# Final Project Proposal

October 24, 2025

**Ethan Senatore** **Gbemiga Oloyede**

**Abstract**

The intersection of fashion and computer vision presents a unique opportunity for fine-grained visual understanding. In this project, we evaluate how model complexity influences performance on a designer classification dataset consisting of 50,000 runway images across 50 fashion labels. Our approach benchmarks traditional architectures, such as ResNet-18, against lightweight convolutional variants to study the trade-offs between depth, accuracy, and interpretability. We hypothesize that smaller, well-structured networks can achieve competitive performance with improved generalization. The ultimate goal is to establish baseline results for this dataset and provide insight into how model capacity affects fine-grained visual categorization in the fashion domain.

## I. Motivation

In recent years, the practical uses of leading computer vision methods has grown increasingly useful for a variety of industries. One of them, surprisingly, is the fashion industry. Various startups have sprung up making use of advanced methods in the computer vision field to tackle design, operation, and shopping challenges. The ability to add images as another another dimension of analysis has piqued the interest of many major companies as well, most notable Amazon, Google, and Meta with their battle over the E-commerce space.

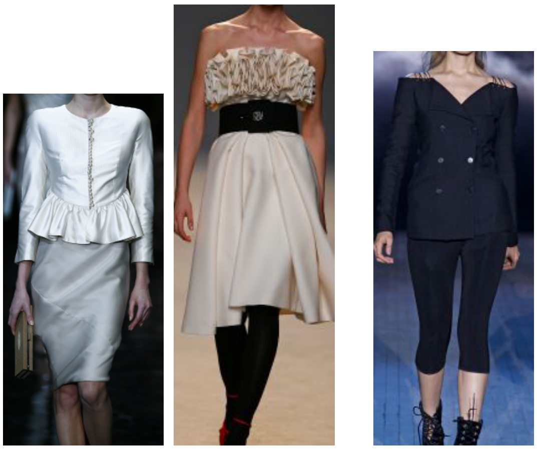
Fashion has proven to be yet another domain touched by the ever-expanding reach of machine learning and computer vision. The data it affords to researchers and engineers is unique in its nature, as fashion imagery offers both structure and abstraction. On one hand, images in this domain contain clearly identifiable and semantically meaningful elements such as, garments, accessories, and they ways they are worn. On the other hand, the subjective and variable nature of fashion introduces significant complexity making it a challenging task to segment, classify, or localize attributes within an image. Furthermore, identifying and modeling relationships between garments based on color, texture, pattern or style remains both a difficult and interesting problem.

## II. Related Works

Recent advances in computer vision for fashion analysis have largely centered around fine-grained visual categorization (FGVC) and attribute recognition. Jia *et al.* [4] introduced **Fashionpedia**, an ontology-driven dataset combining segmentation and fine-grained attribute localization, while Guo *et al.* [2] presented the **iMaterialist Fashion Attribute Dataset**, demonstrating how attribute-level annotations can enhance transfer learning within fashion tasks. Broader surveys such as Cheng *et al.* [1] emphasize the growing intersection between fashion and computer vision, outlining challenges in dataset diversity, scalability, and interpretability. Beyond fashion, methods in general FGVC, such as Zhang *et al.* [6], have focused on improving feature discrimination and sample efficiency through meta-learning strategies, which are relevant to the fine-grained nature of designer classification. More recently, Parekh *et al.* [5] and Han *et al.* [3] explored attribute extraction and multimodal pretraining to better capture detailed visual semantics in apparel imagery. While these works highlight significant progress in fine-grained fashion understanding, we found no prior research utilizing our specific dataset or exploring designer-level identification as a classification task. Unlike large-scale attribute localization or retrieval problems, this work focuses on a more constrained and exploratory setting, aiming to establish baseline performance on what can be viewed as a toy dataset for designer recognition.

## III. Dataset

The dataset we are interested in comes from a FGCV (Fine-Grained Visual Categorization) competiton that was a part of CVPR 2019. The data consists of 50,000 high-resolution runway images from various fashion shows across the world. Each images' label is the designer of the piece depicted. Additionally, we are using a cropped dataset of the images, but another interesting experiment could be looking into the performance of our model on the cropped and uncropped datasets. Figure I showcases three example images from the dataset.



(a) Armani (b) Valli (c) Alexander Wang

Figure I: Three example runway images from the iDesigner dataset.

Designers from this dataset include Alexander Mcqueen, Gucci, Prada, Alexander Wang, and more. Overall, this dataset presents us with a unique challenge and opportunity to uncover latent stylistic patterns associated with specific designers within a high-dimensional feature space.

## IV. Methods

At a high level, our approach involves benchmarking several architectures that are both conventional and lightweight on our fashion dataset. For example, a traditional network such as ResNet-18 will serve as one of our baselines. Our own models will be designed with an emphasis on parameter efficiency and interpretability. We aren't necessarily setting out to design a single best architecure, but to systematically study how reducing network depth or width affects both accuracy and feature localization quality within this specific visual domain.

Since we are still in our initial phase of this project, we have not finalized an exact architecture to to use. Over the next phase of the project, we will conduct various tests on different model designs stemming from known backbones. Ideally, we can progressively adapt them based on validation performance, computational cost, and qualitative behavior. We also may look into how pretraining on ImageNet or other fashion-specific data influences transfer performance when we are constrained by model capacity.

