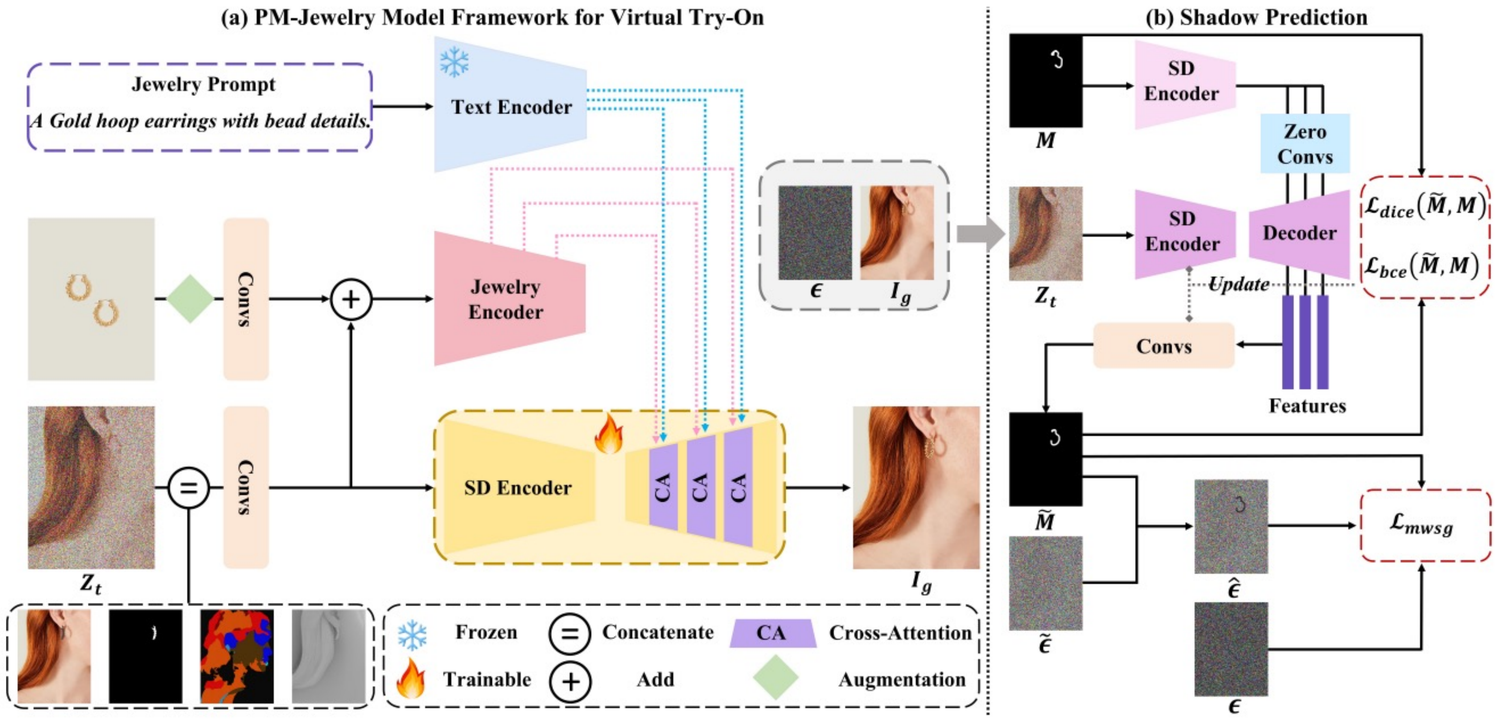
PM-Jewelry: Personalized Multimodal Adaptation for Virtual Jewelry Try-On with Latent Diffusion

Yangfan He, Yinghui Xia, Jinfeng Wei, Tianyu Shi, Jingsong Yang*
ty.shi@mail.utoronto.ca edward.yang@autoagents.ai

Introduction

- **Motivation and Goal**
  - Rapid growth of e-commerce and its transformative impact on virtual try-on (VTO) technologies.
  - Unique challenges in virtual jewelry try-on (e.g., intricate details, reflective surfaces, precise alignment).
  - A novel framework leveraging multimodal learning and latent diffusion models to achieve highly personalized and realistic jewelry try-on experiences.

- **Proposed Framework for PM-Jewelry**
  - Integration of latent diffusion, multimodal learning, and personalized adaptation.
  - Dual-pathway network for combining image-based and latent embeddings.



Categories

Input sources

- **Agnostic maps**
- **Dense pose**
- **Jewelry features**
- **Depth maps**

Key Contributions

- **Personalized Multimodal Adaptation**
  - Integration of user-specific attributes such as skin tone, facial structure, and personal style.
  - Enhancing realism through dual-pathway processing and latent space optimization.

- **User-Centric Features**
  - Style customization (e.g., modern, traditional, minimalistic preferences).
  - Interactive feedback loop for dynamic adjustments.

- **Attention-Based Latent Diffusion**
  - Leveraging depth maps for accurate jewelry alignment with user anatomy.
  - Advanced rendering techniques for reflective surfaces and fine details.

