DS 350 Research on Tradwives Fashion

Introduction

The concept of “tradwives,” short for traditional wives, has recently gained a large traction on TikTok, where curated videos showcasing idealized, and often unattainable, portrayals of motherhood dominate For You Pages. These portrayals frequently reinforce unrealistic expectations for mothers who do not conform to the privileged and romanticized image presented. Some possible factors that contribute to the appeal of these videos is the meticulously crafted aesthetic of their homes, coupled with distinctive fashion choices that often deviate from mainstream trends among younger women.

According to Google, "tradwife" is defined as “a woman who believes in and practices traditional gender roles and marriages. Some may choose to take a homemaking role within their marriage, and others leave their careers to focus on meeting their family’s needs in the home.” While many tradwives adopt 1950s-inspired fashion such as poodle skirts and swing dresses (Fig 1a), the fashion styles observed from some of the tradwive influencers’ content embraced styles that ranged between traditional to modern patterns. TikTok creators such as @naraazizasmith, @esteecwilliams, @ballerinafarm, @jennifer__tate__ exhibit a variety of styles, ranging from retro-inspired outfits to modern, comfortable clothing (Fig 1b).

Figure 1a. 1950s Fashion: Skirts & Dresses

Figure 1b. Different Types of Tradwife Fashion Styles.

With the increasing visibility of the tradwife movement on social media, we are able to investigate to better understand how these portrayals shape the public perceptions of femininity, domesticity, and gender roles. Social media platforms like TikTok are tools that many use to reinforce cultural narratives, and the way in which tradwives present themselves online play a key role in this process for the female audiences. The aesthetic appeal of their content, including their eye-catching outfits, home lifestyle routines, function as a form of self-expression and a strategic mechanism for engagement. This raises questions about how these patterns contribute to the current discussions surrounding traditional values and whether they promote a new vision of womanhood.

Fashion plays a particularly important role in constructing and spreading the tradwife identity. Clothing choices are not just stylistic preferences but serve as visual markers of adherence to traditional gender roles. According to Hanna, a website dedicated to sharing tradwife culture, traditional tradwife dresses typically feature high necklines, ankle length hemlines, and long sleeves, often made from natural fibers with simple patterns. In contrast, modernized tradwife dresses tend to include lower necklines, shorter lengths, and sleeveless designs, utilizing a wider range of materials and bolder patterns. Hanna also recognizes that there are hybrid styles that blend elements of both, such as ankle length dresses with lower necklines or short sleeves. These variations highlight the multifaceted nature of the tradwife community and how the newly emerging tradwife aesthetic is not monolithic but rather an evolving phenomenon.

By applying computer vision techniques to automatically identify and capture such stylistic choices and visual patterns, we can assess how these portrayals influence audience perceptions, and whether they resurge the conversations of gender norms in a digital space.

The concept of “tradwives,” short for traditional wives, has recently gained **a large** traction on TikTok, where curated videos showcasing idealized, and often unattainable, portrayals of motherhood dominate For You Pages. These portrayals frequently reinforce unrealistic expectations for mothers who do not conform to the privileged and romanticized image presented. **A major factor contributing** to the appeal of these videos is the meticulously crafted aesthetic of their homes, coupled with distinctive fashion choices that often deviate from mainstream trends among younger women.

While many tradwives adopt 1950s-inspired fashion, **their styles are far from homogenous**. TikTok creators such as @naraazizasmith, @esteecwilliams, @ballerinafarm, @jennifer__tate__ exhibit a variety of styles, ranging from retro-inspired outfits to modern, comfortable clothing (images below respectively).



This diversity raises questions about whether certain tradwives lean more toward modernized or traditional aesthetics in their daily lives. According to Hanna, a website dedicated to sharing tradwife culture, traditional tradwife dresses typically feature high necklines, ankle length hemlines, and long sleeves, often made from natural fibers with simple patterns. In contrast, modernized tradwife dresses tend to include lower necklines, shorter lengths, and sleeveless designs, utilizing a wider range of materials and bolder patterns. Hanna also recognizes that there are hybrid styles that blend elements of both, such as ankle length dresses with lower necklines or short sleeves. This variation highlights the multifaceted nature of the tradwife community, which can be systematically studied through the lens of computer vision.

