



扩散模型 (Diffusion Model)

作者: Calvin

QQ: 179209347

Mail: 179209347@qq.com

介绍

笔记简介:

- 面向对象: 深度学习初学者
- 依赖课程: **线性代数, 统计概率**, 优化理论, 图论, 离散数学, 微积分, 信息论

知乎专栏:

<https://zhuanlan.zhihu.com/p/693738275>

Github & Gitee 地址:

https://github.com/mymagicpower/AIAS/tree/main/deep_learning

https://gitee.com/mymagicpower/AIAS/tree/main/deep_learning

* 版权声明:

- 仅限用于个人学习
- 禁止用于任何商业用途

ControlNet

ControlNet 是一种用于增强图像生成的神经网络技术，特别适用于稳定扩散模型（Stable Diffusion Model）。它的目的是在不改变生成模型核心结构的情况下，增加对生成内容的控制。

ControlNet 的主要思想是通过增加额外的控制信号来指导生成过程。这些控制信号可以是图像的边缘、草图、深度图或其他形态的信息。通过引入这些信号，ControlNet 可以更好地满足用户对生成图像的特定要求。

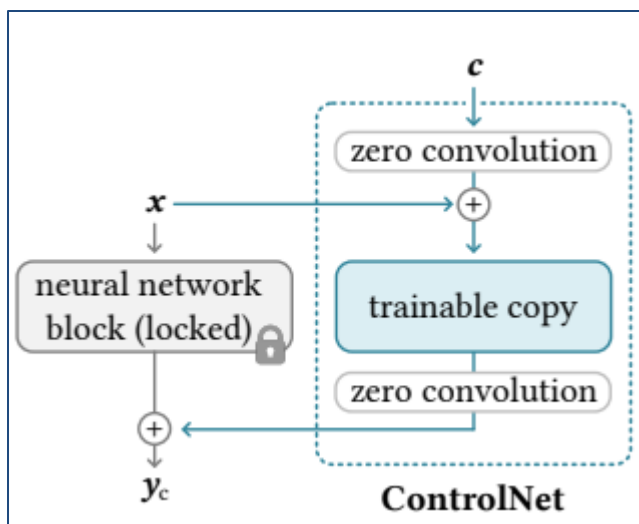


论文: Adding Conditional Control to Text-to-Image Diffusion Models
<https://arxiv.org/pdf/2302.05543>

All rights reserved by www.aias.top , mail: 179209347@qq.com

ControlNet 模型架构

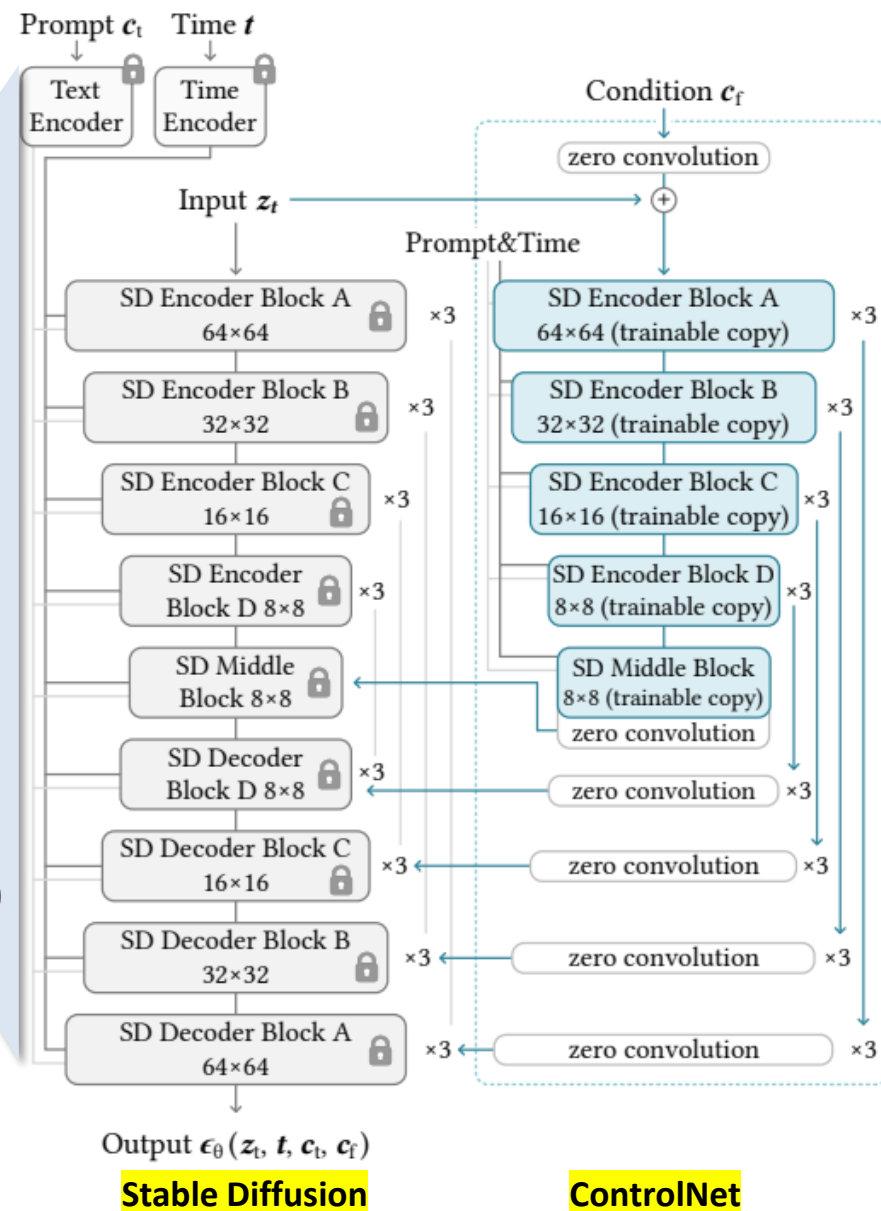
- 首先复制原始稳定扩散模型中 UNet 编码器部分的权重。
- 锁定原始权重以防止过度拟合并方便在较小的数据集上进行训练。
(锁定保留了这些扩散模型生成已在数十亿张图像上训练过的高保真图像的能力。)
- 然后将此可训练副本连接到零卷积,
- 最后将零卷积的输出作为输入发送回冻结块。



