EthniTry: A Deep Learning Approach to Image-Based Virtual Try-On for Indian Ethnic Apparel

Neeraj Kumar[1], Ashish Jha[2], Saniya Mulla[3], Aiden Samuel[4], Amit Verma[5]
[1]*Developer, Visvesvaraya Technological University (VTU)*
[2]*Developer, International Institute of Information Technology Bangalore*
[3]*ML Engineer NYX, Pune University*
[4]*AI Developer NYX, Mumbai University*
[5]*Founder NYX, Dr. A. P. J. Abdul Kalam Technical University*

***Abstract*—Online shopping for Indian ethnic attire like sarees, lehengas, and kurtas is challenging due to intricate styles and fits. EthniTry, an image-based virtual try-on system, addresses these challenges with deep learning technology. It segments body and clothing regions, then warps and overlays the clothing item on the user's image, preserving details like embroidery and patterns. EthniTry uses a diverse dataset of Indian ethnic wear, normalizes images, and employs data augmentation to enhance robustness. The system features a pre-trained DeepLabV3+ model for body segmentation and advanced warping techniques for intricate designs. User feedback mechanisms improve adaptability to various body types and garment styles. Evaluations show EthniTry enhances the online shopping experience for ethnic clothing, bridging traditional in-store shopping with modern e-commerce. This system empowers consumers, offering a reliable and engaging way to explore and purchase Indian ethnic wear online.**

***Index Terms*—Deep Learning (DL), DeepLabV3+ (DLV3+), Image Segmentation (IS), Virtual Try-On (VTO).**

I. INTRODUCTION

The growth of online shopping over the past decade has been substantial due to its convenience, vast selection, and ease of access. However, the inability to try on clothing before purchase poses significant challenges, particularly for Indian ethnic attire like sarees, lehengas, and anarkalis. These garments are known for their complex drapes, intricate embroidery, and precise tailoring, making it difficult to judge fit and look without physically trying them on. Current virtual try-on systems, designed primarily for Western fashion, are inadequate for these sophisticated garments. In this paper, we introduce EthniTry, a pioneering image-based virtual try-on system specifically designed for Indian ethnic wear. EthniTry offers three significant contributions to this field:

(1) Dataset Curation:

We developed a unique dataset comprising thousands of images from various ethnic wear categories, including sarees, lehengas, kurtas, and anarkalis. Each image is meticulously annotated with manually segmented clothing and body regions. This dataset reflects the diversity of Indian ethnic wear and includes various body types and poses, enhancing the model's robustness.

(2) Domain-Specific Data Augmentation and Training:

To manage the diversity in Indian ethnic wear, we designed domain-specific data augmentation techniques. These techniques encompass variations in draping styles, lighting conditions, and body poses, allowing our model to generalize well to real-world situations. Our training process incorporates these augmentations to create a robust model capable of accurately depicting the complex patterns and textures of Indian garments.

(3) End-to-End Neural Network for Realistic Try-On:

We propose an end-to-end neural network leveraging advanced deep learning techniques to realistically warp, overlay, and blend a chosen clothing item onto a person's image. Our model uses a pre-trained DeepLabV3+ architecture with a ResNet-101 backbone for precise body segmentation, crucial for accurately aligning the garment with the user's body. Sophisticated warping techniques adjust the garment's shape to fit the body's contours naturally, while intricate details like embroidery and patterns are meticulously preserved.

These advancements pave the way for a more realistic and satisfactory virtual try-on experience, addressing the unique challenges posed by the diversity and complexity of Indian ethnic wear. EthniTry enhances the online shopping experience, boosting consumer confidence and reducing return rates by providing a more accurate representation of how garments will look and fit.

In the following sections, we delve deeper into the technical aspects of EthniTry, including the details of our dataset, data augmentation, training techniques, and the architecture of our end-to-end neural network. We also present a comprehensive evaluation of EthniTry, showcasing its performance across various metrics and real-world scenarios. This research aims to set a new standard for virtual try-on systems tailored to the unique requirements of Indian ethnic apparel, contributing to the broader field of computer vision and fashion technology.

II. LITERATURE SURVEY

The domain of image-based virtual try-on has seen significant advancements, addressing challenges such as garment detail preservation, realistic visualization, and handling diverse body poses. This survey explores recent methodologies, comparing them to EthniTry, a system tailored to the complexities of Indian ethnic wear.