Add stuff below into methods

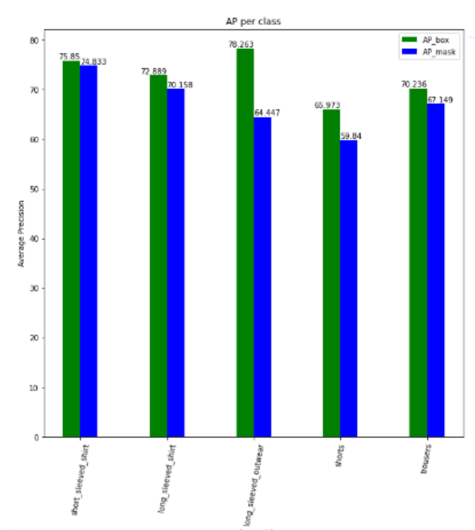
To analyze the visual patterns and fashion diversity within tradwife content, we employed Detectron2, a state-of-the-art computer vision library developed by Facebook AI Research

(FAIR) in 2019. Detectron2 leverages advanced object detection and segmentation algorithms, making it particularly well suited for tasks such as identifying clothing styles in visual image data. We also utilized CLIP (Contrastive Language-Image Pre-Training), a neural network developed by OpenAI. Unlike traditional models optimized for specific tasks, CLIP is trained on a diverse set of image-text pairs, enabling it to predict the most relevant textual descriptions for a given image. In fact, this is not the first application of Detectron2 and CLIP in the domain of fashion analysis. Previous implementations have demonstrated its potential in detecting and segmenting various types of clothing, offering valuable insights into its capabilities and limitations.

How Detectron2 & CLIP was Used for Fashion

One notable implementation is the **Monk_Object_Detection** repository, which utilizes Detectron2 in conjunction with the DeepFashion2 dataset. This dataset is specifically designed for clothing analysis, containing over 801k images annotated with detailed information about apparel categories such as shopping stores vs consumers and other labels. By training Detectron2 on this dataset, the repository demonstrated high accuracy in detecting and segmenting various clothing categories such as dresses, jackets, and tops. The results highlight Detectron2's ability to differentiate between overlapping clothing items and identify subtle features such as patterns and textures.

Another similar repository was **clothing-detection-segmentation-using-detectron2** that also utilized DeepFashion2 and Detectron2. Based on their evaluation results, their findings reported an average precision of approximately 75% per class, highlighting Detectron2's robustness in handling diverse fashion dataset.



For CLIP, its versatility in linking images and text has proven effective across many domains, but [researchers](#) have identified limitations in its ability to analyze retail fashion products. This challenge arises from the complexity of attributes such as style, color, pattern, and texture, which are critical for accurately matching images with text prompts. To address these shortcomings, researchers developed [Fashion CLIP](#), a fine-tuned version of CLIP tailored specifically for retail applications. This modified model accounts for nuanced details often overlooked by the original CLIP, significantly improving its ability to handle ambiguous search queries and enhance alignment between images and text prompts.

