increasing the inference cost.

Method	Training Images	A100 GPU days	Equal Inference Cost	DPG-Bench Score
DPO	1.0M	30	✓	76.74
MaPO	1.0M	25	✓	74.53
SPO	-	5	✓	74.73
ELLA*	34M	112	✗	80.23
Ours	0.24M	6	✓	79.26

However, in contrast, `Syn-Pic` costs $<$ \$20 for labeling preferences using *five* different reward models, since each of them need only a few hours on a single GPU to label the preferences. We also note that using an LLM like GPT-4o to generate the comparisons would take over \$450 to just process all the images from `Syn-Pic`. Here, the bigger cost is in generating 4 images for the 58K prompts from Pick-a-Picv2, which can still be completed in $<$ \$200.

Table 8. Cost comparison of generating and labelling Pick-a-Picv2 *vs.* `Syn-Pic`

Item	**Pick-a-Picv2**	**`Syn-Pic`**
Number of prompts	58 000	58 000
Number of images	1 025 015	232 000
Number of preferences	959 000	N/A
Image generation cost	N/A	\$185.60
Annotation/Labelling cost	\$47 950.00	< \$20.00
Total cost	\$47 950.00	< \$205.60

A.6. MJHQ-30k Evaluation

We evaluate the FID on MJHQ-30k prompts with the images from Midjourney as reference [46] in Tab. 9. We observe consistent improvements over the reported results for SDXL-refiner, indicating that we are able to generate high-fidelity images.

A.7. Additional Examples

We provide further qualitative comparisons against SDXL (Fig. 6) and other preference optimization methods (Fig. 5) from prompts of DPG-Bench. We see improved prompt following: specifically objects mentioned in the prompt which can easily be missed by SDXL are captured by our model. Further, we also see improved modeling on finer details and relations in the generated images. We also provide an example for SD3-Medium in Fig. 7. In addition to the trends observed before, we observe examples where we are able to fix some deformities in the generations of SD3-Medium.

A.8. Pseudo Code

In Sec. 3, we described our two novel components: (a) `Syn-Pic`, and (b) `RankDPO`. Although we provide a method overview in Fig. 2, for completion we also present the detailed workings of these two components in a procedural manner. Algorithm 1 describes the data collection process given the set of prompts, T2I models, and human preference reward models. Algorithm 2 describes the pseudo code to train a diffusion model using `RankDPO`. It takes as input the ranked preference dataset (`Syn-Pic`), reference model θ_{ref}, initial model θ_{init}, and other hyperparameters that control the training and noise-signal schedule in the diffusion process. Finally, Algorithm 3 combines these two procedures to describe our end-to-end data generation and training process.

A.9. Complete Prompts for Figures

Fig. 1 SDXL

- a vibrant garden filled with an array of colorful flowers meticulously arranged to spell out the word 'peace' on the lush green grass. The garden is enclosed by a white picket fence and surrounded by tall trees that sway gently in the breeze. Above, against the backdrop of a blue sky, whimsical clouds have been shaped to form the word 'tensions', contrasting with the tranquil scene below.
- a striking propaganda poster featuring a cat with a sly expression, dressed in an elaborate costume reminiscent of

Table 9. FID Scores on MJHQ-30k Prompts for 10 Categories and Overall. RankDPO consistently outperforms SDXL-Refiner on 9/10 categories.

Category	Animals	Art	Fashion	Food	Indoor	Landscape	Logo	People	Plants	Vehicles	Overall
SDXL-Refiner	28.93	31.05	28.90	30.09	28.83	30.78	36.67	35.56	**28.42**	24.45	9.55
RankDPO SDXL	**24.37**	**27.22**	**22.91**	**25.02**	**26.01**	**28.61**	**29.74**	**27.54**	30.27	**21.83**	**7.99**

Figure 5. Comparison among different preference optimization methods and RankDPO for SDXL. The results illustrate that we generate images with better prompt alignment and visual quality.

