

ControlNet Guide: Precise AI Image Control on Your GPU

February 6, 2026 · by Mark Bartlett

[Download this guide as PDF](#)

Quick Answer: ControlNet lets you control the structure of AI-generated images — feed it an edge map, a pose skeleton, or a depth map alongside your text prompt, and the output follows your layout instead of being random. For SD 1.5, you get 14 official ControlNet models covering every control type. For SDXL, use xinsir's ControlNet Union (one model, 10+ control types). For Flux, Shaker Labs Union Pro 2.0 is the current best option. VRAM overhead is roughly 1-2GB for SD 1.5, 2-4GB for SDXL, and you need 24GB total for Flux + ControlNet. ComfyUI is the recommended interface — install `comfyui_controlnet_aux` for preprocessors and drop ControlNet models into `models/controlnet/`.

 **More on this topic:** [Stable Diffusion Locally](#) · [Flux Locally](#) · [ComfyUI vs A1111 vs Fooocus](#) · [VRAM Requirements](#) · [Planning Tool](#)

Every AI image you generate is a dice roll. Same prompt, same settings, completely different composition. You can't tell Stable Diffusion "put the person HERE in THIS pose." The text prompt controls what's in the image, not where or how.

ControlNet fixes that. It takes a structural guide — an edge map, a body pose, a depth map — and forces the diffusion model to follow that structure. Same prompt, but now the output matches the layout you specified. It's the difference between "a person standing in a room" and "a person standing in exactly this pose in exactly this room layout."

This guide covers what ControlNet does, which preprocessors to use, VRAM requirements, setup in ComfyUI and Automatic1111, Flux ControlNet support, and practical workflows that go beyond generating random images.

What ControlNet Does

ControlNet is a neural network that bolts onto a diffusion model (SD 1.5, SDXL, or Flux) and adds spatial conditioning. You provide two inputs:

1. **A text prompt** — what you want in the image
2. **A control image** — the structure you want it to follow

The control image gets processed through a preprocessor (Canny edge detection, OpenPose skeleton, depth estimation, etc.) to create a control map. ControlNet then guides the diffusion process to respect that map while still following your text prompt.

It doesn't replace your model. It adds a layer of structural guidance on top. Your checkpoint, your LoRAs, your sampler settings all still matter – ControlNet just constrains the spatial layout.

Technically, it works by creating a trainable copy of the model's encoder connected through "zero convolutions" – 1x1 convolution layers initialized to zeros so it starts with zero influence and gradually learns the conditioning. The original model weights stay locked. This is why ControlNet models are roughly half the size of the base model they were trained on.

Control Types: What Each Preprocessor Does

Each preprocessor extracts different structural information from a reference image. Here's what actually matters:

Preprocessor	What It Extracts	Best For	Use When
Canny	Sharp edge outlines from contrast boundaries	Architecture, products, clean line art	You want to preserve exact edges and outlines
OpenPose / DWPose	Human body skeleton (joints, limbs, optionally face and hands)	Character poses, figure drawing	You want a person in a specific pose
Depth (Depth Anything V2)	Grayscale depth map (light = close, dark = far)	Room layouts, scene composition, 3D feel	You want to preserve spatial depth and perspective
Lineart	Clean line drawings from photos	Illustrations, coloring book style, manga	You want a drawn/illustrated version of a photo
Scribble	Rough sketch interpretation	Quick concept art from rough drawings	You drew something by hand and want it polished
MLSD	Straight architectural lines only	Buildings, interiors, man-made structures	You need precise geometric structure
HED / SoftEdge	Soft, smooth edge outlines	Style transfer, recoloring	You want softer structural guidance than Canny

Preprocessor	What It Extracts	Best For	Use When
Normal Map	Surface orientation (RGB-encoded surface normals)	3D-to-2D rendering, surface detail	You're working with 3D models or need surface info
Segmentation	Color-coded region labels (sky, person, building)	Scene composition control	You want to control which regions contain what
Tile	Processes image in chunks for detail enhancement	Upscaling, adding detail at same resolution	You want to upscale while adding new detail
Inpaint	Masked regions for selective regeneration	Fixing specific areas of an image	You want to change only part of an image

The most commonly used: Canny, OpenPose/DWPose, and Depth cover about 80% of ControlNet use cases. Start with these three.