Datasets

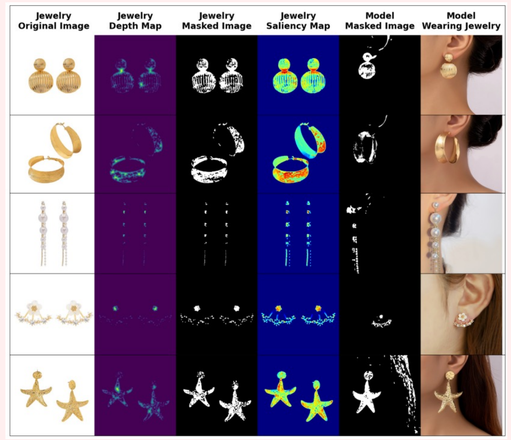
- **A novel jewelry try-on dataset**
  - 6,157 paired images encompassing earrings, necklaces, and rings with diverse styles and fitting conditions.
- **Performance Metrics**
  - Structural Similarity Index (SSIM), Learned Perceptual Image Patch Similarity (LPIPS), and Fréchet Inception Distance (FID).
- **Qualitative Analysis**
  - Comparisons against several baselines as well as ablation studies.

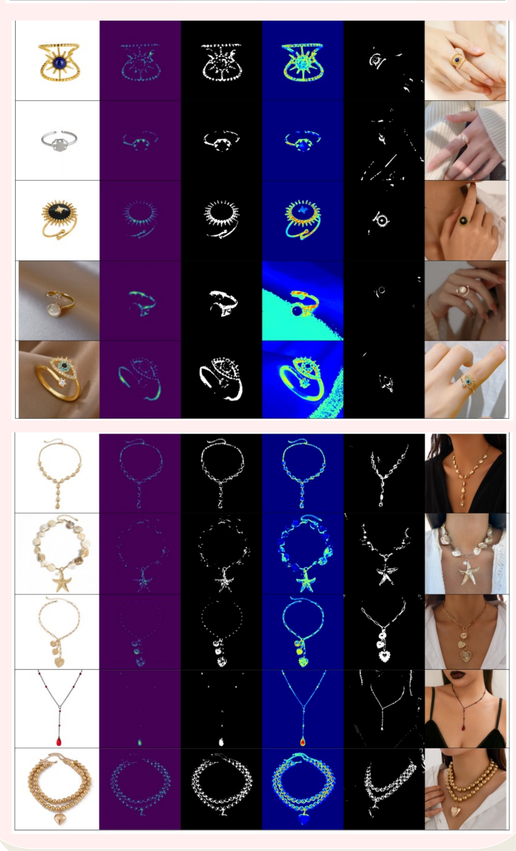
Evaluation

- **Quantitative Evaluation**
  - PM-Jewelry outperforms baselines in SSIM, LPIPS, and FID metrics, indicating superior structural preservation and realism.

Model	Earrings			Necklaces			Rings		
	SSIM (↑)	LPIPS (↓)	FID (↓)	SSIM (↑)	LPIPS (↓)	FID (↓)	SSIM (↑)	LPIPS (↓)	FID (↓)
Stable Diffusion v1.5 [21]	0.7135	0.1278	24.732	0.7018	0.1351	25.823	0.7092	0.1234	25.578
ControlNet [22]	0.7123	0.1292	25.176	0.7068	0.1313	26.512	0.7074	0.1247	25.981
PICTURE [23]	0.7809	0.1223	19.057	0.7607	0.1268	19.984	0.7729	0.1195	19.674
Gal4way/TPD [24]	0.6752	0.1307	28.245	0.6546	0.1419	29.887	0.6629	0.1368	28.582
StableVITON [15]	0.6704	0.1015	30.298	0.6699	0.1116	30.834	0.6718	0.1073	29.621
ComfyUI [25]	0.7942	0.1214	29.769	0.7783	0.1227	30.165	0.7847	0.1192	30.456
Ladi-VTON [12]	0.7134	0.1273	24.612	0.7064	0.1298	26.055	0.7087	0.1246	25.735
GP-VTON [8]	0.7119	0.1264	25.411	0.7046	0.1322	26.213	0.7072	0.1238	25.619
DCI-VTON [10]	0.7295	0.1259	24.831	0.7081	0.1345	26.356	0.7084	0.1211	25.945
Ours	**0.8127**	**0.0972**	**13.529**	**0.9065**	**0.1003**	**9.805**	**0.8913**	**0.1094**	**10.854**

- **Qualitative Evaluation**
  - Visual comparisons demonstrate the model's effectiveness in preserving jewelry details and alignment with facial features.





Conclusion

- **Observations**
  - Integration of latent diffusion and multimodal learning for scalable, robust virtual jewelry try-on.
  - PM-Jewelry represents a significant advancement in virtual jewelry try-on, offering realistic and personalized simulations.
  - Significant improvements in rendering quality, alignment accuracy, and user satisfaction.
- **Future work**
  - Enhancing real-time interaction capabilities.
  - Broadening applicability to diverse jewelry designs and user demographics.
  - Exploring augmented reality integration for immersive experiences.

- **Ablation Study**
  - Highlighting the importance of data augmentation, zero cross-attention mechanism, jewelry encoder, and text encoder in integrating features and maintaining style consistency.