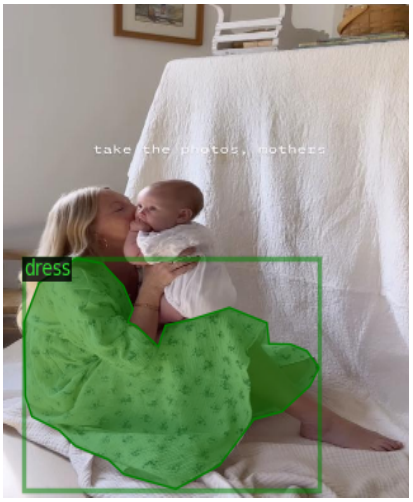
CLIP Model: blue floral dress

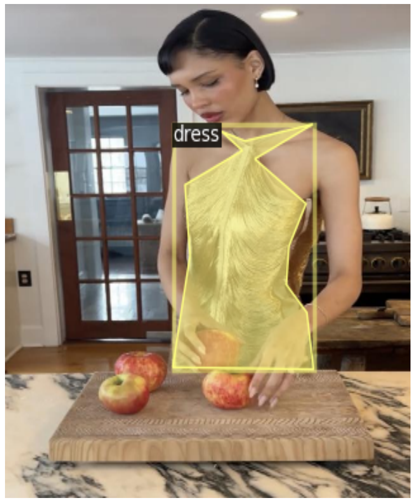
FashionCLIP Model: blue floral dress

Methodology

The primary goal of this analysis was to determine whether Detectron2 is able to learn from TikTok tradwife images and accurately identify various clothing items. To collect data, frames were extracted from TikTok tradwive videos at the midpoint of each detected shot change using Google Cloud Video Intelligence (credit to Sandy for sharing her code and data). Over 20 tradwife creators were sampled to ensure diversity in fashion styles. This process initially yielded 25,767 images. After filtering and cleaning the data, removing duplicates and keeping only the images that clearly show the clothing worn on screen, the final dataset was reduced to 375 images.

The finalized dataset was then uploaded to [Roboflow](#), where object detection annotations were created. The dataset included 10 distinct clothing classes: 1) apron, 2) dress, 3) head garment, 4) long sleeve top, 5) overalls, 6) pants, 7) robe, 8) short sleeve top, 9) skirt, and 10) waist apron. Annotations were created to outline clothing items, avoiding general bounding box methods that might lead to inaccuracies due to overlapping styles. This approach was intended to improve the accuracy and performance of the Detectron2 model during training. Examples below show the annotated images.



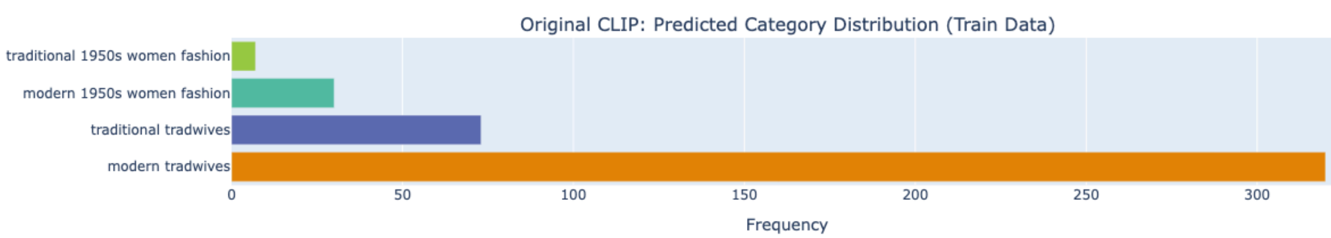


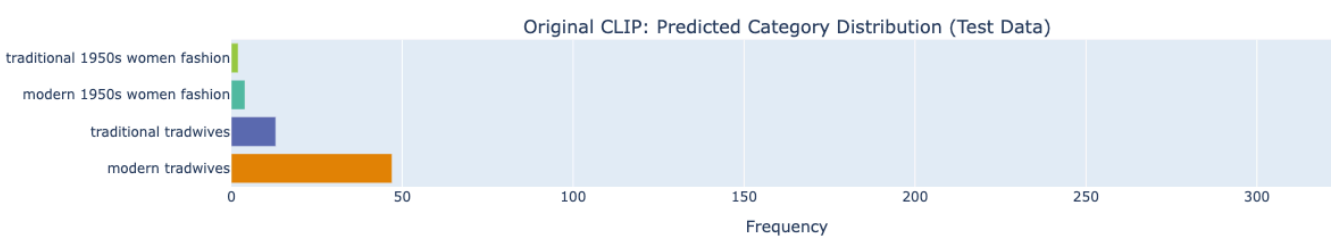
Although I encountered technical difficulties that prevented me from fully testing Detectron2's ability to learn from TikTok images, I had planned to integrate Detectron2 with CLIP for further analysis. Specifically, I intended to use the output of Detectron2's label predictions, such as the bounding boxes, clothes labels, and performance scores, and feed this information into CLIP for additional evaluation. By combining these two models, I aimed to leverage Detectron2's precise object detection with CLIP's image to text matching capabilities, offering deeper insights into how tradwife fashion styles align with modern or traditional aesthetics.

Despite these limitations, I proceeded to evaluate the CLIP and Fashion CLIP models independently using the dataset created through Roboflow. The dataset was formatted in COCO (Common Objects in Context) that is widely used for object detection, segmentation, and recognition tasks with labeled images. The final dataset of 375 images was divided into training, validation, and testing sets with a 70:20:10 split, respectively. See image below to see the distribution of these splits.

