**Introduction:**

At first, I started with **Stable-Diffusion-WebUI**, but my laptop has an **integrated Intel Xe Iris GPU**, which does not support CUDA acceleration. This meant I had to run the generation on **CPU mode**, which was extremely slow, making the process inefficient.

To solve this, I switched to **ComfyUI in Google Colab**, where I could take advantage of GPU acceleration for faster image generation and better results.

**1. The Image Generation Process**

**1.1 Understanding Prompt Construction**

One important observation I made is that **prompt structuring** plays a crucial role in generating high-quality, realistic images. There are **two main approaches** to prompt writing:

1. **Natural Language Descriptions:**
   * Example:

"A realistic 4K photograph of a man and woman in a park during sunset, holding hands.

The man has short dark hair, wearing a casual blue shirt and jeans.

The woman has long wavy brown hair, wearing a white summer dress.

The background consists of golden sunlight filtering through trees, with a soft-focus bokeh effect."

* + **Advantages:** Easier to understand and modify.
  + **Disadvantages:** Sometimes, the AI misinterprets the phrasing.

1. **Keyword-based Prompts:**
   * Example:

"realistic photography, man, woman, outdoor park, sunset, cinematic lighting, 4K,

golden hour, soft-focus background, natural skin texture, detailed faces, perfect hands"

* + **Advantages:** More control over the image generation.
  + **Disadvantages:** Might require trial and error to perfect.

**My approach:** I found that using a **hybrid** of both methods gave the best results.

**1.2 Adjusting Key Parameters for Realism**

To achieve a **high-quality** and **realistic** image, I experimented with **various samplers, CFG scales, and steps**. The best configuration I found is:

* **Sampling Method:** DPM++ 2M Karras
* **CFG Scale:** 7–9 (Balances AI creativity and adherence to the prompt)
* **Sampling Steps:** 40–50 (More steps for better details)
* **Resolution:** 768x1024 or 1024x1024

**2. Addressing the Problem of Hands and Faces**

One of the biggest challenges with **AI-generated humans** is **hand anatomy and facial distortions**. Even with a well-crafted prompt, issues like **extra fingers, distorted hands, and asymmetric faces** often occur.

To fix these, I applied **the following enhancement techniques**:

1. **ControlNet (for Poses and Depth)**
   * **OpenPose Model:** control\_v11p\_sd15\_openpose\_fp16 – Ensures correct hand and body positioning.
   * **Depth Model:** control\_v11f1p\_sd15\_depth\_fp16 – Adds proper depth perception for realism.
2. **GFPGAN (Face Restoration)**
   * **Purpose:** Fixes facial distortions, improves symmetry, and enhances eye details.
   * **Settings:** upscale=1, arch="clean", channel\_multiplier=2.
3. **CodeFormer (Fine-Tuning Facial Details)**
   * **Purpose:** Adds sharper facial features and enhances realistic skin texture.
   * **Settings:** fidelity\_weight=0.8 (Balances enhancement and realism).
4. **MeshGraphormer Hand Refiner** (ComfyUI Node)
   * **Purpose:** Corrects hand structure, eliminates extra fingers, and ensures natural hand poses.
   * **Usage:** Applied as a final refinement step to fix hand anatomy.

**3. Final Image Upscaling for Realistic Details**

After ensuring the **structure and anatomy were correct**, I improved the overall image quality using **Real-ESRGAN**:

* **Model Used:** RealESRGAN\_x4plus.pth
* **Purpose:** Enlarges the image **4x** while preserving and enhancing fine details.

Additionally, I enabled **Hires Fix**:

* **Factor:** 1.5–2x
* **Upscaler:** Latent or Real-ESRGAN
* **Steps:** 20–25
* **Purpose:** Reduces blurriness and increases fine details.

**4. Final Summary of My Optimal Configuration**

|  |  |
| --- | --- |
| **Base Model (Checkpoint)** | sd\_xl\_refiner\_1.0.safetensors, epicrealism\_naturalSinRC1VAE.safetensors |
| **ControlNet - Poses** | control\_v11p\_sd15\_openpose\_fp16 |
| **ControlNet - Depth** | control\_v11f1p\_sd15\_depth\_fp16 |
| **Negative Prompt** | deformed anatomy, fused fingers, extra limbs, blurry faces, distorted features, unnatural hand poses |
| **Face Correction (GFPGAN)** | upscale=1, arch="clean", channel\_multiplier=2 |
| **Face Enhancement (CodeFormer)** | fidelity\_weight=0.8 |
| **Hand Fixing (MeshGraphormer)** | Applied after main generation |
| **Upscaling (Real-ESRGAN)** | RealESRGAN\_x4plus.pth, x4 |
| **High-Resolution Fix (Hires Fix)** | Factor: 1.5–2x, Upscaler: Latent/Real-ESRGAN, Steps: 20–25 |
| **Sampling Method** | DPM++ 2M Karras |
| **CFG Scale** | 7–9 |
| **Resolution** | 768x1024 or 1024x1024 |

**Conclusion**

1. I tested **different prompt styles**, concluding that a **hybrid** approach works best.
2. **Hands and faces are the most problematic areas**, but **MeshGraphormer Hand Refiner, GFPGAN, and CodeFormer** helped significantly.
3. **ControlNet ensured anatomical accuracy**, especially for hand poses and depth perception.
4. **Final Upscaling using Real-ESRGAN and Hires Fix improved realism** while avoiding distortion.