Abstract:

In the realm of fashion, the excitement of bulk clothing purchases often diminishes when faced with the challenge of envisioning and designing these garments. This leads to frustration, particularly when tailors misinterpret designs, resulting in stitched products that stray far from expectations. Additionally, uncertainty looms over the suitability of chosen designs for specific fabrics. To address these issues, we propose FashionVision, an innovative AI-driven recommendation system poised to transform the dress selection process.

FashionVision harnesses the power of advanced image generation techniques, particularly Conditional Generative Adversarial Networks (GANs) and Cyclic GANs, to revolutionize the way customers visualize and select dresses. By utilizing these models, FashionVision can seamlessly transform small fabric design patches into full dress images, empowering users to preview and tailor their chosen styles effortlessly.

The proposed evaluation metric for FashionVision will focus on user satisfaction and accuracy in predicting the desired dress design. Additionally, the system's ability to accurately recommend suitable designs for specific fabrics will be assessed. To train and validate FashionVision, datasets from Kaggle and IndoFashion will be utilized, providing a diverse range of fabric designs and styles for comprehensive model training and testing. Through FashionVision, we aim to alleviate the challenges faced by women in the dress selection process, enhancing their overall shopping experience and satisfaction.

Datasets:

IndoFashion stands out as a distinctive fashion dataset featuring a diverse array of Indian ethnic garments. As far as we know, this dataset represents the first large-scale collection for ethnic cloth classification. With a total of 106k images spanning 15 distinct categories, it encompasses the most renowned Indian ethnic attire. These images were meticulously gathered from Google image searches and various Indian e-commerce platforms, capturing a wide spectrum of styles, colors, and patterns.

Each image in the dataset comes with an annotation denoting its corresponding class category, ensuring clarity and organization. The distribution of images across categories is visually represented in Figure 3, while Figure 1 showcases sample images from each class label. This comprehensive dataset serves as a valuable resource for research and development in the realm of Indian ethnic fashion analysis and recommendation systems.
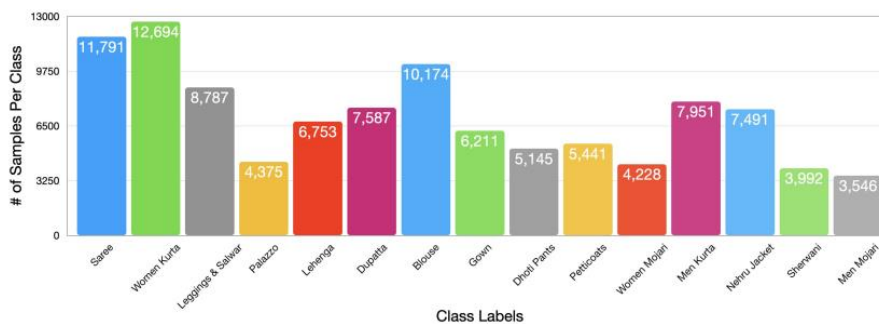
Figure 3: Figure shows category-wise frequency distribution of the 15 cloth categories in our proposed dataset.

The Indian ethnic wear dataset consists of over 8000 annotated images featuring various categories of clothing, including Kurta, Kurti, Flapper, Trouser, and Straight shirt. Each image within the dataset is meticulously annotated with bounding boxes, providing precise localization of the depicted clothing items. This dataset offers a comprehensive collection of Indian ethnic attire, enabling researchers and developers to delve into areas such as computer vision, object detection, and fashion analysis. With its diverse range of clothing categories and detailed annotations, this dataset serves as a valuable resource for advancing research and innovation in the field of computer vision and fashion technology.

Literature review:

Paper 1:

Citation in APA format: Rajput, P. S., & Aneja, S. (2021). IndoFashion: Apparel Classification for Indian Ethnic Clothes. arXiv e-prints. Retrieved from https://indofashion.github.io/assets/paper.pdf

Summary: This paper focuses on the challenge of accurately categorizing Indian ethnic attire due to its distinct styles and patterns compared to Western clothing. It aims to create a large-scale dataset and evaluate models for fine-grained classification of Indian ethnic clothes.

Methods: The authors evaluate various Image Classification models, including ResNet-18, ResNet-50, and ResNet-101, using different augmentations such as jitter and flipping.

Datasets: The primary dataset used is IndoFashion, comprising over 106K images of 15 different categories of Indian ethnic clothes.

Accuracy metrics: Evaluation is based on precision, recall, and F1-score for different ResNet backbone sizes with various augmentations.

Discussion: This study highlights the significance of specialized datasets and advanced augmentation techniques in addressing the challenges of apparel classification for Indian ethnic clothes. It emphasizes the importance of robust evaluation metrics for assessing model performance and guiding future research in this area.