In each annotation JSON file, it contains detailed information about the images, categories, and bounding boxes. Using this information, the images were cropped according to their bounding box boundaries. Each cropped image was paired with 1 of 4 text prompts: 1) modern 1950s women fashion, 2) traditional 1950s women fashion, 3) modern tradwives, and 4) traditional tradwives. The results were stored in a Pandas dataframe for visualization and analysis. Both the training and test datasets were used to evaluate whether sample size impacts its outcomes.

Results





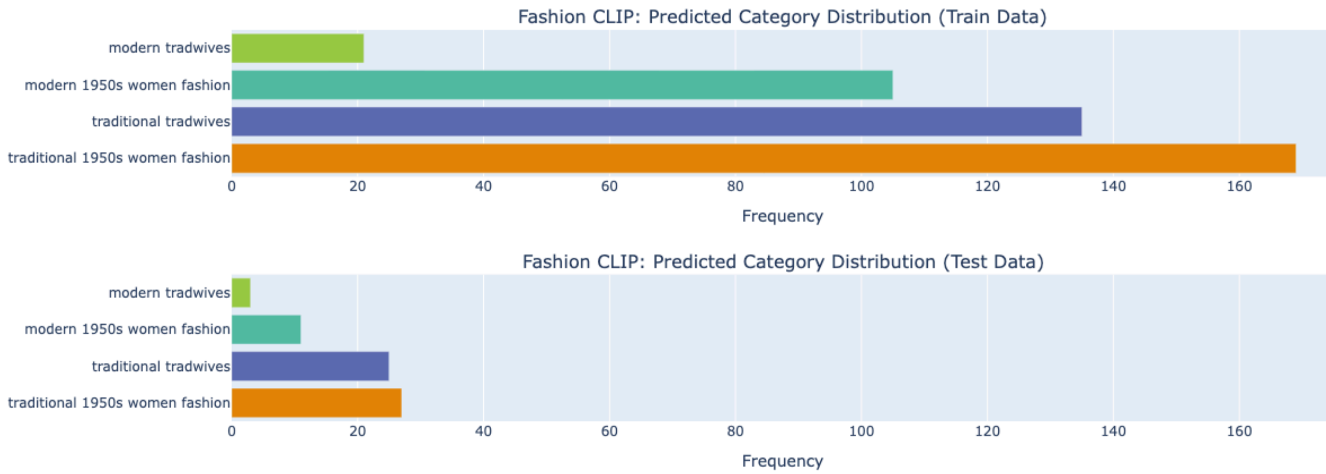
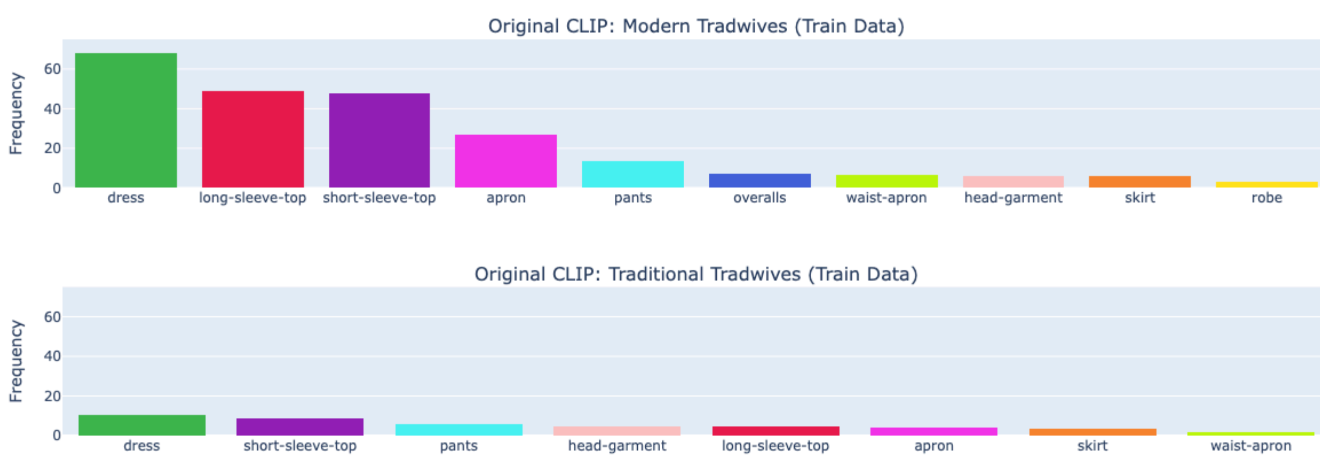


Figure 1a, 1b. Original CLIP vs Fashion CLIP: Predicted Category Distribution of Training and Test Data.

