ACertains

Strategy **Talk**

Do You Want to fine-tune a SD model?

Summary on ACertains

ACertainModel

Main inference model, prompt input has good accuracy.

ACertainThing

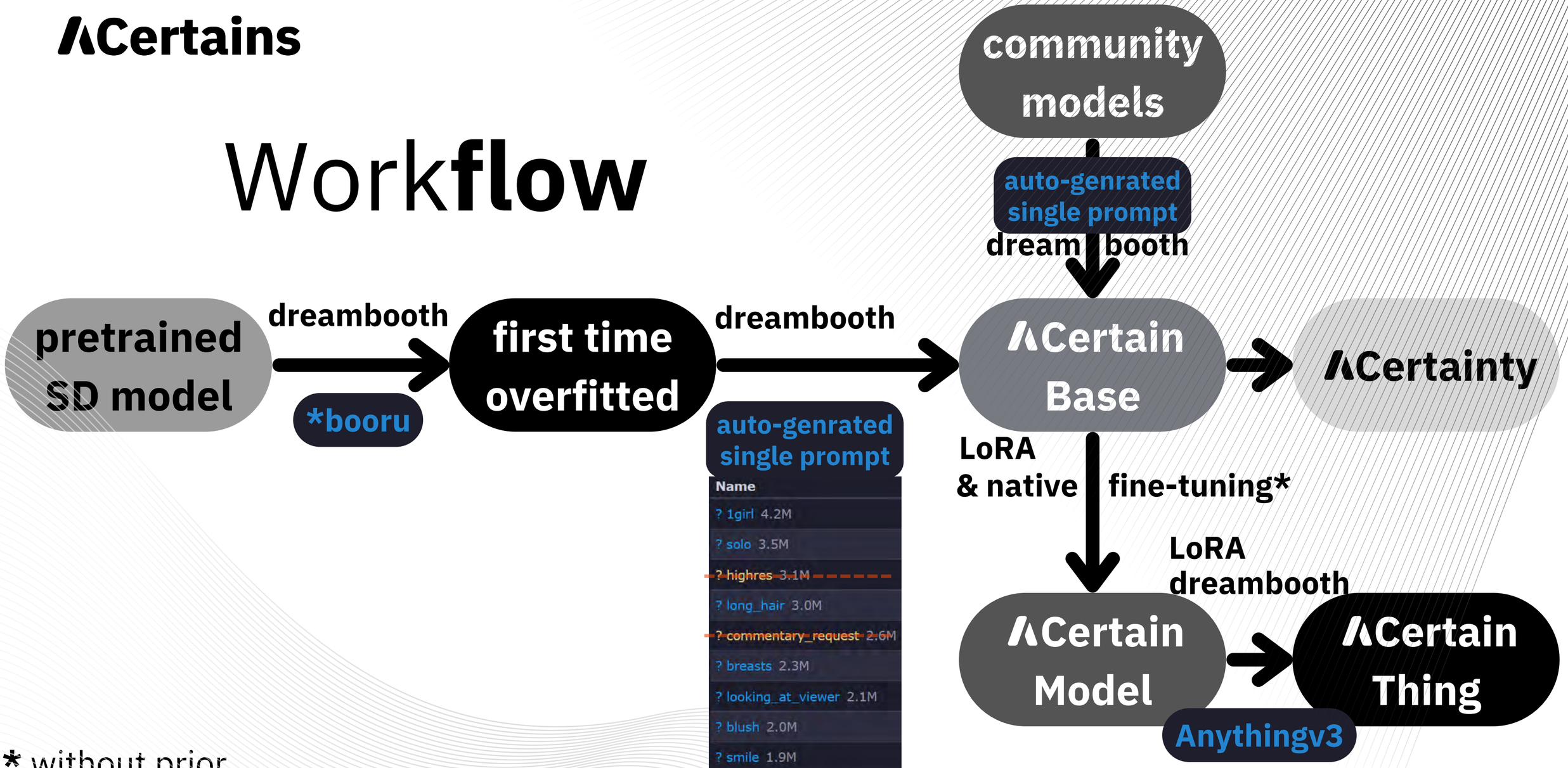
Anythingv3 style (improved for poorly written prompts, but less freedom of creation)

ACertainty

Balanced model better for further Dreambooth and finetuning

ΛCertains

Workflow



* without prior preservation

Certains

Comparable **Analysis**

masterpiece, best quality, 1girl, brown hair, green eyes, colorful, autumn, cumulonimbus clouds, lighting, blue sky, falling leaves, garden
Negative prompt: lowres, bad anatomy, bad hands, text, error, missing fingers, extra digit, fewer digits, cropped, worst quality, low quality, normal quality, jpeg artifacts, signature, watermark, username, blurry, artist name
Steps: 28, Sampler: Euler a, CFG scale: 11, Seed: 114514, Size: 512x768, ENSD: 31337

ACertains

Comparable Analysis

masterpiece, best quality, 1boy, brown hair, green eyes, colorful, autumn, cumulonimbus clouds, lighting, blue sky, falling leaves, garden
Negative prompt: lowres, bad anatomy, bad hands, text, error, missing fingers, extra digit, fewer digits, cropped, worst quality, low quality, normal quality, jpeg artifacts, signature, watermark, username, blurry, artist name
Steps: 28, Sampler: Euler a, CFG scale: 11, Seed: 114514, Size: 512x768, ENSD: 31337

AUTOMATIC1111

Comparable Analysis

masterpiece emphasized

Courtesy of **はかな鳥**

@hcanadli12345

((((masterpiece, best quality, ultra-detailed)))), (official_art, thick_eyebrows, laugh), kawaii, cleavage, (((two side up, white and orangeish and medium streaked hair))), ((tsurime)), Thigh-high socks, Clear vinyl jacket, skindentation, multicolored black bikini, dappled_sunlight, Santorini, geometrical pattern, sport fashion, chaos

Note: the "**masterpiece**" tag is already fine-tuned and it is **not recommended** to emphasize it again.