Table -1: Sample Summarizing table

Paper	Method	Dataset	Evaluation
IndoFashion: Apparel	**ResNet-18, ResNet-50, ResNet-101**	**IndoFashion**	**Precision, Recall, F1-score**

Classification for Indian Ethnic Clothes			

Paper 2:

Summary: The paper introduces a novel fashion design task, aiming to transfer a reference appearance image onto a clothing image while preserving the structure of the clothing. The challenge arises from the absence of reference images for the newly designed output fashion images. The authors propose a diffusion model-based unsupervised structure-aware transfer method to semantically generate new clothes from a given clothing image and a reference appearance image. They decouple the foreground clothing with automatically generated semantic masks and use these masks as guidance in the denoising process to preserve structure information. Additionally, pre-trained vision Transformers (ViT) are employed for appearance and structure guidance.

Methods: The proposed method involves four key components: overview of fashion design with DiffFashion, mask generation by label condition, mask-guided structure transfer diffusion, and ViT feature guidance. Experimental setups include utilizing a label-conditional diffusion model pre-trained on the ImageNet dataset, with specific parameters and threshold settings for mask generation and guidance.

Datasets: The authors introduce a new dataset called OceanBag, comprising real handbag images and ocean animal images as reference appearances for generating new fashion designs. The dataset includes 6,000 photos of handbags and 2,400 marine scene images containing various marine organisms.

Accuracy Metrics: Due to the lack of existing automatic metrics suitable for evaluating fashion design across two natural images, the authors rely on human perceptual evaluation for validation, following existing appearance transfer/fashion design works.

Discussion: This paper contributes to the understanding of reference-based fashion design by proposing a novel diffusion model-based method that addresses the challenge of maintaining structure while transferring appearance. The method outperforms state-of-the-art baseline models, generating more realistic images in the fashion design task. The use of a new dataset and human perceptual evaluation for validation adds to the credibility of the findings. However, limitations may include the subjective nature of human evaluation and potential biases in dataset selection. Nonetheless, the proposed method offers promising solutions for reference-based fashion design with structure-aware transfer.

Paper   3:

Citation: Rostamzadeh, N., Hosseini, S., Boquet, T., Stokowiec, W., Zhang, Y., Jauvin, C., & Pal, C. (2018). Fashion-Gen: The Generative Fashion Dataset and Challenge. arXiv e-prints. Retrieved from https://arxiv.org/pdf/1806.08317v2.pdf

Summary: The paper introduces a new dataset comprising 293,008 high-definition fashion images paired with item descriptions provided by professional stylists. Each item is photographed from various angles. The research aims to address two primary objectives: 1) generating high-resolution images using progressive GAN (P-GAN) growing techniques, and 2) text-to-image synthesis using StackGAN-v1 and StackGAN-v2 models. Baseline results are provided for both tasks, and the paper outlines a challenge based on this dataset.

Methods: Two sets of experiments are presented: first, generating high-resolution images using the P-GAN technique, and second, text-to-image synthesis using StackGAN-v1 and StackGAN-v2 models.

Datasets: The Fashion-Gen dataset consists of 293,008 images, with 260,480 for training, 32,528 for validation, and 32,528 for testing. This dataset is larger than other available datasets for text-to-image translation tasks.

Accuracy Metrics: Inception Score is used as an evaluation metric, with results reported for Fashion Real data, StackGAN-v1, StackGAN-v2, and P-GAN models.

Discussion: The paper contributes to advancing research in high-resolution image generation and text-to-image synthesis by providing a large-scale dataset and baseline results for benchmarking. The use of professional stylist-provided item descriptions enhances the dataset's quality and realism. The experiments conducted demonstrate the effectiveness of P-GAN and StackGAN models in generating realistic fashion images. However, the paper's limitations may include potential biases in stylist-provided descriptions and the need for further exploration of novel techniques to improve image generation quality. Nonetheless, Fashion-Gen dataset and challenge offer valuable resources for researchers to explore and advance the state-of-the-art in fashion image generation.

Paper 4:

Summary: The paper presents MMFashion, an open-source toolbox based on PyTorch for visual fashion analysis. MMFashion supports various fashion analysis tasks, including attribute prediction, recognition and retrieval, landmark detection, parsing and segmentation, and compatibility and recommendation. It follows a modular design principle, making it easily extensible for customized modules. Detailed documentation, demo scripts, and pre-trained models are provided, facilitating the adoption of deep learning-based fashion analysis by non-expert users. MMFashion serves as a comprehensive platform for deploying existing models and developing new ideas in the field of visual fashion analysis.

Methods: MMFashion addresses multiple fashion analysis tasks, each with its specific methods. Fashion Attribute Prediction involves multi-label prediction, with top-k recall rate used for evaluation. Fashion Recognition and Retrieval determines if two images belong to the same clothing item, with variations for in-shop and consumer-to-shop retrieval scenarios. Fashion Landmark Detection detects key-points on clothes, while Fashion Parsing and Segmentation focuses on object detection and instance segmentation on cloth data.

Datasets: The paper utilizes the DeepFashion dataset, which contains over 800,000 diverse fashion images for attribute prediction, clothes retrieval, and landmark detection tasks. Additionally, the authors create annotations for clothes detection and segmentation, serving the Fashion Parsing and Segmentation task. This dataset is considered one of the most accurate open-source datasets in the fashion community.