Above, Figure 1a and 1b present the predicted category distributions for each CLIP model's performance. For the original CLIP model, the top 2 predicted categories were "modern tradwives" and "traditional tradwives," with "modern tradwives" significantly higher than the remaining categories. In contrast, Fashion CLIP's top 2 predicted categories were "traditional 1950s women fashion" and "traditional tradwives." This suggests that the original CLIP model places greater emphasis on identifying tradwife-related categories, while Fashion CLIP focuses more broadly on traditional aesthetics. These results align with previous studies highlighting the limitations of the original CLIP model in capturing nuanced attributes, as Fashion CLIP demonstrates a more refined ability to differentiate traditional styles.



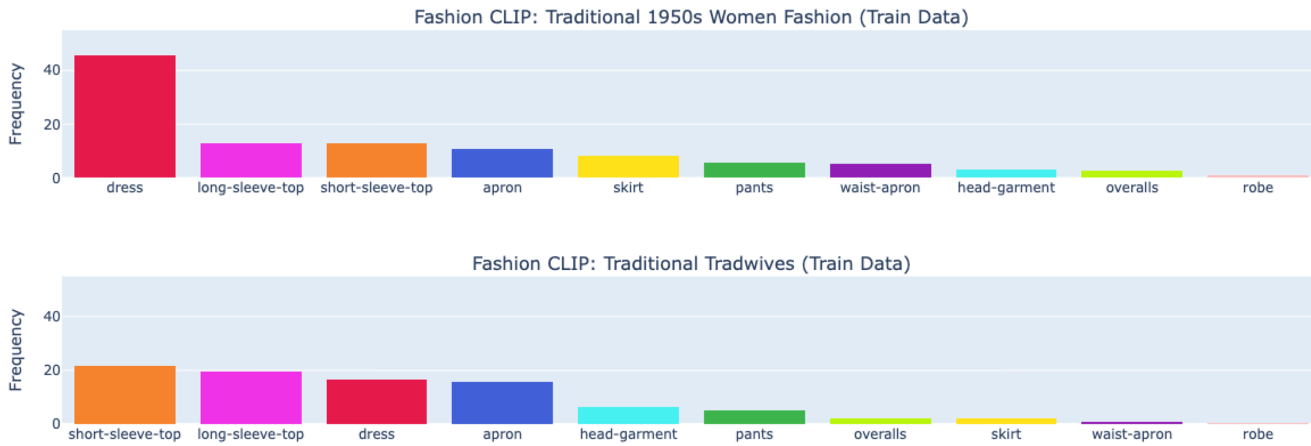
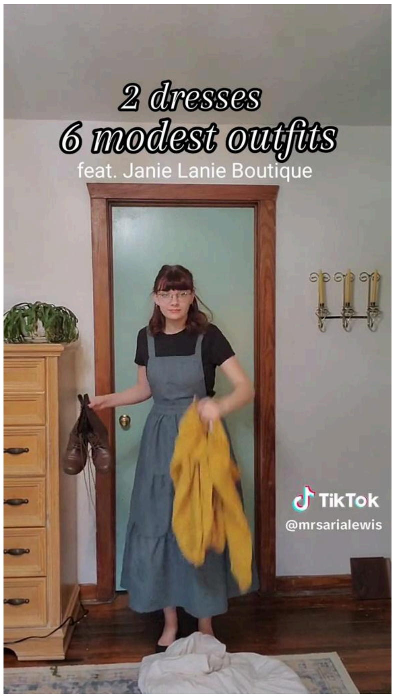
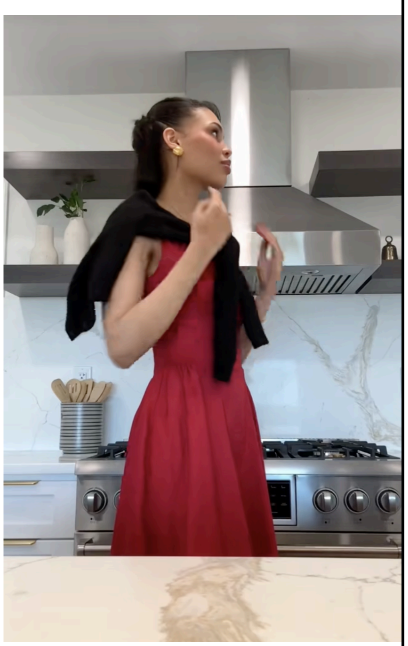