$$y_c = \mathcal{F}(x; \theta) + Z(\mathcal{F}(x + Z(c; \theta_{z1}); \theta_c); \theta_{z2})$$

- Z 表示零卷积, 具有不同的参数,
- \mathcal{F} 是原始 UNet 编码器实现的函数。
- C 是用户给出的附加条件。

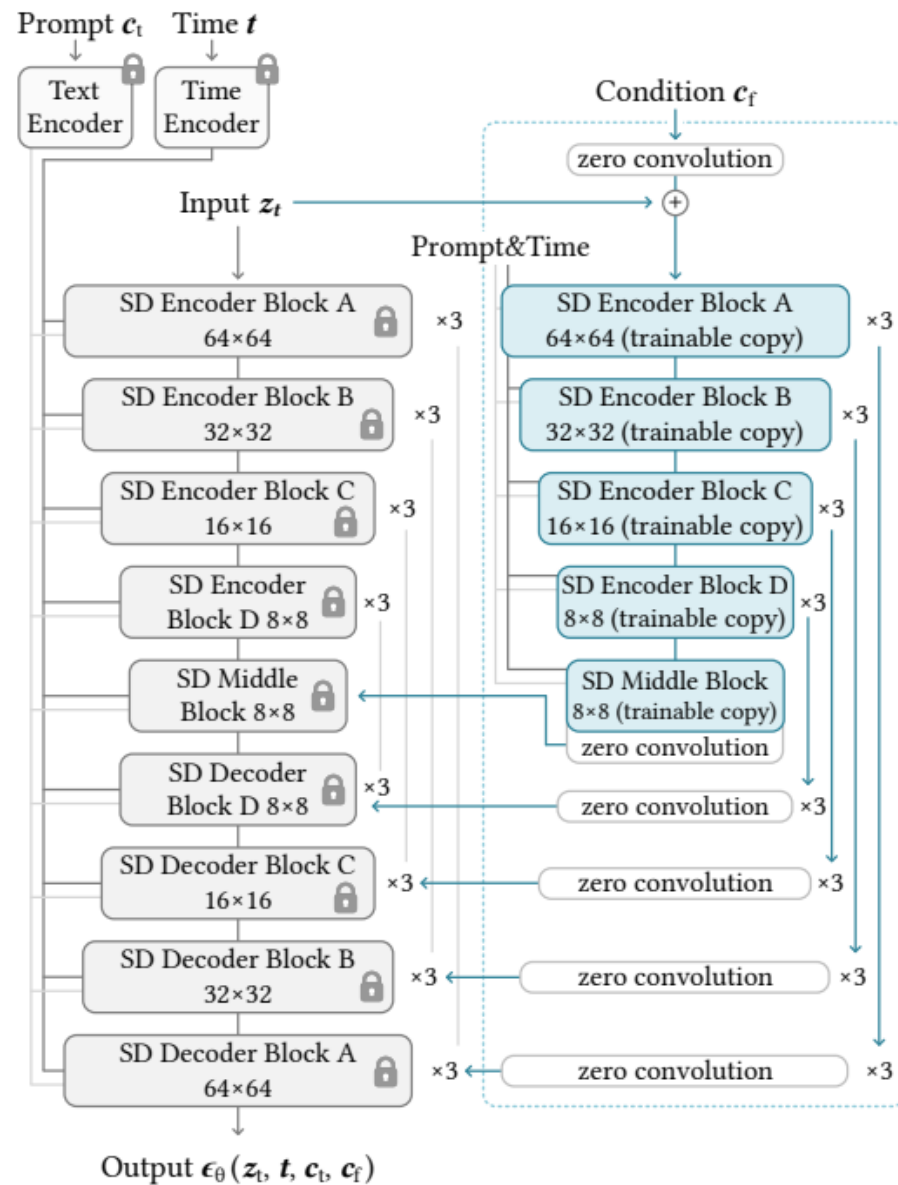
因此, 每当我们训练模型时, 这个可训练副本都会得到更新, 但实际的预训练生成模型仍然保持冻结状态。



