# Artificical Intelligence in Fashion with Computer Vision

Sini Suresh, M1-DATAAI
IP Paris

Supervised by: Vicky Kalogeiton, Damien Rohmer

March-June 2021

# Abstract

Fashion is how we present ourselves to the world and it has become one of the world's largest industries. The remarkable results of Computer Vision with AI have led to advances in fashion, with several real-life applications, such as creating new fashion styles, detecting fashion attributes in images or analysing existing styles, systems recommending fashion compatibility or predicting future fashion trends.

The goal of this research project is to identify the existing models and advancement in the field of computer vision in fashion and explore the the future scope in this area. This project also aims at identifying the best dataset for further research and that can help develop models focused particularly on fashion popularity or fashion trends forecasting.

# Contents

# 1 Introduction

Recent progress in the field of computer vision has advanced machines' ability to recognize and understand the visual world, showing significant impacts in various fields including fashion. Each day billions of photographs are uploaded to photo-sharing services and social media platforms. These images are packed with information about how people live around the world. Researchers across the globe exploit this rich trove of data to understand fashion and style trends worldwide. The opportunity is magnified by the massive scale at which humans are generating cultural artifacts on social media, and by the increasing power of machine learning techniques. For instance, by applying natural language processing to millions of Twitter messages, we can discover relationships between time of day and mood that leverage sample sizes much larger than those of any traditional study.

In this research project the focus is on gathering the information on existing methodologies adopted in fashion with computer vision. This is a literature survey to list out the datasets used, models built and to identify the latest network architectures that can be applied to analyze and predict trends in fashion. The aim of this project is to extract meaningful insights from a set of latest papers on fashion and a set of technical papers that are within the scope of being applied to the fashion industry.

Technically, intelligent fashion is a challenging task because, unlike generic objects, fashion items suffer from significant variations in style and design. Current studies on intelligent fashion cover the research topics not only to detect what fashion items are presented in an image but also to analyze the items, synthesize creative new ones, and finally provide personalized recommendations. Thus, in this report, section 2 summarises the organization of research topics.
Individuals make fashion choices based on many factors, including geography, weather, culture, and personal preference. These factors are studied in the papers listed in the section 3, Fashion Papers.
Despite recent progress, investigating and modeling complex real-world problems when developing intelligent fashion solutions remain challenging. Hence, section 4 details some of the network architectures that can be applied to analyse and synthesise fashion preferences, apart from the commonly applied recurrent neural networks.

| Fashion Papers Inference | | | | | |
|---|---|---|---|---|---|
| # | Paper | Year | Dataset | #of Photos | Key Features |
| 1 | Paris to Berlin [18] | 2020 | GeoStyle | 7.7M | Influence between major cities of the world. Spatio-temporal influences. |
| 2 | Modeling Fashion Influence from Photos [19] | 2020 | GeoStyle, Amazon | 7.7M, 41K | Fashion influence along geolocation and fashion brands. Influential entities in terms of propagating their styles. |
| 3 | Fashion Forward [20] | Aug 2020 | Amazon, DeepFashion | 80000, 200000 | Forecast future of fashion (1-2 years) .Unsupervised manner. |
| 4 | Fashionpedia [4] | July 2020 | Fashionpedia | 48825 | Fashion ontology and dataset - Fashionpedia. Attribute-Mask R-CNN model - instance segmentation and localized attribute recognition. |
| 5 | Knowledge Enhanced Neural Fashion Trend Forecasting [12] | Sept 2020 | FIT, GeoStyle | 680K, 7.7M | Forecast fashion trends of people in various groups. Time series data of fashion elements using the Long-Short Term Memory (LSTM) encoder-decoder framework. |

| Technical Papers Inference | | | |
|---|---|---|---|
| # | Paper | Year | Key Features |
| 6 | Attention is All you Need [23] | Dec 2017 | Novel network architecture model - Transformer, based on self attention mechanisms. DiscreteVariational Encoder (dVAE) for the input. |
| 7 | An Image is Worth 16X16 Words: Transformers For Image Recognition At Scale [1] | Oct 2020 | Inspired by the Transformer, an image is split into fixed-size patches, linearly embed them, add position embeddings and feed the resulting sequence of vectors to a standard Transformer encoder. |
| 8 | Dall-e [43] | Feb 2021 | Zero-Shot Text-to-Image Generation. Text to image generation based on a transformer that autoregressively models the text and image tokens as a single stream of data. |

# 6 Discussion

Fashion-related papers discussed above have given state-of-the-art results in their respective domains. However, datasets used in these experiments such as GeoStyle dataset, like any internet photo dataset, have certain biases in terms of the demographics of the people who have uploaded photos and the location.