Related Tools (Not Technically ControlNet)

IP-Adapter extracts style and content features from a reference image. Where ControlNet controls spatial structure (where things are), IP-Adapter controls appearance (what things look like). They're commonly used together – IP-Adapter for face/style consistency, ControlNet for pose/composition.

T2I-Adapter is a lightweight alternative to ControlNet at only 158MB per model (vs ControlNet's ~1.4GB). It runs the conditioning once for the entire generation instead of every step. Less precise control, but uses 93% less storage and minimal extra VRAM.

VRAM Requirements

ControlNet loads alongside your base model, so you need enough VRAM for both.

Setup	Base Model VRAM	ControlNet Adds	Total Needed	Fits On
SD 1.5 + 1 ControlNet (512x512)	4-6 GB	~1-2 GB	6-8 GB	8 GB GPUs (RTX 3060, 4060)
SD 1.5 + 3 ControlNets (512x512)	4-6 GB	~4-6 GB	10-12 GB	12 GB GPUs (RTX 3060 12GB)

Setup	Base Model VRAM	ControlNet Adds	Total Needed	Fits On
SDXL + 1 ControlNet (1024x1024)	8-10 GB	~2-4 GB	10-14 GB	12-16 GB GPUs
Flux + 1 ControlNet (1024x1024)	~18 GB	~2-3 GB	20-22 GB	24 GB GPUs (RTX 3090, 4090)

Key points:

- Each additional ControlNet stacks. Three ControlNets can add 4-6GB on SD 1.5.
- **T2I-Adapter** adds only ~0.15 GB – use it instead of ControlNet on tight VRAM budgets.
- **Flux on lower VRAM:** Use a GGUF-quantized Flux base model (Q5_K_M fits 8GB GPUs at 1024x1024 with text encoder offloading), which frees room for a ControlNet model. Quality trades off against VRAM savings.
- Both A1111 and ComfyUI have low-VRAM modes that offload ControlNet processing at the cost of speed.

For a complete breakdown of what fits on each GPU tier, see our [VRAM requirements guide](#). For GPU buying advice, see the [GPU buying guide](#).

Setup: ComfyUI (Recommended)

ComfyUI handles ControlNet better than A1111 for most users – its node-based workflow makes multi-ControlNet setups cleaner and it manages VRAM more efficiently. For a full comparison, see our [ComfyUI vs A1111 vs Fooocus guide](#).

Step 1: Install Custom Nodes

If you have ComfyUI Manager installed (you should), open ComfyUI and click **Manager > Install Missing Custom Nodes**. Search for and install:

Package	What It Does
comfyui_controlnet_aux (Fannovel16)	All preprocessors: Canny, OpenPose, Depth Anything, Lineart, MLSD, Scribble, etc.
ComfyUI-Advanced-ControlNet	Per-layer weights, soft weights, advanced multi-ControlNet conditioning
x-flux-comfyui (XLabs-AI)	Required for XLabs Flux ControlNet models specifically

Restart ComfyUI after installing.

Step 2: Download ControlNet Models

Place `.safetensors` files in `ComfyUI/models/controlnet/`.

For SD 1.5 – download from [llyasviel/ControlNet-v1-1](#) on HuggingFace. Start with:

- `control_v11p_sd15_canny.pth` (~1.4 GB)
- `control_v11p_sd15_openpose.pth` (~1.4 GB)
- `control_v11f1p_sd15_depth.pth` (~1.4 GB)

For SDXL – get [xinsir/controlnet-union-sdxl-1.0](#). One model handles 10+ control types.

For Flux – get [Shakker-Labs/FLUX.1-dev-ControlNet-Union-Pro-2.0](#) (~4 GB, FP8 version available for lower VRAM).

Step 3: Build a Basic Workflow

The core ControlNet workflow in ComfyUI:

Load Image → Preprocessor → Apply ControlNet → KSampler → Save Image

1. **Load Image** node – your reference/control image
2. **AIO Preprocessor** or specific preprocessor (Canny, DWPose, etc.) – generates the control map
3. **Load ControlNet Model** node – loads your `.safetensors` ControlNet file
4. **Apply ControlNet** node – connects control map + model to conditioning
5. **KSampler** – generates with normal text prompt + ControlNet conditioning

The Apply ControlNet node connects to your positive conditioning (the same conditioning your text prompt feeds into). The ControlNet and text prompt work together – structure from ControlNet, content from your prompt.