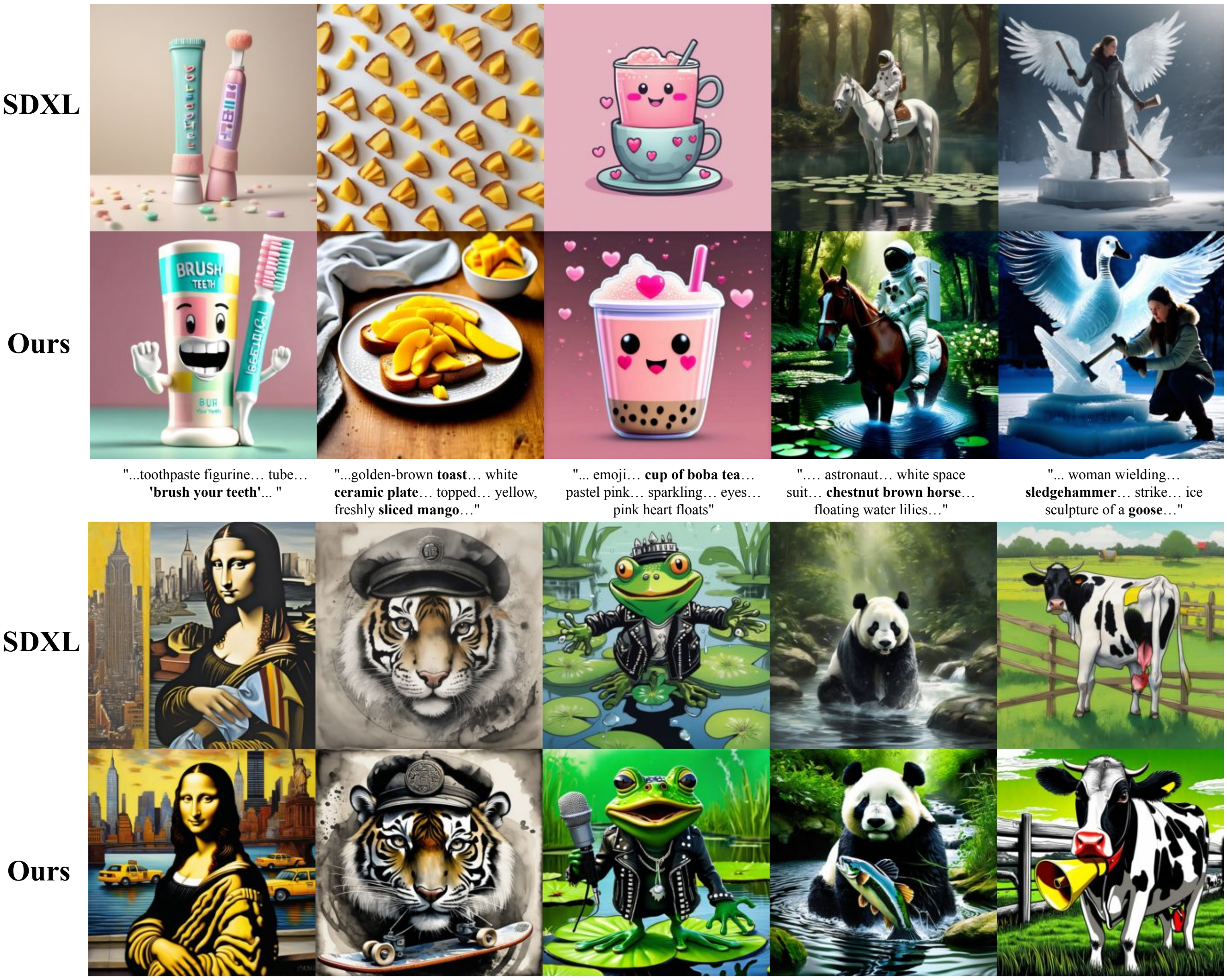

Figure 6. Qualitative comparison between SDXL, before and after our preference-tuning. The results show that our method generates images with better prompt alignment and aesthetic quality.

