A. More Details on Stable Diffusion 3

The text encoder in Stable Diffusion 3 incorporates three language models: CLIP L/14 model $c_\theta^{\text{CLIP-L}}(\cdot)$[40], OpenCLIP bigG/14 model $c_\theta^{\text{CLIP-G}}(\cdot)$[6], and T5-v1.1-XXL model $c_\theta^{\text{T5}}(\cdot)$ [41]. Due to the simplicity of prompts in our work, the T5 model is omitted without performance loss. For a prompt y, the CLIP-based encoders $c_\theta^{\text{CLIP-L}}(\cdot)$ and $c_\theta^{\text{CLIP-G}}(\cdot)$ generate corresponding text embeddings, $c_\theta^{\text{CLIP-L}}(y)$ and $c_\theta^{\text{CLIP-G}}(y)$.

These embeddings are concatenated post-pooling to form a vector conditioning $c_{\text{vec}} \in \mathbb{R}^{2048}$. Additionally, penultimate hidden layer representations from each model are concatenated along the channel dimension, producing a CLIP context conditioning $c_{\text{ctxt}} \in \mathbb{R}^{77\times 2048}$. Our method then exclusively applies operations to c_{vec}.

B. Additional Results

B.1. Additional Comparison Results

We further compare CreTok with other personalization methods designed to produce combination effects through interpolation, such as MagicMix [26] and Diffmorpher [57].

As illustrated in Figure B.1, interpolation-based techniques heavily rely on reference images, which constrains their adaptability. When substantial visual differences exist between the input images, the resulting fusion often appears incoherent, a limitation particularly evident in Diffmorpher.

In contrast, CreTok enables the diffusion model to directly generate combinatorial creativity without relying on image synthesis, producing outputs with enhanced visual coherence and detail.

B.2. Additional Creative Visual Results

We present additional creative images generated by CreTok in Figure B.2, further demonstrating the universality of `<CreTok>` in imparting meta-creativity to diffusion models. For each text pair (t_1, t_2), the corresponding combinatorial creativity is produced using the prompts: "A photo of a `<CreTok>` mixture that resembles t_1 and t_2".

B.3. Additional Styles of Creativity

In addition to the creative image styles presented in Figure 6, Figure B.3 showcases additional results, further illustrating the universality of `<CreTok>`. These results further highlight its seamless integration with natural language, enabling diverse and flexible concept combinations across various styles.

C. More Details on Evaluation

C.1. Prompts Used for GPT-4o Evaluation

We conduct an objective evaluation of the creativity of images generated by CreTok and other methods using GPT-4o, assessing four key dimensions: Conceptual Integration, Alignment with Prompt, Originality, and Aesthetic Quality. The detailed prompts provided to GPT-4o are as follows:

The subject of this evaluation is an image that represents a mixture of a banana and a gorilla. The objective is to assess the creativity of an entity that synthesizes two distinct concepts as delineated in the provided prompt. Accordingly, please evaluate the creativity of images generated by various methodologies for the identical prompt, utilizing the following criteria on a scale from 1 to 10:

1. Conceptual Integration (1-10): This criterion gauges the degree to which the image manifests a coherent and integrated concept, as opposed to merely placing two independent elements side by side. A high score signifies that the elements are intricately merged, creating a new, unified entity.

2. Alignment with Prompt (1-10): This evaluates the extent to which the image conforms to and encapsulates the specific combination of concepts described in the prompt. The image should refrain from including irrelevant elements that detract from the primary concepts. A high score is allocated when the image closely adheres to the specifications of the prompt.

3. Originality (1-10): This assesses the innovativeness of the concept portrayed in the image. The depicted concept should not mimic existing animals, plants, or widely recognized mythical creatures unless specifically mentioned in the prompt. Images that present a distinctive and inventive amalgamation receive a high score.

4. Aesthetic Quality (1-10): This criterion scrutinizes the visual appeal of the image, focusing on color harmony, the balance and arrangement of elements, and the overall visual impact. A high score is awarded when the image is not only conceptually robust but also visually engaging.