ControlNet 模型架构

ControlNet包含了:

- **输入:** 包括Latent Embedding、Time Embedding、Text Embedding以及额外的 Condition。额外的condition是和输入图片一样大小的图像，比如边缘检测图、深度信息图、轮廓图等。
- 12个编码块和1个Stable Diffusion U-Net中间块的“可训练”副本；
- 这12个编码块有 4 种分辨率：分别是 64×64 、 32×32 、 16×16 和 8×8 ；
- 每种分辨率对应3个编码块；
- **输出:** 被添加到 Stable Diffusion U-Net 的 12 个残差结构和 1 个中间块中。

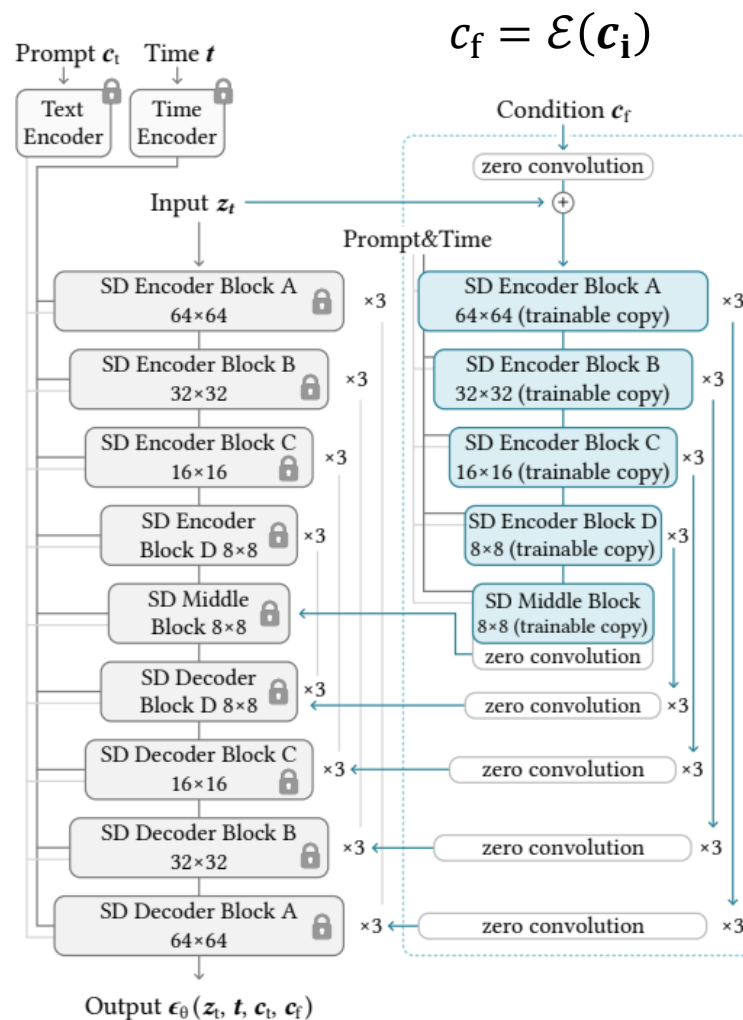


ControlNet 模型架构 - Condition

由于 Stable Diffusion使用类似于 VQ-GAN 的预处理方法，将整个 512x512 图像数据集转换为较小的 64x64 “潜在空间 (潜在图像)” 进行训练。这需要ControlNet将基于图像的条件转换为 64x64 的特征空间，以匹配卷积大小。添加了四层卷积层，这些卷积层的卷积核为4x4，步长为2，通道分别为16, 32, 64, 128，初始化为高斯权重，并与整个ControlNet模型进行联合训练。将图像空间条件 c_i 编码为特征映射：