French Emperor Napoleon Bonaparte. The feline figure is holding a large, yellow wedge of cheese as if it were a precious treasure. The background of the poster is a bold red, with ornate golden details that give it an air of regal authority.

- A deserted park scene illuminated by a soft moonlight where an orange frisbee lies on the grass, slightly tilted to one side. Nearby, a wooden cello and its bow rest in solitude against a weathered park bench, their elegant forms casting long shadows on the pavement. The surrounding trees sway gently in the breeze, indifferent to the forgotten items left in the wake of an earlier emergency rehearsal.
- An anime-style illustration depicts a muscular, metallic tiger with sharp, angular features, standing on a rooftop. The tiger is in a dynamic pose, gripping a sleek, red electric guitar, and its mouth is open wide as if caught in the midst of a powerful roar or song. Above the tiger, a bright spotlight casts a dramatic beam of light, illuminating the scene and creating stark shadows on the surrounding rooftop features.
- A majestic white crane with outstretched wings captured in the act of taking flight from a patch of green grass. In the foreground, an ambulance emblazoned with vibrant red crosses races past, its siren lights ablaze with urgency against the evening sky. The cityscape beyond is silhouetted by the fading hues of dusk, with the outlines of buildings casting long shadows as the day comes to a close.

Fig. 1 SD3

Figure 7. Qualitative comparison between SD3-Medium, before and after our preference-tuning. The results show that our method generates images with better prompt alignment and aesthetic quality.

Algorithm 1 `DataGen`: Generate Synthetically Labeled Ranked Preference Dataset (`Syn-Pic`)

Input: N prompts ($\mathcal{P} = \{\boldsymbol{c}_i\}_{i=1}^{N}$), k T2I Models ($\{\theta_i\}_{i=1}^{k}$), n Reward Models ($\{\mathcal{R}_\psi^i\}_{i=1}^{n}$)
Output: Ranked Preference Dataset $\mathcal{D}$
Initialize: Synthetic dataset $\mathcal{D} = \emptyset$
for $\boldsymbol{c}$ **in** $\mathcal{P}$ **do**
  Generate k images $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^k = \theta_1(\boldsymbol{c}), \theta_2(\boldsymbol{c}), \ldots, \theta_k(\boldsymbol{c})$
  Initialize preference counts $C_i = 0; \;\; \forall i \in \{1, \ldots, k\}$
  for each reward model $\mathcal{R}_\psi^l$ **do**
    Compute scores $R_i^l = \mathcal{R}_\psi^l(\mathbf{x}^i, \boldsymbol{c}); \;\; \forall i \in \{1, \ldots, k\}$
    for each pair (i, j) with $i \neq j$ **do**
      if $R_i^l > R_j^l$ **then**
        Increment preference count $C_i = C_i + 1$
      end if
    end for
  end for
  Compute probabilities $\phi(\mathbf{x}^i) = \frac{C_i}{n \cdot (k-1)}; \;\; \forall i \in \{1, \ldots, k\}$
  Store entry $(\boldsymbol{c}, \mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^k, \phi(\mathbf{x}^1), \phi(\mathbf{x}^2), \ldots, \phi(\mathbf{x}^k))$ in $\mathcal{D}$
end for
return Ranked Preference Dataset $\mathcal{D}$