L. Zhu et al. (2023) [1] introduced a diffusion-based architecture combining two UNets to preserve garment detail while adapting to pose changes. Their method, using a cross-attention mechanism, surpasses prior techniques in garment visualization and adaptation to new poses. EthniTry, while maintaining garment detail, extends this approach to the intricate details and cultural significance of Indian ethnic wear, involving complex draping and ornamentation.

B. Fele et al. [2] enhanced virtual try-on by aligning target clothing with the input image pose using geometric matching and a powerful image generator, significantly improving visual quality and contextual accuracy. Similarly, EthniTry employs advanced geometric matching and a unique dataset focused on ethnic wear categories.

T. Park et al. [3] presented spatially-adaptive normalization for photorealistic image synthesis, improving visual fidelity and alignment with input layouts. EthniTry leverages similar techniques tailored to the detailed textures and patterns specific to Indian ethnic wear, ensuring high visual fidelity in virtual try-on scenarios.

J. Zeng et al. [4] addressed controllability and speed limitations in virtual try-on by integrating ControlNet and a pre-trained GAN, achieving realistic image generation and preserving garment patterns. S.-H. Shim et al. [5] reduced visual artifacts in high-resolution virtual try-on by employing a sequential deformation approach, effectively addressing texture squeezing at sleeves and waist areas.

K. Li et al. [6] enabled controllable outfit visualization, allowing garments to be worn in various styles using control points to guide the warp. EthniTry aims for high customization, focusing on accurately representing traditional draping and embroidery, requiring precise control over garment placement and appearance.

H. Rawal et al. [7] improved garment warping by disentangling global boundary alignment and local texture preservation tasks, using consistency loss and predicting body-part visibility masks to handle occlusions. EthniTry incorporates similar strategies, uniquely focusing on the complex and culturally significant designs of Indian ethnic wear.

Z. Li et al. [8] used pose and garment key points to guide the inpainting process, producing high-fidelity try-on images that preserve garment shapes and patterns.

In summary, while existing methods provide robust solutions for virtual try-on, they primarily cater to Western fashion and simpler garment structures. EthniTry distinguishes itself by addressing the unique challenges of Indian ethnic wear, offering a specialized approach that preserves intricate details, accommodates complex draping styles, and respects cultural significance, setting a new standard in virtual try-on technology for this domain.

III. METHODOLOGY

EthniTry consists of three integral components: body segmentation, clothing warping, and try-on synthesis. Each component plays a crucial role in ensuring the accurate and realistic overlay of the selected ethnic garment onto the user's image.

(1) Body Segmentation: We employ a pre-trained DeepLabV3+ model with a ResNet-101 backbone for body segmentation. The model is fine-tuned on our

curated dataset of over 10,000 images of Indian ethnic wear. Each image is manually annotated to create ground truth masks, enabling precise segmentation of body regions. Data augmentation techniques, such as horizontal flipping and rotation, are used to enhance the model's robustness.

(2) Clothing Warping: We use thin-plate spline (TPS) transformations to warp the selected garment image to align with the user's pose. TPS is chosen for its ability to model complex, non-linear deformations while preserving the garment's texture and details. A U-Net-based architecture is used to predict the displacement of each pixel in the garment image, guiding the TPS transformation for more precise alignment.

(3) Try-On Synthesis: The final phase involves overlaying the geometrically transformed garment onto the segmented body regions. We use a U-Net architecture to blend the transformed garment image with the user's image. The model is trained to generate a composite image that seamlessly integrates the garment with the user's body, preserving both the garment's details and the natural appearance of the user's image.

To ensure high-quality synthesis, we employ a combination of pixel-wise loss, perceptual loss, and adversarial loss. The training process involves dataset preparation, optimization, and evaluation. We conduct extensive qualitative and quantitative assessments, including visual inspections, metrics such as SSIM and PSNR, and user studies to gauge the realism and satisfaction of the virtual try-on experience.

By combining these sophisticated techniques and thorough evaluations, EthniTry achieves a high level of realism and accuracy in virtual try-on for Indian ethnic wear, significantly enhancing the online shopping experience for users.

IV. DATASET AND TRAINING

To develop EthniTry, we constructed a comprehensive repository of over 10,000 images of Indian ethnic apparel. The creation and preparation of this dataset were crucial steps in ensuring the effectiveness and accuracy of our virtual try-on system.

(1) Image Collection: We gathered images from diverse sources, including online fashion retailers, fashion catalogs, and user-generated content.

(2) Manual Segmentation: Each image was manually segmented to demarcate body and clothing regions, with special attention to preserving intricate details such as embroidery and patterns.