Setup: Automatic1111

Step 1: Install the Extension

1. Open A1111 > **Extensions** tab > **Install from URL**

2. Paste: `https://github.com/Mikubill/sd-webui-controlnet.git`
3. Click **Install**, then go to **Installed > Apply and restart UI**

A collapsible “ControlNet” section appears in the txt2img and img2img tabs.

Step 2: Download Models

Place ControlNet model files in:

```
stable-diffusion-webui/extensions/sd-webui-controlnet/models/
```

Same models as listed in the ComfyUI section above. Make sure ControlNet model versions match your base checkpoint – SD 1.5 ControlNets only work with SD 1.5 models, SDXL ControlNets with SDXL.

Step 3: Use the ControlNet Panel

1. Upload your reference image in the ControlNet section
2. **Enable** the checkbox
3. Select a **Preprocessor** (Canny, OpenPose, etc.)
4. Select the matching **Model** from the dropdown
5. Check **Pixel Perfect** – this auto-sets the preprocessor resolution to match your output
6. Click **Generate**

Key settings:

- **Control Weight** (0-2, default 1.0): How strongly ControlNet influences the output. 0.7-1.0 is the sweet spot for most tasks.
- **Starting/Ending Control Step** (0-1): When ControlNet applies during generation. Ending at 0.8 lets the model “clean up” in the final steps.
- **Control Mode**: “Balanced” for most use cases. “ControlNet is More Important” when you need strict adherence to the structure.

Multi-ControlNet

A1111 supports up to 10 simultaneous ControlNets. Configure the max in **Settings > ControlNet > Multi ControlNet: Max models amount**. Each unit gets its own preprocessor, model, weight, and control mode.

ControlNet for Flux

Flux ControlNet is still maturing but usable. SD 1.5 has the most complete and battle-tested ControlNet ecosystem (14 official models). Flux is catching up.

Current Best Options

Model	Developer	Control Types	Size	Notes
Union Pro 2.0	Shakker Labs	Canny, soft edge, depth, pose, gray	3.98 GB	Best all-in-one for Flux. FP8 version available.
XLabs v3	XLabs-AI	Canny, depth, HED (separate models)	1.49 GB each	Best single-mode quality.
Union	InstantX	Canny, tile, depth, blur, pose, gray, lq	6.6 GB	Beta. More modes but larger.
Flux Tools	Black Forest Labs	Canny, depth	Varies	Official, not technically ControlNet.

Recommended settings for Shakker Labs Union Pro 2.0:

- Conditioning scale: 0.7-0.9 (lower than SD 1.5 ControlNet's typical 1.0)
- Guidance end: 0.65-0.8 depending on control type

In ComfyUI: Use the `x-flux-comfyui` nodes for XLabs models, or native ComfyUI Flux ControlNet nodes for InstantX/Shakker Labs models.

The reality: If ControlNet is your primary workflow, SD 1.5 still offers the most control types, the smallest model files, the most community resources, and the lowest VRAM requirements. Flux produces better base images but has fewer ControlNet options. SDXL sits in the middle – xinsir's Union model covers most use cases in one file.

Practical Workflows

Pose Matching

Turn a reference photo into a completely different style while keeping the exact pose.

1. Load your reference photo (person in desired pose)
2. Run through **OpenPose** or **DW Pose** preprocessor → skeleton map

3. Apply ControlNet with the pose map
4. Write a prompt for the style you want: “oil painting of a knight in armor, dramatic lighting”
5. Generate – output follows the skeleton but renders in your chosen style

Pro tip: Combine IP-Adapter (for face/style consistency) with ControlNet OpenPose (for pose). This gives you the same character in different poses – useful for character sheets and comic panels.

Architecture and Interior Design

Preserve room layout while completely changing the style.

1. Photograph your room or building
2. Run through **Depth** preprocessor → depth map (preserves spatial layout)
3. Optionally add a second ControlNet with **Canny** for edge detail
4. Prompt: “modern minimalist living room, warm lighting, hardwood floors”
5. Generate – same room dimensions and furniture placement, different style

MLSD works better than Canny for buildings specifically – it detects straight architectural lines and ignores organic details.

Sketch to Polished Image

Turn a rough hand drawing into a polished result.

1. Draw your concept on paper or in a drawing app (doesn't need to be good)
2. Use **Scribble** preprocessor for rough sketches, **Lineart** for cleaner drawings
3. Apply ControlNet
4. Describe the final result in your prompt
5. Generate – the AI follows your composition and structure

This is the lowest-barrier ControlNet workflow. A stick figure with basic shapes becomes a fully rendered scene.