$$c_f = \mathcal{E}(c_i)$$

c_f 为转换后的特征图。该网络将512x512的图像条件转换为64x64的特征图。如图所示，使用ControlNet来控制U-net的各个层。



零卷积

零卷积，是简单的 1x1 卷积层，其权重和偏差初始化为零。如此，在训练ControlNet之前，所有零卷积模块的输出都为零，使得ControlNet完全在原有Stable Diffusion基底模型的能力上进行微调，不会产生大的偏差。

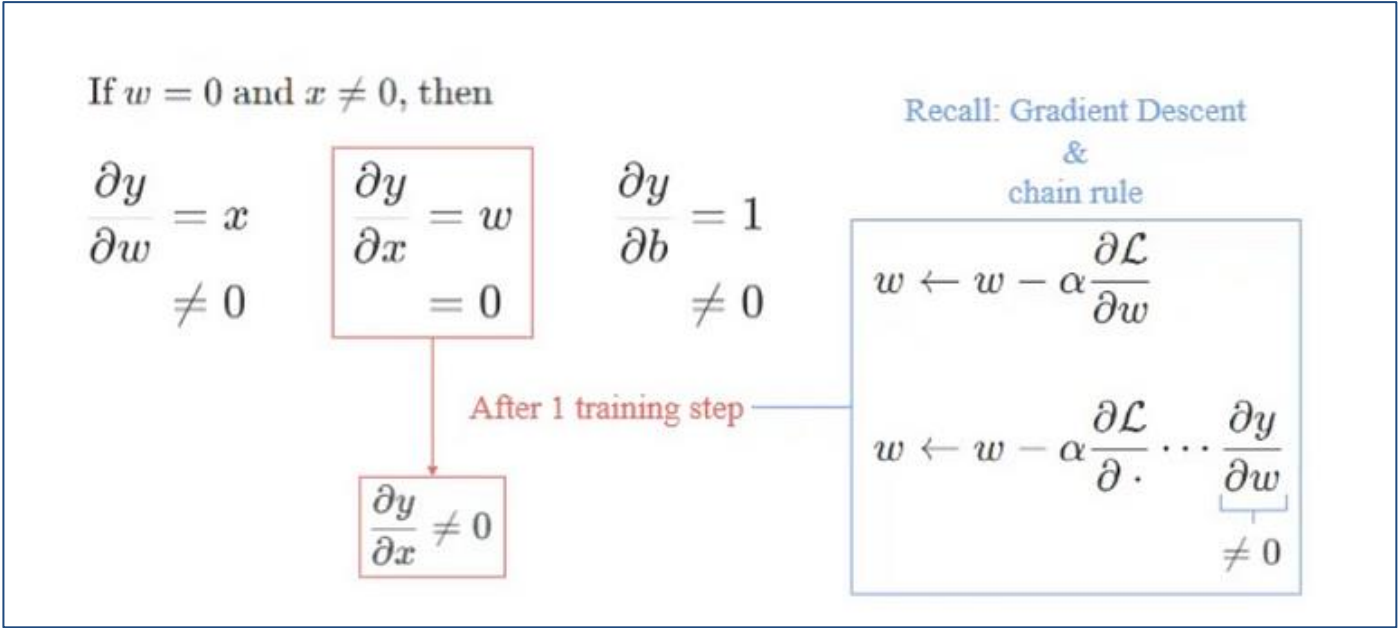
零卷积模块的初始权重为 0，但梯度不会为 0，为什么零卷积的梯度不会为 0？我们以一般层为例（注意卷积运算可以写成矩阵运算）

