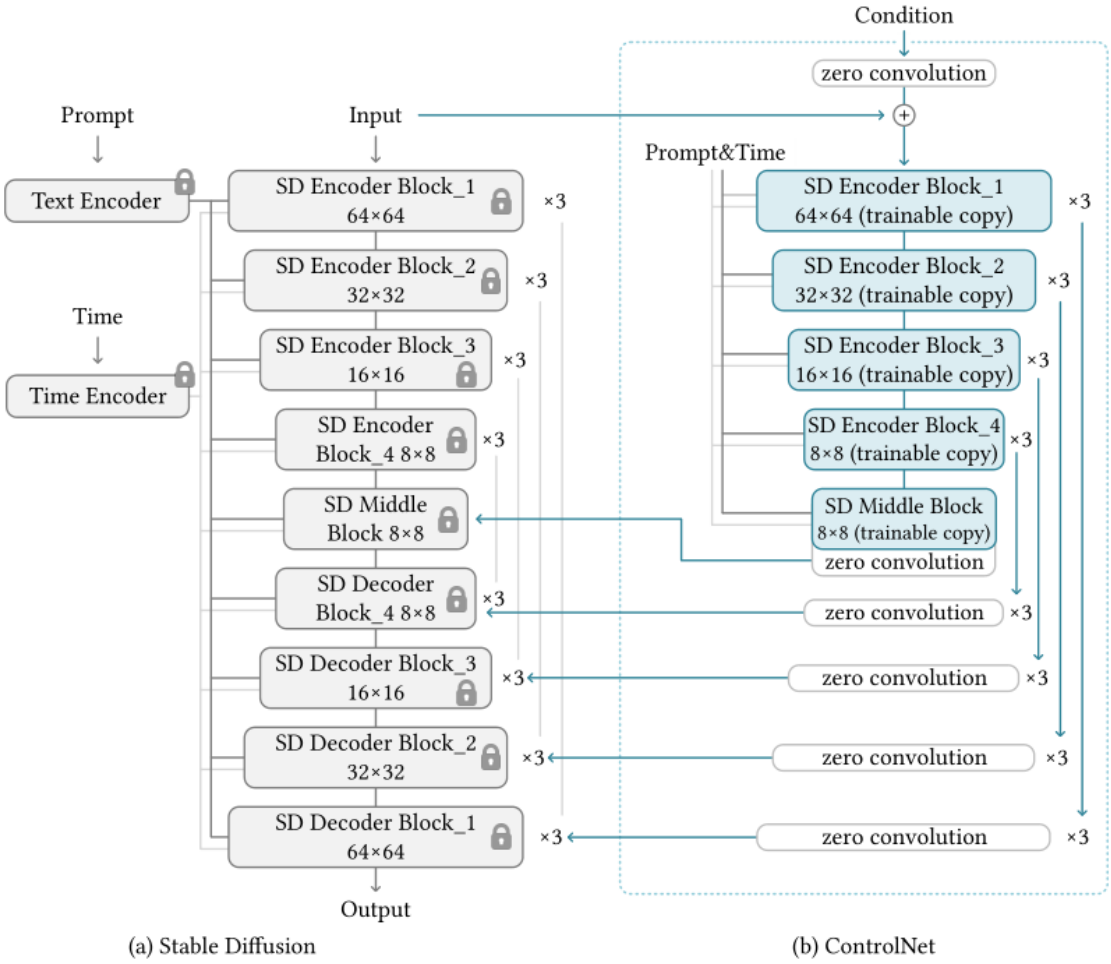
SD3 Controlnet 调研

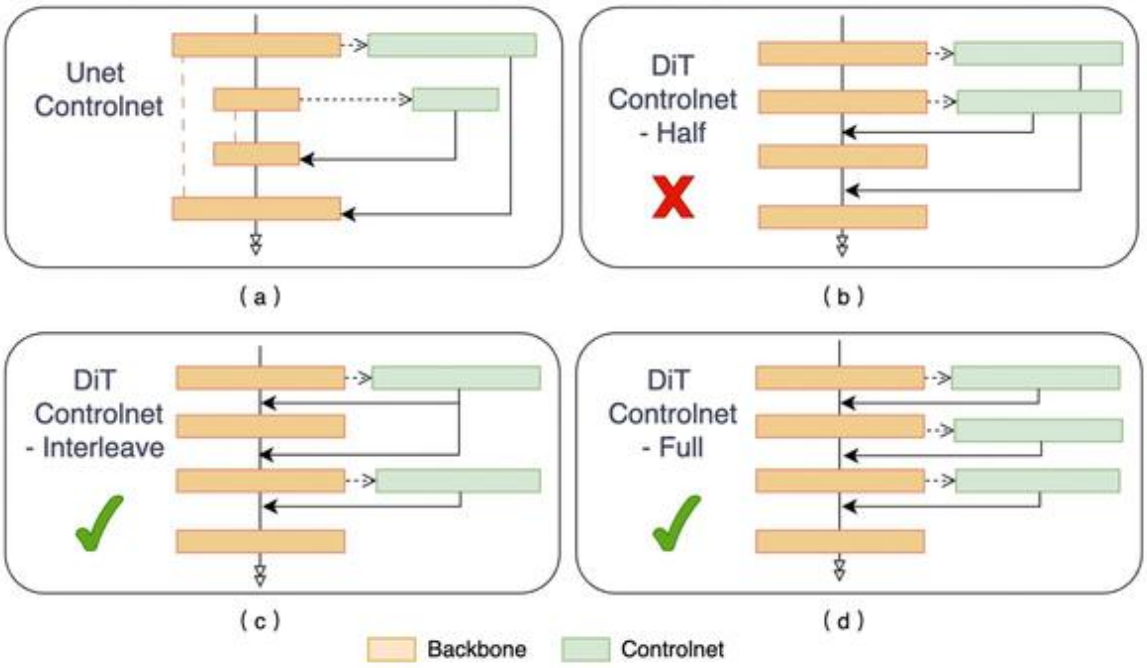
Controlnet 原始论文：https://arxiv.org/abs/2302.05543

Hugging Face controlnet_sd3： https://huggingface.co/docs/diffusers/api/pipelines/controlnet_sd3