- A beautifully aged antique book is positioned carefully for a studio close-up, revealing a rich, dark brown leather cover. The words "Knowledge is Power" are prominently featured in the center with thick, flowing brushstrokes, gleaming in opulent gold paint. Tiny flecks of the gold leaf can be seen scattered around the ornately scripted letters, showcasing the craftsmanship that went into its creation. The book is set against a plain, uncluttered background that focuses all attention on the intricate details of the cover's design.
- A pristine white bird with a long neck and elegant feathers stands in the foreground, with a towering dinosaur sculpture positioned behind it among a grove of trees. The dinosaur, a deep green in color with textured skin, contrasts sharply with the smooth plumage of the bird. The trees cast dappled shadows on the scene, highlighting the intricate details of both the bird and the prehistoric figure.
- A striking portrait photograph showcasing a fluffy, cream-colored hamster adorned with a vibrant orange beanie and oversized black sunglasses. The hamster is gripping a small white sign with bold black letters that proclaim "Let's PAINT!" The background is a simple, blurred shade of grey, ensuring the hamster remains the focal point of the image.
- A whimsical scene unfolds in a lecture hall where a donkey, adorned in a vibrant clown costume complete with a ruffled collar and a pointed hat, stands confidently at the podium. The donkey is captured in a high-resolution photo, addressing an audience of attentive students seated in rows of wooden desks. Behind the donkey, a large blackboard is filled with complex mathematical equations, hinting at the serious nature of the lecture juxtaposed with the humorous attire of the lecturer.
- A spacious living room features an unlit fireplace with a sleek, flat-screen television mounted above it. The television screen displays a heartwarming scene of a lion embracing a giraffe in a cartoon animation. The mantle of the fireplace is adorned with decorative items, including a small clock and a couple of framed photographs.

Algorithm 2 `RankDPO`: Ranking-based Preference Optimization using `Syn-Pic`

Input: Ranked Preference Dataset $\mathcal{D}$, Initial model θ_{init}, Reference model θ_{ref}
Input: Pre-defined signal-noise schedule $\{\alpha_t, \sigma_t\}_{t=1}^T$
Hyper-parameters: # Optimization Steps (m), Learning Rate (η), Divergence Control β
Initialize: $\theta = \theta_{\text{init}}$
Output: Fine-tuned model θ^{RankDPO}
for iter $= 0$ **to** m **do**
  Sample entry $(\boldsymbol{c}, \mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^k, \phi(\mathbf{x}^1), \phi(\mathbf{x}^2), \ldots, \phi(\mathbf{x}^k)) \sim \mathcal{D}$
  Sample timestep $t \sim \mathcal{U}(1, T)$, and noise $\boldsymbol{\epsilon}^i \sim \mathcal{N}(0, I)$
  Compute noisy image $\mathbf{x}_t^i = \alpha_t \mathbf{x}^i + \sigma_t \boldsymbol{\epsilon}^i$
  Compute model scores $\mathbf{s}_i \triangleq \mathbf{s}(\mathbf{x}^i, \boldsymbol{c}, t, \boldsymbol{\theta}) = \|\boldsymbol{\epsilon}^i - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t^i, \boldsymbol{c})\|_2^2 - \|\boldsymbol{\epsilon}^i - \boldsymbol{\epsilon}_{\text{ref}}(\mathbf{x}_t^i, \boldsymbol{c})\|_2^2$
  Determine ranking τ by sorting images based on $\phi(\mathbf{x}^i)$ in descending order
  for each pair (i, j) with $i > j$ in τ **do**
    Compute pairwise gains: $G_{ij} = 2^{\phi(\mathbf{x}^i)} - 2^{\phi(\mathbf{x}^j)}$
    Compute discount factors: $D(\tau(i)) = \log(1 + \tau(i))$ and $D(\tau(j)) = \log(1 + \tau(j))$
    Compute pairwise DCG weights: $\Delta_{ij} = |G_{ij}| \cdot \left| \frac{1}{D(\tau(i))} - \frac{1}{D(\tau(j))} \right|$
    Compute pairwise loss: $\mathcal{L}_{ij} = \Delta_{i,j} \log \sigma \left(-\beta \left(\mathbf{s}(\mathbf{x}^i, \boldsymbol{c}, t, \boldsymbol{\theta}) - \mathbf{s}(\mathbf{x}^j, \boldsymbol{c}, t, \boldsymbol{\theta}) \right) \right)$
  end for
  Sum pairwise losses: $\mathcal{L}_{\text{RankDPO}} = -\sum_{i>j} \mathcal{L}_{ij}$
  Compute gradients $\text{grad}_{\text{iter}} = \nabla_\theta \mathcal{L}_{\text{RankDPO}}$
  Update model parameters: $\theta = \theta - \eta \cdot \text{grad}_{\text{iter}}$
end for
Final $\theta^{\text{RankDPO}} = \theta$
return Fine-tuned model θ^{RankDPO}