- 在反向传播时，我们看到发生右侧梯度流，并且我们可以看到梯度不一定为 0。

这样就能让零卷积模块逐渐成为具有非零权重的公共卷积层，并不断优化参数权重。

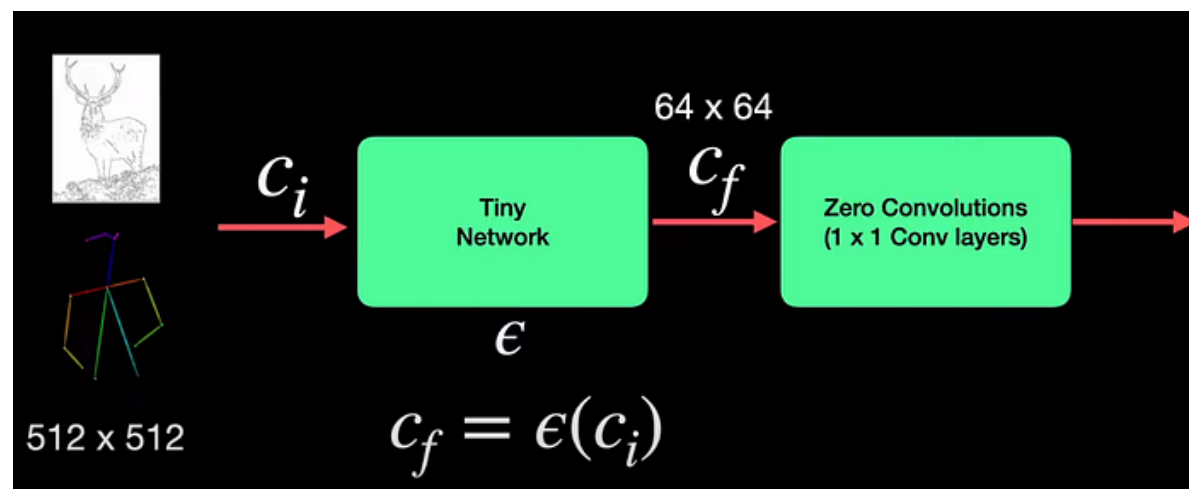
$$y = wx + b$$

$$\frac{\partial y}{\partial w} = x \quad \frac{\partial y}{\partial x} = w \quad \frac{\partial y}{\partial b} = 1$$

