REPORT

I have reached the goal of creating a character which is a consistent hybrid between Emma Watson and Taylor Swift.

Task 1: Training Loras

1. Collect images of Emma Watson and Taylor Swift from the internet (about 20 images per character). Do preprocess (detect face, square crop, sanity check)
2. Training Loras:
    a. Since **Emma Watson** and **Taylor Swift** are well-known public figures, the pretrained model might already possess some prior knowledge about them. This makes the training process less challenging to achieve the desired outcome (compared to an ordinary face, like mine or someone else's).
    b. Training on **FLUX.1-dev** base is a strategic choice. While FLUX.1-schnell generates images faster (requiring only about 4 steps compared to the 10-20 steps of FLUX.1-dev), making it more suitable for production environments, FLUX.1-dev serves as an excellent starting point due to its ability to produce more consistent and higher-quality images.
    c. As mentioned earlier, the training process is relatively straightforward, so the focus has shifted to effectively merging two faces. The detailed training parameters are:

Parameters	Value	Comments
dtype	float16	try bfloat16 if needed
input size	1024	
train text_encoder?	No	text_encoder works well with SD. But does not work with FLUX
optimizer	adamw8bit	
learning rate	1e-4	Try drop to 1e-3 if results are not good
step	2000	Try from 500-4000
lora_rank	16	Increase if needed
Resources	RTX-4090 24GB VRAM (personal PC)	

The results are organized into the following folders:

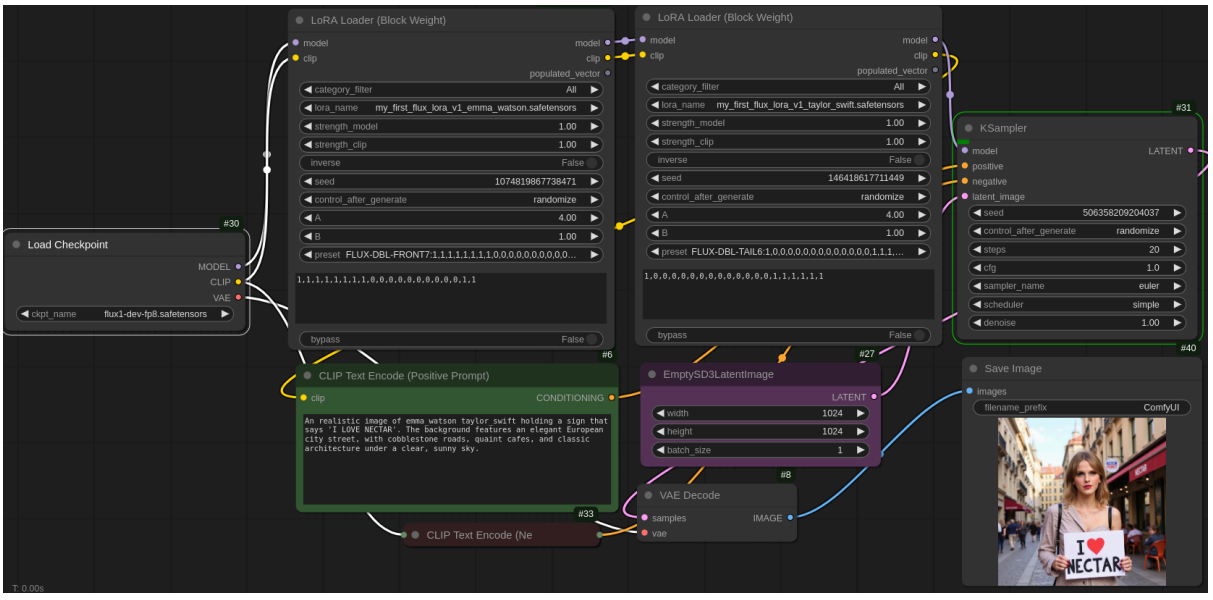
- **images/taylor_swift**: Contains the output images of single Lora of Taylor Swift.
- **images/emma_watson**: Contains the output images of single Lora of Emma Watson.

d. For better merge results, I think we need to focus more on **trigger word** when training Loras

Taylor Swift Lora	Emma Watson Lora	Comments
taylor_swift	emma_watson	Simply name of characters
emma_watson_taylor_swift	emma_watson_taylor_swift	Give the common trigger word for both Loras
zxcvbn	zxcvbn	Strange string, eliminate all references or connections to the names Taylor Swift and Emma Watson.

Task 2: Combine 2 Loras:

1. Install and design workflow for generating images using FLUX. Load Lora using *https://github.com/ltdrdata/ComfyUI-Inspire-Pack* as you recommened.