- An ornate representation of the Taj Mahal intricately positioned at the center of a gold leaf mandala, which showcases an array of symmetrical patterns and delicate filigree. Surrounding the central image, the mandala's design features accents of vibrant blues and reds alongside the gold. Below this striking visual, the words "Place of Honor" are inscribed in an elegant, bold script, centered meticulously at the bottom of the composition.

Fig. 4

- A plump wombat, adorned in a crisp white panama hat and a vibrant floral Hawaiian shirt, lounges comfortably in a bright yellow beach chair. In its paws, it delicately holds a martini glass, the drink precariously balanced atop the keys of an open laptop resting on its lap. Behind the relaxed marsupial, the silhouettes of palm trees sway gently, their forms blurred into the tropical backdrop.

Algorithm 3 Generate `Syn-Pic` and Train `RankDPO`

Input: N prompts $(\mathcal{P} = \{\boldsymbol{c}_i\}_{i=1}^N)$, k T2I Models $(\{\theta_i\}_{i=1}^k)$, n Reward Models $(\{\mathcal{R}_\psi^i\}_{i=1}^n)$
Input: Initial model θ_{init}, Reference model θ_{ref}, Pre-defined signal-noise schedule $\{\alpha_t, \sigma_t\}_{t=1}^T$
Hyper-parameters: # Optimization Steps (m), Learning Rate (η), Divergence Control β
Output: Fine-tuned model θ^{RankDPO}

// Generate Synthetically Labeled Ranked Preference dataset $\mathcal{D}$ using Algorithm 1
$\mathcal{D} = \texttt{DataGen}(\mathcal{P}, \{\theta_i\}_{i=1}^k, \{\mathcal{R}_\psi^i\}_{i=1}^n)$

// Train θ using Algorithm 2
$\theta^{\text{RankDPO}} = \texttt{RankDPO}(\mathcal{D}, \theta_{\text{init}}, \theta_{\text{ref}}, \{\alpha_t, \sigma_t\}_{t=1}^T, m, \eta, \beta)$

return Fine-tuned model θ^{RankDPO}

- a whimsical scene featuring a bright orange fruit donning a miniature brown cowboy hat with intricate stitching. The orange sits atop a wooden table, its textured peel contrasting with the smooth surface beneath. To the side of the orange, there's a small cactus in a terracotta pot, completing the playful western theme.
- A creative studio photograph featuring tactile text spelling 'hello' with vibrant, multicolored fur that stands out boldly against a pure white background. This playful image is showcased within a unique frame made of equally fluffy material, mimicking the texture of the centerpiece. The whimsical arrangement is perfectly centered, lending a friendly and inviting vibe to the viewer.
- An intricately detailed oil painting depicts a raccoon dressed in a black suit with a crisp white shirt and a red bow tie. The raccoon stands upright, donning a black top hat and gripping a wooden cane with a silver handle in one paw, while the other paw clutches a dark garbage bag. The background of the painting features soft, brush-stroked trees and mountains, reminiscent of traditional Chinese landscapes, with a delicate mist enveloping the scene.
- A vibrant yellow rabbit, its fur almost glowing with cheerfulness, bounds energetically across a sprawling meadow dotted with a constellation of wildflowers. The creature's sizeable, red-framed glasses slip comically to the tip of its nose with each jubilant leap. As the first rays of sunlight cascade over the horizon, they illuminate the dew-draped blades of grass, casting the rabbit's exuberant shadow against the fresh green canvas.
- A whimsical scene unfolds in a lecture hall where a donkey, adorned in a vibrant clown costume complete with a ruffled collar and a pointed hat, stands confidently at the podium. The donkey is captured in a high-resolution photo, addressing an audience of attentive students seated

in rows of wooden desks. Behind the donkey, a large blackboard is filled with complex mathematical equations, hinting at the serious nature of the lecture juxtaposed with the humorous attire of the lecturer.