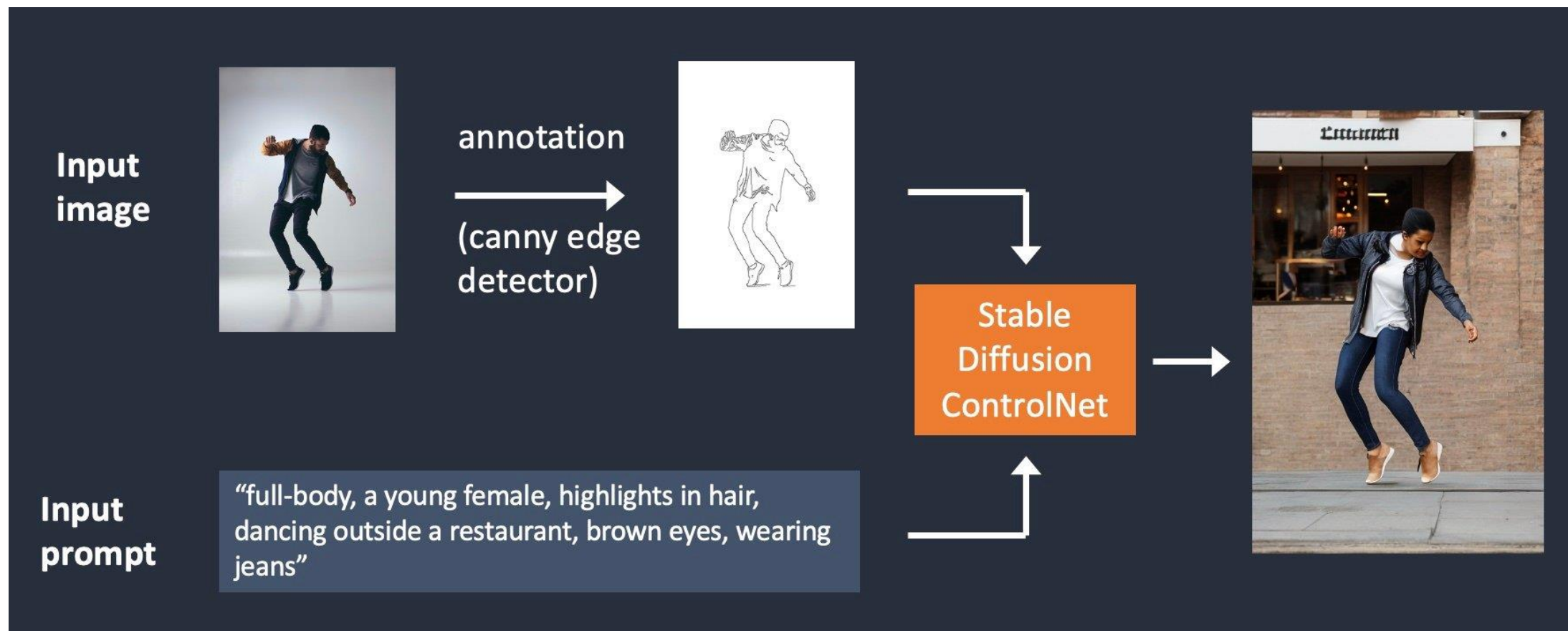


ControlNet 模型架构 – Condition - 边缘图或姿势

- 作为边缘图或姿势的条件输入具有 512×512 的尺寸，这对于这些 1×1 卷积层来说是一个非常高的维度。因此，为了克服这个问题，引入了一个具有 4 个卷积层的微型神经网络，该网络将这些图像从图像条件空间 c_i 转换为特征条件空间 c_f ，从而将尺寸减小到 64×64 。这意味着条件的维度足够小，可以愉快地使用零卷积。
- 每当有可选的文本提示 c_t 作为条件输入时，它们都会由 CLIP 编码器与时间位置编码一起处理，并直接输入到稳定扩散模型的冻结权重中。

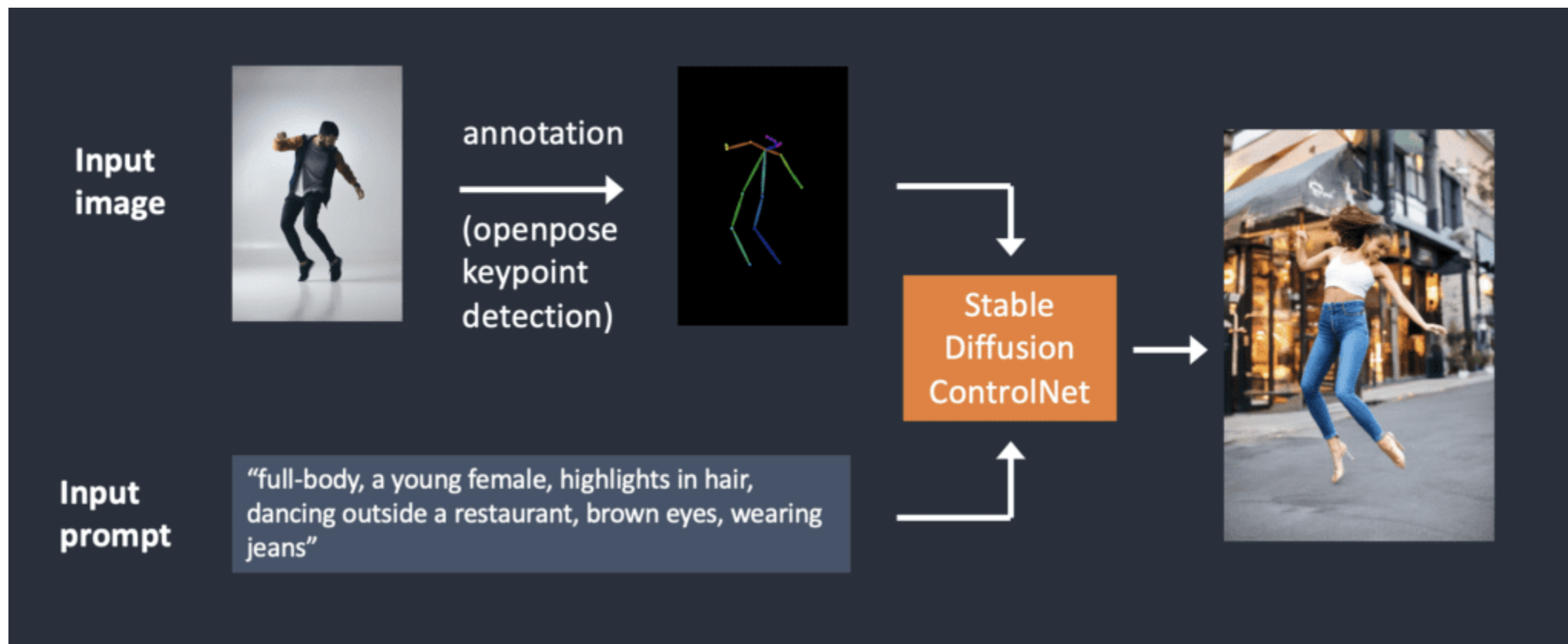


ControlNet 模型架构 – Condition - 边缘图



<https://www.ai-bites.net/controlnet-take-complete-control-of-images-from-the-generative-model/>

ControlNet 模型架构 – Condition - 姿势



<https://www.ai-bites.net/controlnet-take-complete-control-of-images-from-the-generative-model/>