Accuracy Metrics: Evaluation metrics include the standard top-k recall rate and accuracy for attribute prediction models. The recall rate and accuracy are computed for each attribute separately, comparing assigned attributes with ground-truth attributes.

Discussion: MMFashion contributes significantly to the field of visual fashion analysis by providing a comprehensive, flexible, and user-friendly toolbox. Its modular design and extensive documentation make it accessible to both experts and non-experts, fostering collaboration and innovation in the fashion analysis community. The paper's strengths lie in its detailed methodology, use of a diverse dataset, and clear evaluation metrics. However, potential limitations may include the need for further validation across different datasets and scenarios.

Overall, MMFashion serves as a valuable resource for researchers and practitioners, advancing the understanding and solutions in visual fashion analysis.

Paper 5:

Summary: The paper addresses the challenge of generating 3D human body models with clothing, which can accurately represent the complexity of dressed individuals in images and videos. Existing models lack the ability to generalize to dressed people and fail to capture the intricate geometry of clothing. The authors propose a novel approach called CAPE, a conditional Mesh-VAE-GAN model, which learns clothing deformation from 3D scans of clothed individuals. CAPE extends the SMPL body model to incorporate clothing as an additional term, allowing it to generate clothing for different body shapes and styles. The model preserves fine details like wrinkles using patchwise discriminators for 3D meshes.

Methods: Clothing is represented as a displacement layer on a graph topology similar to the SMPL body model. A conditional Mesh-VAE-GAN, incorporating graph convolutional neural networks, is used to learn the clothing deformation. The model's GAN component encourages the generation of visually plausible wrinkles, improving the quality of fine structures in the clothing.

Datasets: A dataset of 3D clothing is built by capturing temporal sequences of 3D human body scans, with corresponding minimally-clothed scans used to estimate body shapes under clothing. Noisy frames and failed registrations are removed manually.

Accuracy Metrics: Mean square error (MSE) is used to measure the reconstruction accuracy of the model compared to ground truth vertices, with adjustments made to eliminate the influence of factors like focal length and camera translation.

Discussion: The paper introduces a significant advancement in generative clothing models for 3D human body meshes, enabling realistic clothing generation for various body shapes and poses. CAPE addresses limitations of existing models and contributes to the understanding of clothing representation in 3D models, with potential applications in computer graphics, virtual try-on systems, and fashion industry simulations.

Papers	Methods	Datasets	Evaluations
DiffFashion: Reference-based Fashion Design with Structure-aware Transfer by Diffusion Models	DiffuseIT SpliceViT WCT2 STROTSS Ours	Handbag Marine life	
MMFashion: An Open-Source Toolbox for Visual Fashion Analysis	VGG-16 VGG-16 ResNet50 ResNet50	Polyvore-D Polyvore-D Polyvore	
Fashion-Gen: The Generative Fashion Dataset and Challenge	Fashion Real data 256 × 256 StackGAN-v1 StackGAN-v2 P-GAN (Karras et al., 2017)	-CelebA -CelebA-HQ -DeepFashion - Fashion ImageSynthesis -MS COCO -Caltech-UCSD Birds-200-2011 -Flowers Oxford-102 -Fashion dataset (ours)	
IndoFashion : Apparel Classification for Indian Ethnic Clothes		Atlas dataset [16] Jain et al. [14] IndoFashion(Ours)	
Learning to Dress 3D People in Generative Clothing	PCA CoMA-1 CoMA-4 Ours	CAPE Dataset	

References:

Rajput, P. S., & Aneja, S.: IndoFashion: Apparel Classification for Indian Ethnic Clothes. arXiv e-prints. Retrieved from https://indofashion.github.io/assets/paper.pdf (2021)

Cao, S., Chai, W., Hao, S., Zhang, Y., Chen, H., & Wang, G. (2023). DiffFashion: Reference-based Fashion Design with Structure-aware Transfer by Diffusion Models. arXiv e-prints. Retrieved from https://arxiv.org/pdf/2302.06826.pdf

Rajput, P. S., & Aneja, S.: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2021, pp. 3935-3939

Rostamzadeh, N., Hosseini, S., Boquet, T., Stokowiec, W., Zhang, Y., Jauvin, C., & Pal, C. (2018). Fashion-Gen: The Generative Fashion Dataset and Challenge. arXiv e-prints. Retrieved from https://arxiv.org/pdf/1806.08317v2.pdf

Liu, X., Li, J., Wang, J., & Liu, Z. (2020). MMFashion: An Open-Source Toolbox for Visual Fashion Analysis. arXiv e-prints. Retrieved from https://arxiv.org/pdf/2005.08847v2.pdf

Ma, Q., Yang, J., Ranjan, A., Pujades, S., Pons-Moll, G., Tang, S., & Black, M. J. (2020). Learning to Dress 3D People in Generative Clothing. arXiv e-prints. Retrieved from https://arxiv.org/pdf/1907.13615v3.pdf