Fig. 5

- a vibrant garden filled with an array of colorful flowers meticulously arranged to spell out the word 'peace' on the lush green grass. The garden is enclosed by a white picket fence and surrounded by tall trees that sway gently in the breeze. Above, against the backdrop of a blue sky, whimsical clouds have been shaped to form the word 'tensions', contrasting with the tranquil scene below.
- a reimagined version of the Mona Lisa, where the iconic figure is depicted with a brown cowboy hat tilted rakishly atop her head. In her hand, she grips a silver microphone, her mouth open as if caught mid-scream of a punk rock anthem. The background, once a serene landscape, is now a vibrant splash of colors that seem to echo the intensity of her performance.
- A deserted park scene illuminated by a soft moonlight where an orange frisbee lies on the grass, slightly tilted to one side. Nearby, a wooden cello and its bow rest in solitude against a weathered park bench, their elegant forms casting long shadows on the pavement. The surrounding trees sway gently in the breeze, indifferent to the forgotten items left in the wake of an earlier emergency rehearsal.
- An anime-style illustration depicts a muscular, metallic tiger with sharp, angular features, standing on a rooftop. The tiger is in a dynamic pose, gripping a sleek, red electric guitar, and its mouth is open wide as if caught in the midst of a powerful roar or song. Above the tiger, a bright spotlight casts a dramatic beam of light, illuminating the scene and creating stark shadows on the surrounding rooftop features.
- A majestic white crane with outstretched wings captured in the act of taking flight from a patch of green grass. In the foreground, an ambulance emblazoned with vibrant red crosses races past, its siren lights ablaze with urgency against the evening sky. The cityscape beyond is silhouetted by the fading hues of dusk, with the outlines of buildings casting long shadows as the day comes to a close.

Fig. 6

- A digitally rendered image of a whimsical toothpaste tube figurine that boasts a candy pastel color palette. The figurine is set against a soft, neutral background, enhancing its playful charm. On the body of the toothpaste tube, bold letters spell out the reminder 'brush your teeth,' inviting a sense of dental care responsibility. The tube cap is carefully designed to exhibit a realistic, shiny texture, creating a striking contrast with the matte finish of the tube itself.
- A piece of golden-brown toast resting on a white ceramic plate, topped with bright yellow, freshly sliced mango. The mango slices are arranged in a fan-like pattern, and the plate sits on a light wooden table with a few crumbs scattered around. The texture of the toast contrasts with the soft, juicy mango pieces, creating an appetizing snack.
- An intricately designed digital emoji showcasing a whimsical cup of boba tea, its surface a glistening shade of pastel pink. The cup is adorned with a pair of sparkling, heart-shaped eyes and a curved, endearing smile, exuding an aura of being lovestruck. Above the cup, a playful animation of tiny pink hearts floats, enhancing the emoji's charming appeal.
- An intricately designed digital emoji showcasing a whimsical cup of boba tea, its surface a glistening shade of pastel pink. The cup is adorned with a pair of sparkling, heart-shaped eyes and a curved, endearing smile, exuding an aura of being lovestruck. Above the cup, a playful animation of tiny pink hearts floats, enhancing the emoji's charming appeal.
- A surreal image capturing an astronaut in a white space suit, mounted on a chestnut brown horse amidst the dense greenery of a forest. The horse stands at the edge of a tranquil river, its surface adorned with floating water lilies. Sunlight filters through the canopy, casting dappled shadows on the scene.
- a focused woman wielding a heavy sledgehammer, poised to strike an intricately carved ice sculpture of a goose. The sculpture glistens in the light, showcasing its detailed wings and feathers, standing on a pedestal of snow. Around her, shards of ice are scattered across the ground, evidence of her previous strikes.
- A detailed painting that features the iconic Mona Lisa, with her enigmatic smile, set against a bustling backdrop of New York City. The cityscape includes towering skyscrapers, a yellow taxi cab, and the faint outline of the Statue of Liberty in the distance. The painting merges the classic with the contemporary, as the Mona Lisa is depicted in her traditional attire, while the city behind her pulses with modern life.
- An intricate Chinese ink and wash painting that depicts a majestic tiger, its fur rendered in delicate brush strokes, wearing a traditional train conductor's hat atop its head. The tiger's piercing eyes gaze forward as it firmly grasps a skateboard, which features a prominent yin-yang symbol in its design, symbolizing balance. The background of the painting is a subtle wash of grays, suggesting a misty and timeless landscape.
- An animated frog with a rebellious punk rock style, clad in a black leather jacket adorned with shiny metal studs, is energetically shouting into a silver microphone. The frog's vibrant green skin contrasts with the dark jacket, and it stands confidently on a large green lily pad floating on a pond's surface. Around the lily pad, the water is calm, and other pads are scattered nearby, some with