ControlNet 应用: Canny 边缘检测图像生成

Canny 识别出图像内各对象的边缘轮廓, 根据提示词生成新的图片。
对应ControlNet模型: control_canny

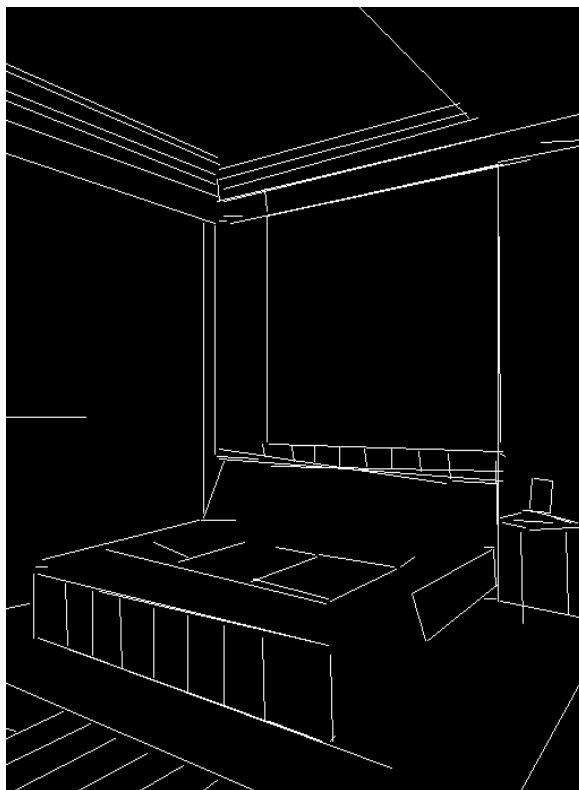


prompt = "masterpiece, best quality, ultra detailed, extremely detailed CG unity 8k wallpaper, best illumination, best shadow, an extremely delicate and beautiful, dynamic angle, finely detail, depth of field, bloom, shine, glinting stars, classic, illustration, painting, highres, original, perfect lighting";

ControlNet 应用: MLSD 线条检测图像生成

MLSD 线条检测用于生成房间、直线条的建筑场景效果比较好。
对应ControlNet模型: control_mlsd

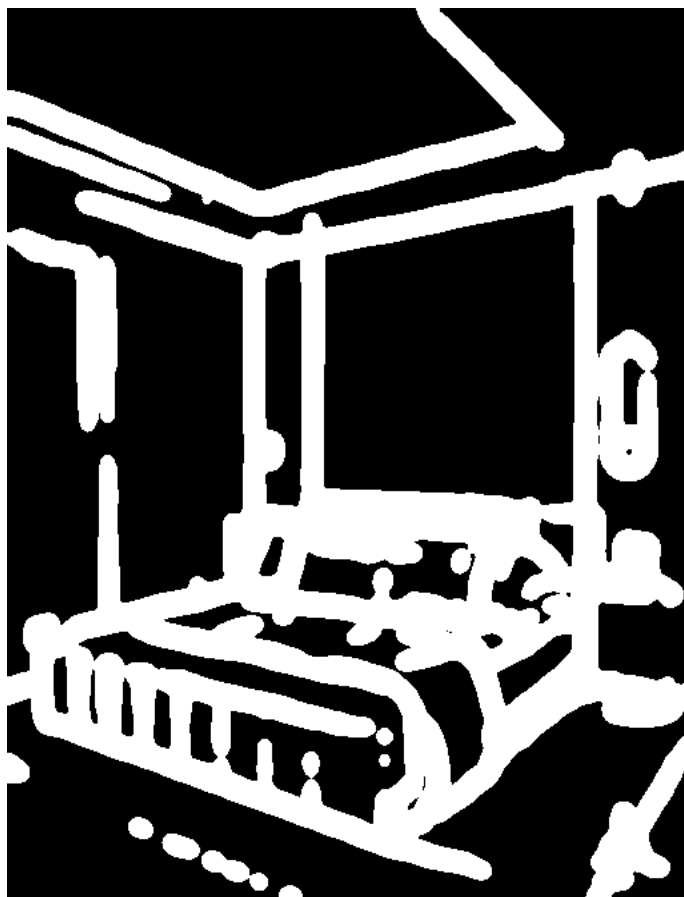
提示词: "royal chamber with fancy bed";



ControlNet 应用: Scribble 涂鸦图像生成

根据涂鸦效果的草图线条生成新的图片。
对应ControlNet模型: control_scribble。

提示词: "royal chamber with fancy bed";