Fashion Forward focuses on forecasting the future of fashion over a 1- 2 year time course. In this horizon, we expect consumer purchase behavior to be the foremost indicator of fashion trends. This is again limited to a very short term forecasting. Fashion in Computer Vision should aim at predicting/ forecasting the trends over a longer period which will help the fashion industry and the related textile industry to better manage their production.

Also, Knowledge Enhanced Neural Fashion Trend Forecasting (KERN) is dependent on various external tools for object detection. As the tags come from an existing tagging tool, this might contain some noise and result in a small bias of real fashion trends.
Again, as mentioned, the method performs better in forecasting shorter-term (half-year) prediction than a longer-term (one-year) prediction on the FIT dataset, including the KERN model. This is mostly because future prediction requires forecasting data with a longer time horizon and, such a setting, in the case of FIT dataset, reduces the quantity of training data.

Even though, Fashionpedia is one of the most recent and better datasets, the number of photos is comparatively less. However, the ability of Fashionpedia dataset to detect fine grained attributes can be utilised for further research using novel approaches in computer vision such as Transformers or a more improved model, Dall-e.

As the transformers prove to perform better the recurrent neural networks in various scenarios, this can be applied in fashion to see if we can obtain better results in fashion trend forecasting. Transformers can be applied, for instance, in the place or LSTMs in [12] along with time series information to retrieve better results in prediction. Moreover, as we have seen the example of "an illustration of a baby hedgehog in a Christmas sweater walking a dog" in Dall-e [43], this can be applied by Fashion Designers to obtain unseen designs and better enhance their ideas.

# 7 Conclusion

With the significant advancement of information technology, research in computer vision and its applications in fashion have become an important topic and received great attention. Meanwhile, the enormous amount of data generated by social media platforms and e-commerce websites provide an opportunity to explore knowledge relevant to the development of intelligent fashion techniques.

This research project details the different research areas in fashion and also presents some of the prominent and latest papers that attempt to analyse and forecast fashion trends. This project also covers the major datasets and ontology like Fashionpedia that can be seen as one of the major advancements in this area. At present all the papers in fashion with computer vision revolve around recurrent neural networks. With innovative models like Dall-e [43], it would also be interesting to combine visual data with these types of architectures and new datasets to explore new types of as-yet-unseen connections.

The combination of big data, machine learning, computer vision, and automated analysis algorithms, would make for a very powerful analysis tool more broadly in the visual discovery of fashion and many other areas.

# References


[1] Jakob Uszkoreit Neil Houlsb Alexey Dosovitskiy Lucas Beyer Alexander Kolesnikov Dirk Weissenborn Xiaohua Zhai Thomas Unterthiner Mostafa Dehghan Matthias Minderer Georg Heigold Sylvain Gelly. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale". In: (2020).

[2] U. Mall K. Matzen B. Hariharan N. Snavely K. Bala. "GeoStyle: Discovering Fashion Trends and Events". In: (2019).

[3] A. Veit B. Kovacs S. Bell J. McAuley K. Bala S. Belongie. "Learning Visual Clothing Style with Heterogeneous Dyadic Co-Occurrences". In: (2015).

[4] Menglin Jia Mengyun Sh Mikhail Sirotenko Yin Cui Claire Cardie Bharath Hariharan Hartwig Adam Serge Belongie. "Fashionpedia: Ontology, Segmentation, and an Attribute Localization Dataset". In: (2020). DOI: https://arxiv.org/pdf/2004.12276.pdf.

[5] M. H. Kiapour X. Han S. Lazebnik A. C. Berg T. L. Berg. "Where to Buy It: Matching Street Clothing Photos in Online Shops". In: (2015).

[6] G. Pons-Moll S. Pujades S. Hu M. J. Black. "ClothCap: Seamless 4D Clothing Capture and Retargeting". In: (2017).

[7] P. Guan L. Reiss D. A. Hirshberg A. Weiss M. J. Black. "DRAPE: DRessing Any PErson". In: (2012).

[8] Z. Wu G. Lin Q. Tao J. Cai. "M2E-Try On Net: Fashion from Model to Everyone". In: (2019).

[9] Y. Ma X. Yang L. Liao Y. Cao and T. Chua. "Who, Where, and What to Wear?: Extracting Fashion Knowledge from Social Media". In: (2019).