blooming pink flowers.

- A sizable panda bear is situated in the center of a bubbling stream, its black and white fur contrasting with the lush greenery that lines the water's edge. In its paws, the bear is holding a glistening, silver-colored trout. The water flows around the bear's legs, creating ripples that reflect the sunlight.
- In a grassy field stands a cow, its fur a patchwork of black and white, with a bright yellow megaphone attached to its red collar. The grass around its hooves is a lush green, and in the background, a wooden fence can be seen, stretching into the distance. The cow's expression is one of mild curiosity as it gazes off into the horizon, the megaphone positioned as if ready to amplify the cow's next "moo".

Fig 7

- On a rainy day, three umbrellas with bright and varied colors—yellow, red, and blue—are opened wide and positioned upright on a worn, wooden table. Their fabric canopies are dotted with fresh raindrops, capturing the soft, diffused light of a hazy morning. Beside these umbrellas lies a classic round watch with a leather strap and a polished face that reflects the muted light. The watch and umbrellas share the table's space, hinting at a paused moment in a day that has just begun.
- An aerial view of Toronto's skyline dominated by the iconic CN Tower standing tall amongst the surrounding buildings. The image is taken from the window of an airplane, providing a clear, bird's-eye perspective of the urban landscape. Across the image, the words "The CN Tower" are prominently displayed in the playful Comic Sans font. The cluster of city structures is neatly bisected by the glistening blue ribbon of a river.
- A vibrant scene featuring a punk rock platypus, its webbed feet firmly planted on an old tree stump. The creature is clad in a black leather jacket, embellished with shiny metal studs, and it's passionately shouting into a silver microphone. Around its neck hangs a bright red bandana, and the stump is situated in a small clearing surrounded by tall, green grass.
- A sleek gray cat balances on the roof of a polished black car. The car is situated in a driveway, flanked by neatly trimmed hedges on either side. Sunlight reflects off the car's surface, highlighting the cat's poised stance as it surveys its surroundings.
- The iconic Statue of Liberty, with its verdant green patina, stands imposingly with a torch raised high in front of the Big Ben Clock Tower, whose clock face is clearly visible behind it. The Big Ben's golden clock hands contrast against its aged stone façade. In the surrounding area, tourists are seen marveling at this unexpected juxtaposition of two renowned monuments from different countries.
- Two vibrant red jugs are carefully positioned below a trio of open black umbrellas, which stand stark against the backdrop of a grey, stormy sky. The jugs rest on the wet, glistening concrete, while the umbrellas, with their smooth, nylon fabric catching the breeze, provide a sharp contrast in both color and texture. Each umbrella casts a protective shadow over the jugs, seemingly safeguarding them from the impending rain.