ControlNet 应用: SoftEdge图像生成

SoftEdge 边缘检测可保留更多柔和的边缘细节, 类似手绘效果。

对应ControlNet模型: control_softedge。

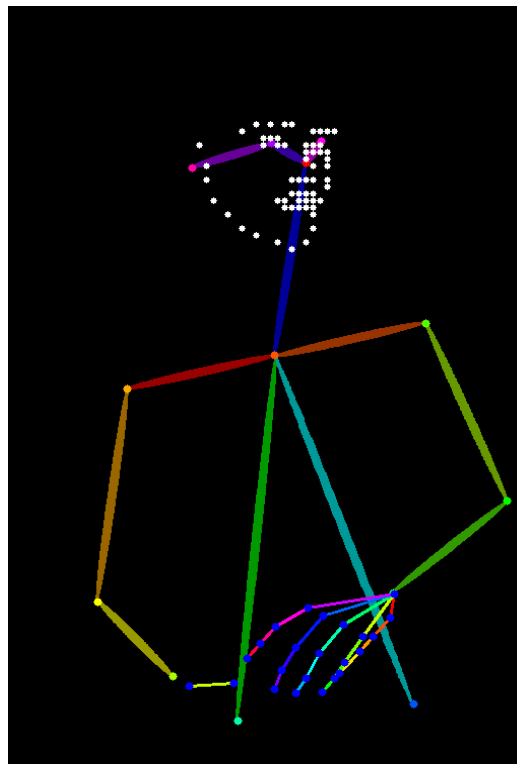
提示词: "royal chamber with fancy bed";



ControlNet 应用： OpenPose 姿态检测图像生成

OpenPose 姿态检测可生成图像中角色动作姿态的骨架图(含脸部特征以及手部骨架检测)，这个骨架图可用于控制生成角色的姿态动作。对应ControlNet模型：
control_openpose。

提示词： "chef in the kitchen";



ControlNet 应用：语义分割图像生成

语义分割可多通道应用，原理是用颜色把不同类型的对象分割开，让AI能正确识别对象类型和需求生成的区界。

提示词： "old house in stormy weather with rain and wind";

对应ControlNet模型： control_seg。



ControlNet 应用: Depth 深度检测图像生成

通过提取原始图片中的深度信息，生成具有原图同样深度结构的深度图，越白的越靠前，越黑的越靠后。

对应ControlNet模型: control_depth。

提示词: "Stormtrooper's lecture in beautiful lecture hall";



- Midas
- DPT

ControlNet 应用: Normal Map法线贴图图像生成

根据图片生成法线贴图, 便于AI给图片内容进行更好的光影处理, 它比深度模型对于细节的保留更加的精确。

对应ControlNet模型: control_normal。

提示词: "A head full of roses";

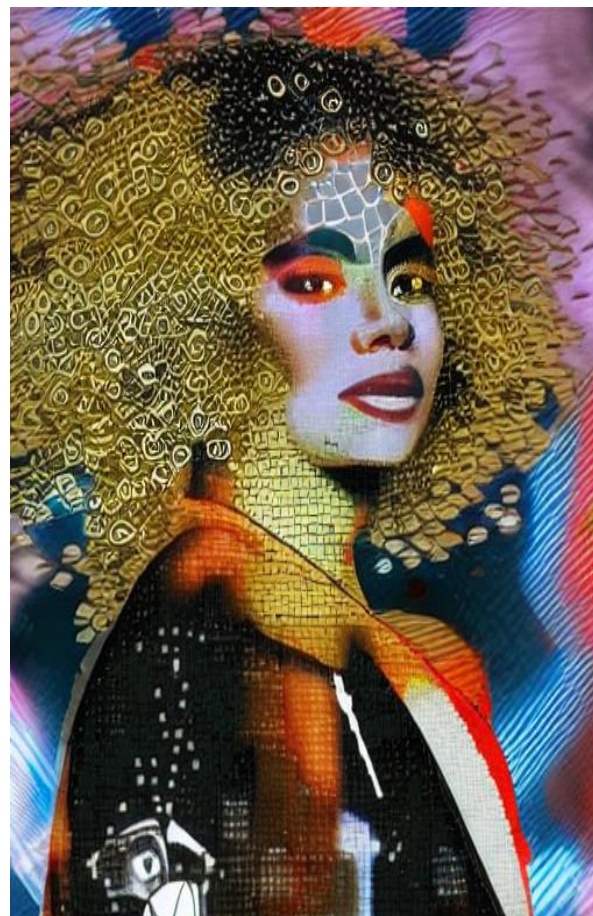


ControlNet 应用: Lineart图像生成

Lineart 边缘检测预处理器可很好识别出图像内各对象的边缘轮廓, 用于生成线稿。

对应ControlNet模型: control_lineart。

提示词: "michael jackson concert";



ControlNet 应用: Content Shuffle 图像生成

Content Shuffle 图片内容变换位置, 打乱次序, 配合模型 control_v11e_sd15_shuffle 使用。

对应ControlNet模型: control_shuffle。

提示词: "New York"





Thank

You