In conclusion, based on the aforementioned criteria, provide a comprehensive creative assessment (1-10) and articulate specific justifications for your rating.

C.2. More Details on User Study

This section provides a detailed overview of our User Study. The interface used for the study is illustrated in Figure C.4. We utilize 27 text pairs sourced from the original BASS paper to ensure a fair comparison. Creative images generated by CreTok and other methods are displayed in Figure C.1, C.2, and C.3. This user study involves 50 participants who evaluate and rank the creative outputs for each text pair (1-5). Their responses are summarized in Table C.1.

D. More Details on Proposed *CangJie* Dataset

We introduce *CangJie*, the first dataset specifically designed for the combinatorial creativity proposed in the TP2O task. Named after 仓颉, the creator of Chinese characters, *CangJie* symbolizes the dataset's focus on generat-

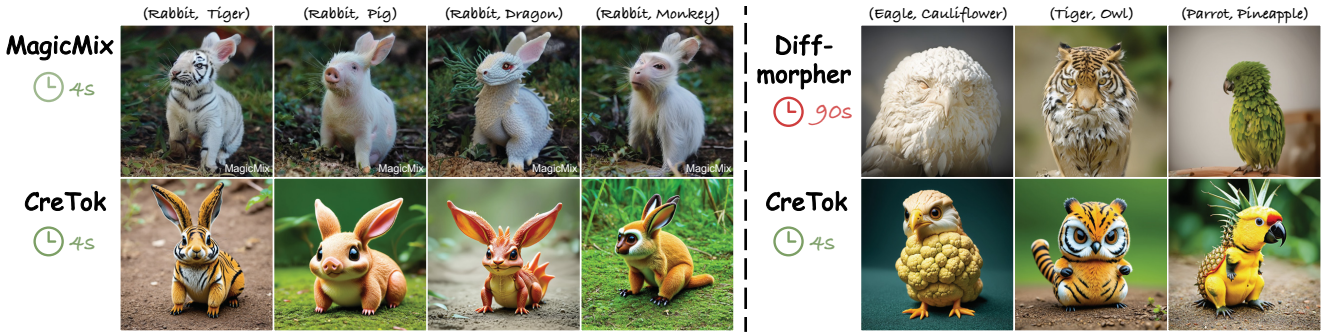

Figure B.1. More visual comparisons of combinatorial creativity. We further compare CreTok with MagicMix [26] and DiffMorpher [57], to further highlight CreTok's superior performance.

Table C.1. Average ranking of each text pair evaluated in the User Study.

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14
Stable Diffusion 3	1.86*	2.50*	2.74*	4.04	3.74	**2.26**	**2.22**	3.04	4.58	2.62*	2.98*	3.66	**2.54**	2.84
Stable Diffusion 3.5	3.52	4.20	3.76	**1.98**	3.02	3.02	2.48*	2.96	2.88	3.50	3.00	2.88	2.90	3.60
Kandinsky 3	3.90	2.64	3.40	2.70*	2.68*	3.00	3.78	**2.26**	**1.70**	2.86	3.30	**2.14**	3.82	**2.54**
BASS	3.90	3.82	2.98	3.52	**2.08**	3.76	2.86	4.38	3.44	3.70	2.98	3.64	2.92	3.44
CreTok	**1.82**	**1.84**	**2.12**	2.76	3.48	2.96*	3.66	2.36*	2.40*	**2.32**	**2.74**	2.68*	2.82*	2.58*

	#15	#16	#17	#18	#19	#20	#21	#22	#23	#24	#25	#26	#27
Stable Diffusion 3	4.28	3.16	3.00	3.78	2.92	4.20	3.40	2.54*	2.56*	4.64	4.18	3.78	4.36
Stable Diffusion 3.5	3.90	**1.88**	3.00	3.32	2.56*	2.12*	3.22	**2.30**	2.76	3.04	3.62	1.96*	3.04
Kandinsky 3	3.08	3.70	3.58	3.50	3.98	3.32	4.14	3.66	3.40	2.52*	2.96	4.28	2.68*
BASS	**1.56**	3.14	**2.64**	**2.14**	3.44	3.38	2.20*	2.74	3.74	2.62	2.94*	3.28	**2.20**
CreTok	2.18*	3.12*	2.78*	2.26*	**2.10**	**1.98**	**2.04**	3.76	**2.54**	**2.18**	**1.30**	**1.70**	2.72

ing novel concepts. *CangJie* comprises 200 text pairs, combining common animals (e.g., dogs, cats, pigs) and plants (e.g., bananas, pineapples, lettuce). Detailed composition of the dataset are presented in Table D.1.