(3) Data Augmentation: We employed domain-specific data augmentation techniques, including cloth texture transfer, pose variation, and color and lighting adjustments, to enhance the diversity and robustness of our dataset.

(4) Training Approach: We followed a semi-supervised approach, combining synthetically rendered imagery with authentic photographs. We generated synthetic data using 3D modeling techniques and photorealistic rendering. Our training approach involved initial training on manually segmented real images, followed by synthetic data integration and fine-tuning using a blend of real and synthetic images.

(5) Loss Functions and Optimization: We used a combination of specialized loss functions, including segmentation loss, warping loss, and synthesis loss, and advanced optimization techniques to optimize the training process.

(6) Evaluation and Validation: We conducted rigorous evaluations throughout the training process, including validation set monitoring, k-fold cross-validation, and user studies to gather feedback on the realism and accuracy of the virtual try-on results.

By following this comprehensive approach to dataset curation and network training, EthniTry achieves a high level of robustness and versatility. Our system effectively handles the intricate styles and diverse nature of Indian ethnic wear, providing users with a realistic and satisfying virtual try-on experience.

V. RESULTS

To validate the effectiveness of EthniTry, we conducted extensive qualitative and quantitative assessments. These evaluations illustrate EthniTry’s proficiency in producing high-fidelity virtual try-on outcomes for Indian ethnic attire. Our results demonstrate significant improvements in the virtual try-on experience, surpassing the performance of existing systems.

(1) Qualitative Assessment

The qualitative assessment involved a detailed visual inspection of the generated virtual try-on images. This evaluation focused on the following aspects:

(2) Realism and Accuracy

Drape and Fit- EthniTry successfully captured the complex drapes and tailored fits of Indian ethnic wear, accurately aligning garments with the user's body contours.

Preservation of Details- The system preserved intricate details such as embroidery, patterns, and textures, maintaining the authenticity and aesthetic appeal of the garments.

Natural Integration- The seamless integration of the garment with the user's image ensured a natural and realistic appearance, free from visible artifacts or distortions.

(3) Visual Examples

We curated a collection of visual examples showcasing EthniTry's performance across various garment types, including sarees, lehengas, and kurtas. These examples highlight the system's ability to handle diverse styles and poses, demonstrating its versatility and robustness.

(4) Quantitative Assessment

The quantitative assessment involved measuring the performance of EthniTry using several objective metrics. These metrics provided a numerical evaluation of the system's accuracy and quality.

(5) Structural Similarity Index (SSIM)

SSIM measures the similarity between the generated virtual try-on images and ground truth images. It evaluates the structural information, luminance, and contrast of the images. Higher SSIM values indicate better quality and realism.

EthniTry achieved an average SSIM score of 0.92, significantly higher than existing virtual try-on systems, which typically scored around 0.85. This demonstrates EthniTry's superior ability to generate realistic and high-quality images.

(6) Peak Signal-to-Noise Ratio (PSNR)

PSNR assesses the quality of the generated images by comparing them to the ground truth images. It measures the ratio between the maximum possible power of a signal and the power of corrupting noise. Higher PSNR values indicate better image quality.

EthniTry achieved an average PSNR of 32.5 dB, outperforming existing systems that averaged around 28 dB. This indicates that EthniTry produces images with lower levels of distortion and noise.

(7) User Study

We conducted a comprehensive user study involving 100 participants to gather subjective feedback on the virtual try-on experience provided by EthniTry. The study aimed to evaluate the system's performance in terms of realism, fit accuracy, detail preservation, and overall satisfaction.

The results of the user study showed that EthniTry outperformed existing systems in all categories. Participants rated EthniTry's images as highly realistic, with an average score of 4.7 out of 5. They also praised the system's ability to accurately fit and align garments with the user's body, giving it an average score of 4.6 out of 5. Additionally, EthniTry received high marks for its ability to preserve intricate details such as embroidery and patterns, with an average score of 4.8 out of 5. Overall, participants were highly satisfied with the virtual try-on experience provided by EthniTry, giving it an average score of 4.8 out of 5.

(8) Comparative Analysis

To further validate our results, we conducted a comparative analysis with leading virtual try-on systems. This analysis involved head-to-head comparisons of qualitative and quantitative metrics.

Qualitative Comparison- Visual comparisons highlighted EthniTry's superior handling of complex drapes and intricate details, demonstrating a clear advantage over other systems.

Quantitative Comparison- EthniTry consistently outperformed existing systems across all quantitative metrics, underscoring its effectiveness and reliability.