Upscaling with Tile ControlNet

Add real detail when upscaling (not just blur removal).

1. Start with your low-resolution image
2. Upscale 2-4x with a standard upscaler (ESRGAN, Real-ESRGAN)
3. Run through **Tile** ControlNet – it processes in 512x512 chunks, adding new detail to each tile

4. Result: higher resolution with genuinely new detail, not just interpolation

In A1111, combine with the **Ultimate SD Upscale** script for automatic tile processing. In ComfyUI, use a tiled processing workflow.

QR Code Art

Generate artistic QR codes that are functional and visually appealing.

1. Generate a standard QR code
2. Use it as the control image with a dedicated QR ControlNet model (`DionTimmer/controlnet_qrcode`)
3. Set weight to 1.0-1.5 (high enough to keep the code scannable)
4. Prompt with your desired art style
5. Test scannability – lower the weight if the code breaks, raise it if the art overwhelms the pattern

Troubleshooting

“ControlNet Has No Effect”

The most common beginner issue. Check these in order:

1. **Is ControlNet enabled?** The checkbox must be checked.
2. **Is the preprocessor set correctly?** If set to “none” without a pre-processed control map, nothing happens.
3. **Does the ControlNet model match the preprocessor?** A Canny model needs a Canny preprocessor (or pre-generated Canny map).
4. **Does the ControlNet model match your checkpoint?** SD 1.5 ControlNet + SDXL checkpoint = error or no effect.
5. **Is the control weight above zero?** Default 1.0 is fine. Below 0.3 you’ll barely see an effect.
6. **Is the preprocessor resolution correct?** In A1111, check “Pixel Perfect” to auto-match. In ComfyUI, ensure the preprocessor output resolution matches your generation resolution.

Model Version Mismatch Errors

```
RuntimeError: mat1 and mat2 shapes cannot be multiplied
```

This means you're loading a ControlNet model built for a different base model architecture. SD 1.5 ControlNets are ~1.4GB, SDXL ControlNets are larger. They're not interchangeable.

VRAM Out of Memory

- **Reduce active ControlNets** – each one adds 1-4GB
- **Enable Low VRAM mode** in ControlNet settings (A1111) or use `--medvram-sdxl`
- **Switch to T2I-Adapters** where possible – 93% less memory
- **Lower the generation resolution** – VRAM scales with resolution
- **ComfyUI handles multi-ControlNet more efficiently** than A1111 for memory management

Weak Control / Output Ignoring Structure

- **Raise the control weight** from 1.0 to 1.2-1.5 (above 1.5 may cause artifacts)
- **Use "ControlNet is More Important" mode** in A1111
- **Increase the ending control step** – at 0.5 the model has half the steps to deviate
- **Check your preprocessor output** – preview the control map before generating. If the map is messy, the output will be messy.

The Bottom Line

ControlNet turns text-to-image generation from "hope for the best" into "follow this structure." Three preprocessors cover most needs: Canny for edges, OpenPose/DWPose for poses, and Depth for spatial layout.

Start here:

1. Install [ComfyUI](#) with `comfyui_controlnet_aux`
2. Download 3 ControlNet models matching your base model (Canny, OpenPose, Depth)
3. Load a reference photo, pick a preprocessor, write a prompt, generate

For SD 1.5 (6-8GB VRAM): The complete ecosystem. 14 official models, smallest files, most tutorials online.

For SDXL (12-16GB VRAM): Use xinsir's ControlNet Union – one model for 10+ control types.

For Flux (24GB VRAM): Shakker Labs Union Pro 2.0 is the best all-in-one. XLabs v3 individual models for highest single-mode quality.

If you're already running [Stable Diffusion locally](#) or [Flux](#), ControlNet is the single biggest upgrade to your workflow. It takes 10 minutes to install and immediately makes every generation more useful.

Related Guides

- [Stable Diffusion Locally: Getting Started](#)
 - [Flux Locally: Complete Guide](#)
 - [ComfyUI vs Automatic1111 vs Fooocus](#)
 - [VRAM Requirements for Local LLMs](#)
 - [GPU Buying Guide for Local AI](#)
 - [What Can You Run on 8GB VRAM](#)
 - [AI Art Styles & Workflows Guide](#)
 - [Local AI Planning Tool – VRAM Calculator](#)
-

Source: <https://insiderllm.com/guides/controlnet-guide-beginners/>

Free guides for running AI locally