Unet Controlnet 结构



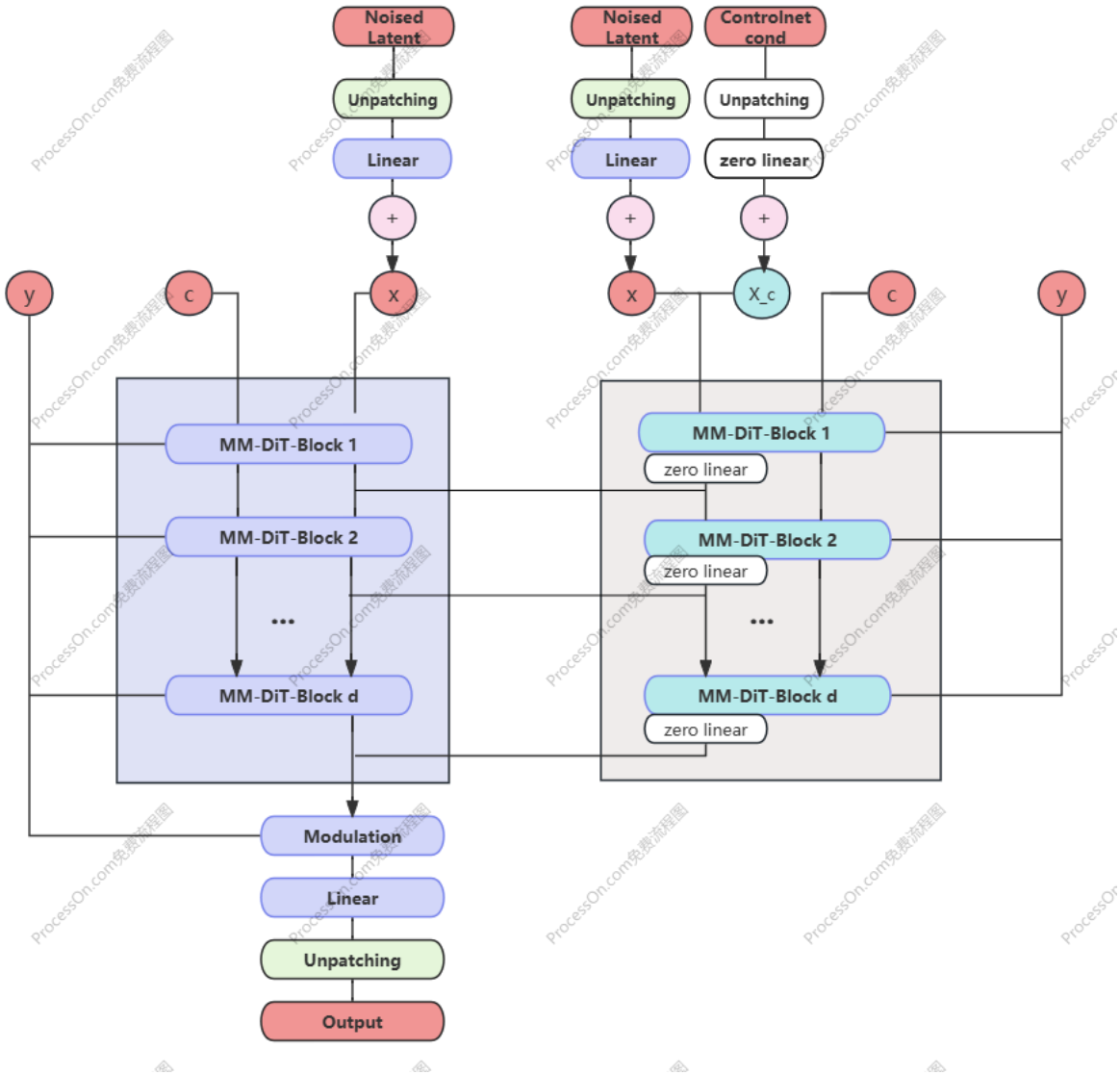
DiT Controlnet 结构



社区普遍使用的 Unet ControlNet (图a)， 直接迁移到 DiT 架构上需要将前一半的特征加到后一半上面（图b）。不同于传统的基于 Unet 的文生图结构，DiT 的文生图模型由一系列的 Transformer Block 堆叠而成，没有显式的 Encoder-Decoder 结构，因此直接迁移Unet的Controlnet 结构效果不好。因此目前主流采用 图c 或 图 d 的方式。

SD3 Controlnet 结构

https://huggingface.co/InstantX/SD3-Controlnet-Canny



上图完整版：https://www.processon.com/diagraming/6718ae1961fdee7d75eab48d

代码实现

需迁移的diffusers代码包括以下部分：

- src\diffusers\models\transformers\transformer_sd3.py： residual部分

```
class SD3Transformer2DModel(ModelMixin, ConfigMixin, PeftAdapterMixin, FromOriginalModelMixin):
    def forward(
                ckpt_kwargs: Dict[str, Any] = {"use_reentrant": False} if is_torch_version(">=", "1.11.0") else {}
                encoder_hidden_states, hidden_states = torch.utils.checkpoint.checkpoint(
                    create_custom_forward(block),
                    hidden_states,
                    encoder_hidden_states,
                    temb,
                    **ckpt_kwargs,
                )

            else:
                encoder_hidden_states, hidden_states = block(
                    hidden_states=hidden_states, encoder_hidden_states=encoder_hidden_states, temb=temb
                )

            # controlnet residual
            if block_controlnet_hidden_states is not None and block.context_pre_only is False:
                interval_control = len(self.transformer_blocks) // len(block_controlnet_hidden_states)
                hidden_states = hidden_states + block_controlnet_hidden_states[index_block // interval_control]
```

- src\diffusers\models\controlnet_sd3.py： 类似于transformer_sd3.py

```