Figure 2a, 2b. Original CLIP vs Fashion CLIP: Class Distribution of Top 2 Categories of Training Data.

When examining the probability distributions for each label (Figure 2a and 2b), both models consistently identified "dress" as the category with the highest probability, aligning well with the text prompts. To further compare, I selected examples representing the lowest, middle, and highest probability scores under the "dress" label for the top category predicted by each model. The table below illustrates these scores alongside the corresponding images.

<table>
<tr><th colspan="3">Original CLIP: Modern Tradwives</th></tr>
<tr><td>Prob = 0.357</td><td>Prob = 0.798</td><td>Prob = 0.981</td></tr>
<tr><td>
</td><td>
</td><td>
</td></tr>
<tr><th colspan="3">Fashion CLIP: Traditional 1950s Women Fashion</th></tr>
<tr><td>Prob = 0.465</td><td>Prob = 0.759</td><td>Prob = 0.956</td></tr>
<tr><td></td><td></td><td></td></tr>
</table>

Interestingly, while the original CLIP model showed less differentiation across the probability scores, Fashion CLIP revealed clear stylistic progressions at each level. This highlights Fashion CLIP's potential for deeper analysis of fashion styles, enabling researchers to explore how specific image attributes align with various text prompts.

Figure 3a, 3b. Original CLIP vs Fashion CLIP: Class Distribution of Top 2 Categories of Testing Data.

Figure 3a and 3b display the test dataset results for both models, showing the variations in the distribution of category matches. For the original CLIP model, the top 4 highest matching categories (e.g. "dres", "long-sleeve-top", "short-sleeve-top", "apron") differed slightly in order between the training and testing datasets, suggesting that larger datasets may improve generalization. In contrast, Fashion CLIP exhibited more substantial shifts in category rankings, with lower ranking classes from the training data becoming more prominent in the testing data. This may reflect Fashion CLIP's greater sensitivity to variations in fashion styles, as opposed to the original CLIP's reliance on textual dependencies.

Limitations and Future Steps

Currently, the Detectron2 model is running on a Colab Notebook. To successfully complete training, significantly more time (approximately 3.5 days with the current dataset) and computing power are required. The notebook environment has presented a few errors and warnings during execution, indicating the need for a more robust setup. A practical next step would be to transfer the code to the lab's computer and submit a job request for more efficient and reliable execution. Alternatively, setting up an SSH connection to the Wellesley's remote server could provide a flexible and scalable solution for running the model, offering greater computational resources without relying solely on local hardware.

Additional limitations include the potential for implicit bias during the initial data screening stage. Increasing the number of human reviewers involved in filtering and cleaning the dataset could help ensure that the dataset better represents the diversity of tradwives' fashion. Another critical improvement would be to expand the dataset size significantly. Testing with 1,000 images or more could yield better results and enhance model performance.

Future steps also involve exploring a wider variety of text prompts to analyze CLIP's responses more comprehensively. This exploration could include prompts focusing on aesthetic intentions or occasion specific purposes to refine the official text prompts for larger datasets. In fact, depending on the focus of the text prompts, the general CLIP model may outperform Fashion CLIP in certain cases. Investigating these cases further could provide valuable insights into the strengths and limitations of both models.

Lastly, to better understand the relationships between the tradwives fashion, some statistical tests could be applied such as regression analysis or ANOVA to determine whether differences in performance scores are statistically significant across different models or dataset splits. We could also try correlation analysis to assess the relationship strengths between bounding box characteristics/labels and model performance metrics. Incorporating these statistical analysis would provide more credibility in my findings and guide us further while exploring the two models.

Colab Notebook of my code: fashion train.ipynb