ACertainThing

Comparable **Analysis**

no masterpiece

Courtesy of **はかな鳥**

@hcanadli12345

((((best quality, ultra-detailed)))), (official_art, thick_eyebrows, laugh), kawaii, cleavage, (((two side up, white and orangeish and medium streaked hair))), ((tsurime)), Thigh-high socks, Clear **vinyl jacket,** skindentation, multicolored black bikini, dappled_sunlight, Santorini, geometrical pattern, sport fashion, chaos

Note: the "**masterpiece**" tag is already fine-tuned and it is **not recommended** to emphasize it again.

SimilarityCalculatior

Relation with other **SD** models

Facts speak louder than words

Calculate the cosine similarity to compare the similarity with other models

SimilarityCalculatior made by *Nyanko Lepsoni* and *RcINS*.
Danke schön!

Base:	SD1.4	SD1.2	NovelAI
SD1.4	100%	71.5%	83.8%
SD1.2	71.5%	100%	67.9%
NovelAI	83.8%	67.9%	100%
CModel	80.9%	74.7%	90.3%
CThing	80.5%	74.7%	89.9%
Certinty	80.9%	74.7%	90.3%
AnythingV3	84.1%	67.7%	98.3%
WD1.3.5	87.8%	68.9%	95.5%
Emis1.0	83.9%	68%	99.9%

Problems and **Proposed Solutions**

1.Language Drift

Fine-tuning a diffusion model on a small set of subject images causes it to lose the ability to generate generic images of the same class and forget the class-specific prior.

Solution 1：Dreambooth use the model's own generated samples by adding a relative weight of the prior-preservation loss. However the ratio of prior-preservation is not easy to determine.

Solution 2：This is a method that requires a lot of GPU time - during the regular training process, we add auto-generated images from the current model with prompt of a single word, with words chosen from a pre-estimated word frequency list randomly according to a certain ratio (we chose our word list from Danbooru Tags). To avoid overfitting, each auto-generated image is used only once.

Problems and **Proposed Solutions**

2.Overfitting

When a model has learned to focus too much on the specific characteristics of the training data, rather than generalizing to new situations.

Solution 1：Dreambooth used prior-preservation loss, and the ratio of prior-preservation is never easy to determine.

Solution 2：It is training data that caused the overfitting. Thus we use a subset of the training data to train an overfitted model, select the previous checkpoint which and use it to generate images by prompt for a single word. These images can be placed in the regular training data according to the word frequency ratio, and the subset of the which that caused the overfitting can be removed, and then retrain the model from the very beginning.

Problems and **Proposed Solutions**

3.Merged Models?

Some merged models have good performance, such as AnythingV3. Should I continue to merge?
This is not scientifically sound and will ultimately result in a model that is overfitted in some cases and normal in others. This model looks good at first, but you will find that it is not faithful to the input prompt and suffers from the problem of language drift mentioned above.

Solution： We can use the method mentioned above to train two models together using a word frequency list with Dreambooth. We can add or replace the training data with the images generated by the model we want to merge, according to the calculated ratio, and maintain a dynamic training dataset during training to prevent overfitting, as mentioned above. We get a balanced model that does not overfit in certain directions. Then choose a checkpoint that is about to be overfitted but not yet as the final version. This type of model is popular in the community because it has good output even under poorly written prompt inputs, such as the **ACertainThing**.

Certains

Future **Outlook**

To be honest, I have already achieved my satisfactory model, and the remaining work will only be labor-intensive if no new transformative methods is introduced.
Such work as adding all the data from Danbooru2022 to the dataset and continuing to train, I think this is something anyone with GPU power and spare time can do. I will further improve it when I have more leisure time.

Certains

Unsolved
Issues & Advice

There are still problems with the portrayal of details such as eyes, hands, and feet, which is largely due to the fact that the LAION dataset used by Stable Diffusion contains a lot of inappropriate data, such as Tom and Jerry, who are not five-fingered creatures.

In the three models I released this time, I tried to eliminate some of the negative effects brought by the "laion-aesthetics v2 5+" used by Stable Diffusion 1.4. Can you believe that in the "laion-aesthetics v2 5+" data, if you search for Apple iPod, you will get many meme images of apples that can be eaten with headphones inserted, with a high aesthetics score. This is obviously wrong; on the other hand, Stable Diffusion 1.2 is a better version.

If you want to continue training a Stable Diffusion model with anime style, please avoid Stable Diffusion 1.4+ version as possible. When you see through its releasing history, you will find that the subsequent versions mostly involve struggles and interests among the founding teams, and there is no good quality assurance. It is not that the higher the version, the better. My personal opinion is that Stable Diffusion 1.2 is the best version.

Certains

Get In **Touch**

Huggingface

https://huggingface.co/JosephusCheung

Twitter

@RealJosephus