Figure B.2. Additional results of combinatorial creativity generated by CreTok.

Figure B.3. More styles of combinatorial creativity generated by CreTok.

Table D.1. Deatils of *CangJie*.

(Alpaca, Lion)	(Alpaca, Zebra)	(Ant, Cat)	(Ant, Deer)	(Ant, Gibbon)
(Ant, Horse)	(Ant, Rhinoceros)	(Antelope, Gibbon)	(Antelope, Gorilla)	(Antelope, Rhinoceros)
(Antelope, Walrus)	(Antelope, Zebra)	(Apricot, Bear)	(Asparagus, Owl)	(Banana, Giraffe)
(Banana, Seal)	(Bat, Dog)	(Bat, Donkey)	(Bat, Grape)	(Bat, Lobster)
(Bat, Mantis)	(Bat, Octopus)	(Bat, Rhinoceros)	(Bat, Zebra)	(Bear, Elephant)
(Bear, Gorilla)	(Bear, Persimmon)	(Bear, Zebra)	(Bee, Carrot)	(Bee, Fish)
(Bee, Frog)	(Bee, Lizard)	(Bee, Penguin)	(Bee, Pig)	(Bee, Tiger)
(Beetroot, Crab)	(Beetroot, Eagle)	(Beetroot, Octopus)	(Blackberry, Spider)	(Blueberry, Seal)
(Butterfly, Chicken)	(Butterfly, Elephant)	(Butterfly, Fox)	(Butterfly, Giraffe)	(Butterfly, Parrot)
(Butterfly, Squid)	(Butterfly, Zucchini)	(Carrot, Lizard)	(Cat, Elephant)	(Cat, Squirrel)
(Cat, Toad)	(Cat, Zebra)	(Cauliflower, Sheep)	(Cherry, Owl)	(Chicken, Leopard)
(Chicken, Octopus)	(Chicken, Pineapple)	(Chicken, Tiger)	(Coconut, Owl)	(Cow, Giraffe)
(Cow, Hippopotamus)	(Cow, Lion)	(Cow, Lobster)	(Cow, Zebra)	(Crab, Pear)
(Crab, Potato)	(Crab, Raccoon)	(Crab, Sheep)	(Cricket, Cucumber)	(Cricket, Fish)
(Cricket, Leopard)	(Cricket, Lizard)	(Cricket, Octopus)	(Cricket, Shark)	(Cricket, Squid)
(Crocodile, Rhinoceros)	(Crocodile, Turtle)	(Crocodile, Zebra)	(Deer, Dog)	(Deer, Fox)
(Deer, Gibbon)	(Deer, Hippopotamus)	(Deer, Monkey)	(Deer, Whale)	(Deer, Wolf)
(Deer, Zucchini)	(Dog, Gibbon)	(Dog, Lion)	(Dolphin, Elephant)	(Dolphin, Leopard)
(Donkey, Gorilla)	(Donkey, Leopard)	(Donkey, Turtle)	(Donkey, Zebra)	(Duck, Fox)
(Duck, Leopard)	(Duck, Monkey)	(Duck, Parrot)	(Duck, Potato)	(Duck, Wolf)
(Eagle, Grapefruit)	(Eagle, Lizard)	(Eagle, Owl)	(Eagle, Parrot)	(Eagle, Tiger)
(Earthworm, Snake)	(Elephant, Penguin)	(Elephant, Raccoon)	(Falcon, Fox)	(Falcon, Peacock)
(Fish, Kiwi)	(Fish, Parrot)	(Fish, Seal)	(Fish, Shrimp)	(Fish, Tiger)
(Fish, Wolf)	(Flamingo, Giraffe)	(Fox, Giraffe)	(Fox, Leopard)	(Frog, Hippopotamus)
(Frog, Leopard)	(Frog, Tiger)	(Frog, Turtle)	(Frog, Whale)	(Frog, Wolf)
(Garlic, Shark)	(Gibbon, Hippopotamus)	(Gibbon, Owl)	(Gibbon, Tiger)	(Giraffe, Peacock)
(Giraffe, Squirrel)	(Gorilla, Lion)	(Gorilla, Lizard)	(Gorilla, Lobster)	(Hippopotamus, Horse)
(Horse, Leopard)	(Horse, Shrimp)	(Kale, Octopus)	(Kale, Penguin)	(Kiwi, Owl)
(Kiwi, Shark)	(Lemon, Octopus)	(Leopard, Lion)	(Leopard, Pig)	(Leopard, Rhinoceros)
(Leopard, Squirrel)	(Leopard, Turtle)	(Leopard, Whale)	(Leopard, Wolf)	(Lettuce, Sheep)
(Lettuce, Snail)	(Lion, Pineapple)	(Lion, Pumpkin)	(Lizard, Peacock)	(Lizard, Rhinoceros)
(Lizard, Shark)	(Lizard, Turtle)	(Locust, Monkey)	(Locust, Peacock)	(Locust, Raccoon)
(Mantis, Shark)	(Monkey, Owl)	(Monkey, Zebra)	(Moth, Octopus)	(Moth, Penguin)
(Moth, Shrimp)	(Moth, Starfruit)	(Octopus, Peach)	(Octopus, Pineapple)	(Octopus, Potato)
(Octopus, Raccoon)	(Octopus, Rambutan)	(Octopus, Squirrel)	(Octopus, Starfruit)	(Octopus, Watermelon)
(Octopus, Zucchini)	(Ostrich, Owl)	(Owl, Parrot)	(Owl, Raccoon)	(Owl, Strawberry)
(Peach, Penguin)	(Penguin, Pineapple)	(Penguin, Raccoon)	(Pig, Raccoon)	(Pig, Sheep)
(Pig, Squirrel)	(Raccoon, Sheep)	(Raccoon, Spider)	(Raccoon, Walrus)	(Seal, Spider)
(Shark, Wolf)	(Shrimp, Toad)	(Snail, Tiger)	(Snail, Watermelon)	(Snail, Wolf)
(Snail, Zebra)	(Spinach, Wolf)	(Starfruit, Toad)	(Strawberry, Wolf)	(Toad, Turtle)