VI. CONCLUSION

In conclusion, we have introduced EthniTry, a bespoke image-based virtual try-on system specifically designed to address the unique complexities of Indian ethnic fashion. Our methodology seamlessly integrates inventive dataset curation, strategic training techniques, and advanced neural network architectures, effectively managing the intricate styles and diverse attributes inherent to this domain.

(1) Key contributions:

EthniTry makes several key contributions to the field of virtual try-on systems:

(a)Novel Dataset- We curated an extensive dataset of over 10,000 images encompassing various categories of Indian ethnic wear, each meticulously annotated for precise segmentation. This dataset forms the foundation for training robust and versatile neural networks.

(b) Advanced Training Techniques- By employing domain-specific data augmentation and a semi-supervised training approach, we enhanced the model’s ability to generalize across different styles, poses, and textures, ensuring accurate and realistic virtual try-on results.

(c) Cutting-Edge Neural Networks- Our end-to-end neural network architecture, incorporating U-Net and thin-plate spline transformations, enables realistic warping, overlaying, and blending of garments onto user images while preserving intricate details such as embroidery and patterns.

(2) Future Directions:

This pioneering work marks a significant stride towards revolutionizing the interface of technology and fashion. Future research and development can build on the foundation laid by EthniTry in several ways:

Enhanced Personalization- Incorporating user-specific measurements and preferences to further refine the virtual try-on experience.

Broader Apparel Categories- Extending the system to include other types of garments and fashion accessories.

Real-Time Try-On- Developing real-time virtual try-on capabilities, allowing users to see the effects of different garments in real-time through augmented reality applications.

REFERENCES

[1] L. Zhu, D. Yang, T. Zhu, F. Reda, W. Chan, C. Saharia, M. Norouzi, and I. Kemelmacher-Shlizerman, "TryOnDiffusion: A Tale of Two UNets," arXiv, Jun. 14, 2023. [Online]. Available: https://arxiv.org/abs/2306.08276.

[2] B. Fele, A. Lampe, P. Peer, and V. Štruc, "C-VTON: Context-Driven Image-Based Virtual Try-On Network," in 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Jan. 2022, doi: 10.1109/wacv51458.2022.00226.

[3] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic Image Synthesis with Spatially-Adaptive Normalization," arXiv, Mar. 18, 2019. [Online]. Available: https://arxiv.org/abs/1903.07291

[4] J. Zeng, D. Song, W. Nie, H. Tian, T. Wang, and A. Liu, "CAT-DM: Controllable Accelerated Virtual Try-on with Diffusion Model," arXiv, Nov. 30, 2023. [Online]. Available: https://arxiv.org/abs/2311.18405

[5] S.-H. Shim, J. Chung, and J.-P. Heo, "Towards Squeezing-Averse Virtual Try-On via Sequential Deformation," arXiv, Dec. 26, 2023. [Online]. Available: https://arxiv.org/abs/2312.15861

[6] K. Li, J. Zhang, S.-Y. Chang, and D. Forsyth, "Wearing the Same Outfit in Different Ways -- A Controllable Virtual Try-On Method," arXiv, Nov. 29, 2022. [Online]. Available: https://arxiv.org/abs/2211.16989

[7] H. Rawal, M. J. Ahmad, and F. Zaman, "GC-VTON: Predicting Globally Consistent and Occlusion Aware Local Flows with Neighborhood Integrity Preservation for Virtual Try-on," arXiv, Nov. 7, 2023. [Online]. Available: https://arxiv.org/abs/2311.04932

[8] Z. Li, P. Wei, X. Yin, Z. Ma, and A. C. Kot, "Virtual Try-On with Pose-Garment Keypoints Guided Inpainting," in 2023 IEEE/CVF International Conference on Computer Vision (ICCV), 2023. [Online]. Available: https://openaccess.thecvf.com/content/ICCV2023/html/Li_Virtual_Try-On_with_Pose-Garment_Keypoints_Guided_Inpainting_ICCV_2023_paper.html

[9] A. Baldrati, D. Morelli, G. Cartella, M. Cornia, M. Bertini, and R. Cucchiara, "Multimodal Garment Designer: Human-Centric Latent Diffusion Models for Fashion Image Editing," arXiv, Apr. 4, 2023. [Online]. Available: https://arxiv.org/abs/2304.02051

[10] Z. Huang, H. Li, Z. Xie, M. Kampffmeyer, Q. Cai, and X. Liang, "Towards Hard-Pose Virtual Try-On via 3D-Aware Global Correspondence Learning," arXiv, Nov. 25, 2022. [Online]. Available: https://arxiv.org/abs/2211.14052