

# 2024/10/01

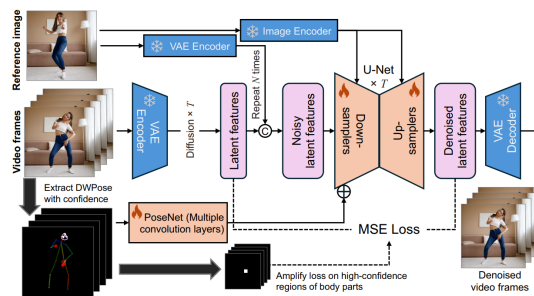
- Jiu

## MimicMotion: High-Quality Human Motion Video Generation with Confidence-aware Pose Guidance

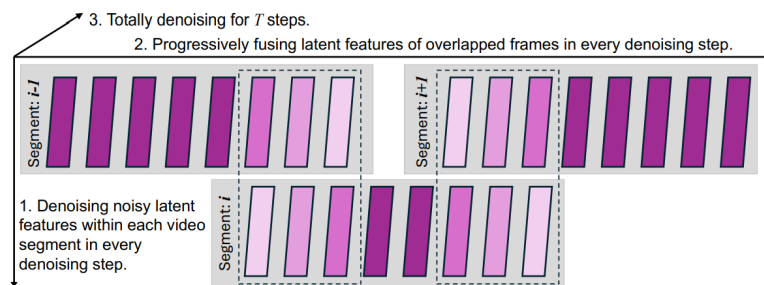
<https://arxiv.org/pdf/2406.19680>

(NeurIPS 2024?)

- Diffusionによる動画生成手法の論文
- pose情報をガイドとした動画生成
- confidenceを導入することで信頼度の高い部分の損失を重要視→pose推定モデルに依存しない



- 長時間の動画を生成するため動画をセグメントに分割
- 隣接セグメントはoverlapし、重み $\lambda_{fusion}$ を導入することで融合時の一貫性を保持する



### Algorithm 1: Progressive frame-level latent fusion for long video generation.

**Input:**  $I_{ref}$ : Reference image.  $P_i^j$ : Pose frame corresponding to  $j$ -th frame in  $i$ -th video segment;  $z_i^j$ : The latent feature of  $j$ -th frame in  $i$ -th video segment;  $N$ : the number of frames in a video segment;  $C$ : the number of overlapped frames.

**Output:**  $z'$ : A long sequence of latent features of video frames.

```

 $z \sim \mathcal{N}(0, I)$ ; // Random initialization of video latent features.
 $\lambda_{fusion} \leftarrow 1/(C+1)$ ; // Set a scale of latent fusion.
for  $t = T$  to 1 do // Denoising from noisy latent features step by step.
  for  $i = 1, 2, \dots$  do
     $z_i \leftarrow \text{DM}(z_i, I_{ref}, P_i, t)$  // Separately denoise each segment.
  for  $i = 1, 2, \dots$  do // Within each video segment.
    for  $j = 1$  to  $N$  do // Start latent fusion for each frame.
      if  $i > 1$  and  $j \leq C$  then // Latent fusion with the previous segment.
         $z_i^j \leftarrow j\lambda_{fusion}z_i^j + (1-j\lambda_{fusion})z_{i-1}^{N-C+j}$ 
      else if  $j > N-C$  then // Latent fusion with the next segment.
         $z_i^j \leftarrow (N+1-j)\lambda_{fusion}z_i^j + (C-N+j)\lambda_{fusion}z_{i+1}^{C-N+j}$ 
    return  $z' = \text{Merge}(z)$ ; // Merge multi-segment features following Listing 1.

```

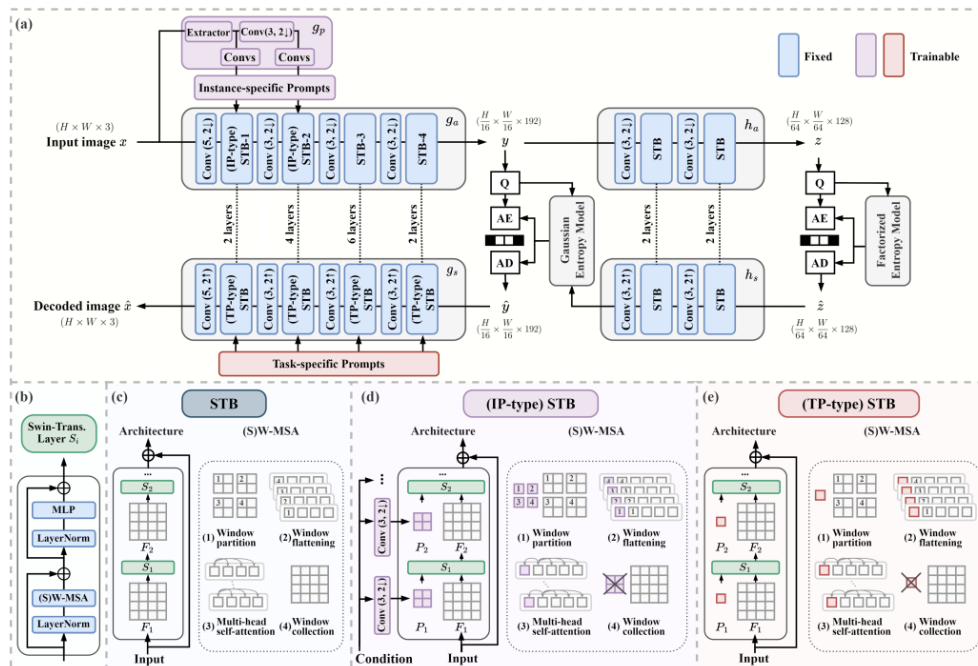
- Tatsumi

## TransTIC: Transferring Transformer-based Image Compression from Human Perception to Machine Perception Image

<https://arxiv.org/pdf/2306.05085>

(ICCV 2023)

- for human→for machineのScalable Image Coding Model(SICM)の論文
- プロンプト（学習パラメタ）を使用したtransformerベースのSICMを提案
  - ベースラインはfor human modelをパラメタ固定で使用
  - EncoderとDecoderのプロンプトのみを学習することでマルチタスクに対応
  - Encoder側：タスクと画像に依存
  - Decoder側：タスクのみに依存



- Tanaka

## Hybrid Implicit Neural Image Compression with Subpixel Context Model and Iterative Pruner

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10402791>

(VCIP 2023)

- コンテキストモデルを使用したINRによる画像圧縮手法の論文
- 従来のピクセルベースのコンテキストモデルは計算量が多い
- サブピクセルベースのコンテキストモデル(SCM)を提案
  - 行列演算に基づいて並列に処理
  - マルチスケールな洗剤変数のため、PixelShuffleによりサイズを合わせる
- 学習パラメタの枝刈、量子化によるビットレート削減性能を上げるために  $\mathcal{L}_1((\theta, \phi))$  を損失に入れ、パラメタを0に近づける