Figure C.1. Images generated by CreTok and other methods used in the User Study.

Figure C.2. Images generated by CreTok and other methods used in the User Study (continued).

Figure C.3. Images generated by CreTok and other methods used in the User Study (continued).

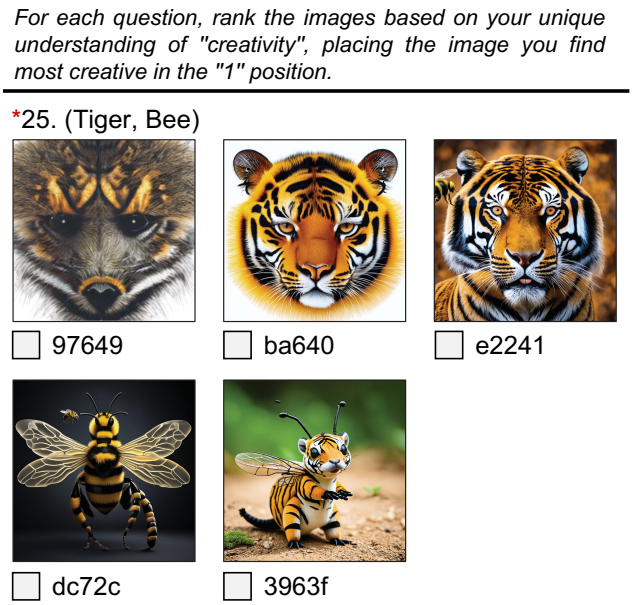

Figure C.4. Interface of the User Study.