For now, our initial hypothesis is that a carefully designed smaller CNN can reach comparable performance to deeper architectures on the designer classification task, due to the strong visual regularities and localized features present in fashion imagery. We will seek to ask more questions as we dive deeper into the data.

## V. Analysis

To analyze performance, we will evaluate each model using standard classification metrics such as top-1 accuracy, validation loss, and confusion matrices. Beyond raw accuracy, we will compare the parameter count of each of the models, training time, and computational efficiency. We will also seek to study the feature localization and interpretability across the architectures. Finally, combining both analyses we will attempt to determine how model depth, width, and pretraining influences not only classification accuracy but also the quality of learned representations within the fashion domain.

## VI. Preliminary Testing

To obtain an initial baseline, we trained a ResNet-18 model from scratch on a subset of the dataset containing 10,000 training images and 5,000 validation images. Each image was resized to (224 × 224). The model was trained for 20 epochs using the AdamW optimizer and cross-entropy loss.

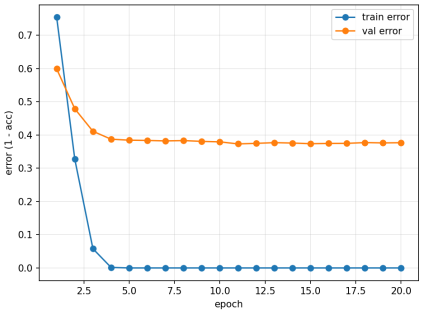


Figure II: Training and validation error for ResNet-18 trained from scratch.

As can be seen in Figure II, the training error rapidly decreases within the first 4-5 epochs, reaching nearly zero thereafter. However, the validation error plateaus around 0.38 (i.e., ≈ 62% accuracy), showing a clear performance gap between training and validation. However, this could be due to the small amount of training data we provided given the complex task. The good news we took from this was that the quick convergence also implies that the model capacity is sufficient for this dataset subset.

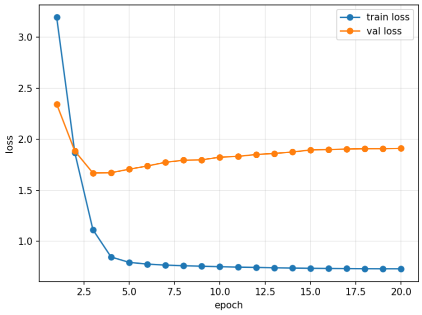
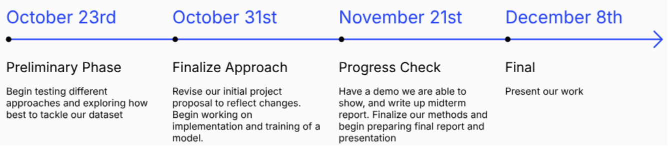


Figure III: Training and validation loss for ResNet-18 trained from scratch.

In Figure III, we see that the training loss decreases sharply and stabilizes around 0.7, while the validation loss bottoms out near epoch 4 and then gradually increases. This behavior is consistent with the validation error trend and further confirms overfitting. The increase in validation loss despite stable validation accuracy suggests the model's predictions become increasingly confident on misclassified samples, a common symptom of poor generalization.

## VII. Timeline



October 23rd

Preliminary Phase

Begin testing different approaches and exploring how best to tackle our dataset

October 31st

Finalize Approach

Revise our initial project proposal to reflect changes. Begin working on implementation and training of a model.

November 21st

Progress Check

Have a demo we are able to show, and write up midterm report. Finalize our methods and begin preparing final report and presentation

December 8th

Final

Present our work

Figure IV: Proposed Timeline

# References

[1] Wen-Huang Cheng, Sijie Song, Chieh-Yun Chen, Shintami Chusnul Hidayati, and Jiaying Liu. Fashion meets computer vision: A survey. *ACM Computing Surveys (CSUR)*, 54(4):1–41, 2021.

[2] Sheng Guo, Weilin Huang, Xiao Zhang, Prasanna Srikhanta, Yin Cui, Yuan Li, Matthew R. Scott, Hartwig Adam, and Serge Belongie. The imaterialist fashion attribute dataset, 2019.

[3] Yunpeng Han, Lisai Zhang, Qingcai Chen, Zhijian Chen, Zhonghua Li, Jianxin Yang, and Zhao Cao. Fashionsap: Symbols and attributes prompt for fine-grained fashion vision-language pre-training. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 15028–15038, 2023.

[4] Menglin Jia, Mengyun Shi, Mikhail Sirotenko, Yin Cui, Claire Cardie, Bharath Hariharan, Hartwig Adam, and Serge Belongie. Fashionpedia: Ontology, segmentation, and an attribute localization dataset. In *European conference on computer vision*, pages 316–332. Springer, 2020.

[5] Viral Parekh, Karimulla Shaik, Soma Biswas, and Muthusamy Chelliah. Fine-grained visual attribute extraction from fashion wear. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3973–3977, 2021.

[6] Yabin Zhang, Hui Tang, and Kui Jia. Fine-grained visual categorization using meta-learning optimization with sample selection of auxiliary data. In *Proceedings of the european conference on computer vision (ECCV)*, pages 233–248, 2018.