2. Experiments in Lora Block Weight:
    a. In any case, the weight at the first layer must be 1.
    b. 4 cases:
        i. FLUX-DBL-ALL: 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1
        ii. FLUX-DBL-FRONT7: 1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0
        iii. FLUX-DBL-MID6: 1,0,0,0,0,0,0,0,1,1,1,1,1,1,0,0,0,0,0,0,0
        iv. FLUX-DBL-TAIL6: 1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1
        v. FLUX-DBL-FIRSTHALF: 1,1,1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0,0,1
        vi. FLUX-DBL-SECONDHALF: 1,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,1,1,1

c. The images results will be in folder **images/mixed/{emma_type}-{taylor_type}**
For example: **ALL_FRONT** emma loras use weight FLUX-DBL-ALL and taylor loras use weight FLUX-DBL-FRONT7.
d. Conclusions:

	Emma type	Taylor type	Comments	Mark
ALL_FRONT	FLUX-DBL-ALL	FLUX-DBL-FRONT7	Emma features dominate Taylor features	Bad
ALL_MID	FLUX-DBL-ALL	FLUX-DBL-MID6		
ALL_TAIL	FLUX-DBL-ALL	FLUX-DBL-TAIL6	Balance ratio of 2 features. The consistency is quite good.	Good
FRONT_ALL	FLUX-DBL-FRONT7	FLUX-DBL-ALL	Emma features dominate Taylor features	Bad
MID_ALL	FLUX-DBL-MID6	FLUX-DBL-ALL		
TAIL_ALL	FLUX-DBL-TAIL6	FLUX-DBL-ALL	Balance ratio of 2 features. The consistency is quite good.	Good
FRONT_MID	FLUX-DBL-FRONT7	FLUX-DBL-MID6	The results do not include features of either Emma or Taylor.	Bad
MID_FRONT	FLUX-DBL-MID6	FLUX-DBL-FRONT7		
TAIL_FRONT	FLUX-DBL-TAIL6	FLUX-DBL-FRONT7	The features of TAIL weight dominate the features of FRONT weight	Bad
FRONT_TAIL	FLUX-DBL-FRONT7	FLUX-DBL-TAIL6		
MID_TAIL	FLUX-DBL-MID6	FLUX-DBL-TAIL6		
TAIL_MID	FLUX-DBL-TAIL6	FLUX-DBL-MID6		
SECOND_FIRST	FLUX-DBL-SECONDHALF	FLUX-DBL-FIRSTHALF	Balance ratio of 2 features. But the consistency is not good.	Medium
FIRST_SECOND	FLUX-DBL-FIRSTHALF	FLUX-DBL-SECONDHALF		

3. Experiments in trigger words:
   The demonstration images are in folder **images/trigger_words/emma_watson_taylor_swift** and **images/trigger_words/zxcvbn.** For case different trigger words for each Loras taylor_swift/emma_watson (first row in below table), all the images in section 2 (Lora Block Weight) were generated using this technique, so I will not show them again.

Taylor lora	Emma lora	What I use in prompt	Comments	Mark
taylor_swift	emma_watson	taylor_swift emma_watson	Quality is quite good. The consistency is good. Carefully refine the prompt and conduct tests to ensure the images contain only one face. If the prompt is too basic, there may be cases where the image ends up with two existing faces from the two LoRAs.	Good
emma_watson_taylor_swift	emma_watson_taylor_swift	emma_watson_taylor_swift	The result is good and consistency is quite good (a small percentage that the face will be different). For me, these results are better than using different trigger words.	Good+
zxcvbn	zxcvbn	zxcvbn	The result is good, consistency is good. But I think the face in images are a little more like Taylor. If have more time, I will try to adjust the weight to balance 2 features	Good-

Task 3: Answer the question

1. **What would be the equivalent lora block weight settings for both loras if I just used the emma watson lora in the default way and not the taylor swift lora?**

   You can set the weight for emma watson whatever you want and put all zeros for weight of taylor swift. Or you can turn of “bypass” option in Lora Loader Block of taylor swift.

2. **What would be the equivalent lora block weight settings if I applied both loras naively with weight 0.8?**

   Multiply all numbers in weight with 0.8. Or you can keep weight stable and change the lora strength to 0.8 for both loras

3. **Which character seems to dominate the resemblance of the hybrid character (i.e. where most of the lora block weight settings cause the hybrid to look more like 1 character than the other)**

   For this question, I have already presented in the above section (Experiments in Lora Block Weight and Experiments in trigger words).

Task 4 (bonus): Generate videos

1. The video generate by EchoMimicV2 have bad quality. Even I try with the background on street or the background on studio (same with demo of authors). The animation is quite good but the consistency in video is not good.