class SD3ControlNetModel(ModelMixin, ConfigMixin, PeftAdapterMixin, FromOriginalModelMixin):
    def forward(
                    **ckpt_kwargs,
                )

            else:
                encoder_hidden_states, hidden_states = block(
                    hidden_states=hidden_states, encoder_hidden_states=encoder_hidden_states, temb=temb
                )

            block_res_samples = block_res_samples + (hidden_states,)

        controlnet_block_res_samples = ()
        for block_res_sample, controlnet_block in zip(block_res_samples, self.controlnet_blocks):
            block_res_sample = controlnet_block(block_res_sample)
            controlnet_block_res_samples = controlnet_block_res_samples + (block_res_sample,)

        # 6. scaling
        controlnet_block_res_samples = [sample * conditioning_scale for sample in
        controlnet_block_res_samples]
```

- src\diffusers\pipelines\controlnet_sd3： 整体pipeline

  pipeline_stable_diffusion_3_controlnet.py

  pipeline_stable_diffusion_3_controlnet_inpainting.py

问题

ppdiffusers：

- self.inference_optimize 分支：暂不考虑

```
        if self.inference_optimize:
            out = self.simplified_sd3(
                hidden_states=hidden_states, encoder_hidden_states=encoder_hidden_states, temb=temb
            )
            hidden_states = out[1]
            encoder_hidden_states = None
        else:
            encoder_hidden_states, hidden_states = self.sd3_origin_transformer(
                hidden_states=hidden_states, encoder_hidden_states=encoder_hidden_states, temb=temb
            )
```

其他：

- PaddleMIX 版本：develop 后端： GPU
- 目录结构与 diffusers保持一致 ✔
- 只推理 or 训练+推理： 只推理
- tests、examples： examples 暂不考虑