[10] H.-J. Chen K.-M. Hui S.-Y. Wang L.-W. Tsao H.-H. Shuai W.-H. Cheng. "BeautyGlow: On-Demand Makeup Transfer Framework with Reversible Generative Network". In: (2019).

[11] X. Yang Y. Ma L. Liao M. Wang T. Chua. "TransNFCM: Translation-Based Neural Fashion Compatibility Modeling". In: (2019).

[12] Yunshan Ma Yujuan Ding Xun Yang Lizi Liao Wai Keung Wong Tat-Seng Chua. "Knowledge Enhanced Neural Fashion Trend Forecasting". In: (2020). DOI: https://arxiv.org/pdf/2005.03297.pdf.

[13] X. Han Z. Wu P. X. Huang X. Zhang M. Zhu Y. Li Y. Zhao L. S. Davis. "Automatic Spatially-aware Fashion Concept Discovery". In: (2017).

[14] X. Han Z. Wu Z. Wu R. Yu L. S Davis. "VITON: An Image-based Virtual Try-on Network". In: (2018).

[15] P. Isola J. Zhu T. Zhou A. A. Efros. "Image-to-Image Translation with Conditional Adversarial Networks". In: (2017).

[16] S. Jiang Y. Wu Y. Fu. “Deep Bi-directional Cross-triplet Embedding for Cross-Domain Clothing Retrieval”. In: (2016).

[17] Z. Al-Halah R. Stiefelhagen K. Grauman. “Fashion Forward: Forecasting Visual Style in Fashion”. In: (2017).

[18] Ziad Al-Halah Kristen Grauman. “From Paris to Berlin: Discovering Fashion Style Influences Around the World”. In: (2020). DOI: https://openaccess.thecvf.com/content_CVPR_2020/papers/Al-Halah_From_Paris_to_Berlin_Discovering_Fashion_Style_Influences_Around_the_CVPR_2020_paper.pdf.

[19] Ziad Al-Halah Kristen Grauman. “Modeling Fashion Influence from Photos”. In: (2020). DOI: https://www.cs.utexas.edu/~ziad/papers/tmm_2020_fashion_influence_from_photos.pdf.

[20] Ziad Al-Halah Rainer Stiefelhage Kristen Grauman. “Fashion Forward: Forecasting Visual Style in Fashion”. In: (2020). DOI: https://openaccess.thecvf.com/content_ICCV_2017/papers/Al-Halah_Fashion_Forward_Forecasting_ICCV_2017_paper.pdf.

[21] G. Balakrishnan A. Zhao A. V. Dalca F. Durand J. Guttag. “Synthesizing Images of Humans in Unseen Poses”. In: (2018).

[22] E. Simo-Serra H. Ishikawa. “Fashion Style in 128 Floats: Joint Ranking and Classification using Weak Data for Feature Extraction”. In: (2016).

[23] Ashish Vaswani Noam Shazeer Niki Parmar Jakob Uszkoreit Llion Jones Aidan N. Gomez Łukasz Kaiser. “Attention is All You Need”. In: (2017).

[24] A. Veit S. Belongie T. Karaletsos. “Conditional Similarity Networks. In CVPR”. In: (2017).

[25] K. Vaccaro S. Shivakumar Z. Ding K. Karahalios R. Kumar. “The Elements of Fashion Style”. In: (2016).

[26] WZ. Lin H. Xie P. Kang Z. Yang W. Liu Q. Li. “Cross-Domain Beauty Item Retrieval via Unsupervised Embedding Learning”. In: (2019).

[27] Y. Kalantidis L. Kennedy L. Li. “Getting the Look: Clothing Recognition and Segmentation for Automatic Product Suggestions in Everyday Photos”. In: (2013).

[28] K. Gong Y. Gao X. Liang X. Shen M. Wang L. Lin. “Graphonomy: Universal Human Parsing via Graph Transfer Learning”. In: (2019).

[29] T. Li R. Qian C. Dong S. Liu Q. Yan W. Zhu L. Lin. “BeautyGAN: Instance-level Facial Makeup Transfer with Deep Generative Adversarial Network”. In: (2018).

[30] Wen-Huang Cheng Sijie Song Chien-Yun Chen Shintami Chusnul Hidayati Jiaying Liu. “Fashion Meets Computer Vision: A Survey”. In: (2021). DOI: https://arxiv.org/pdf/2003.13988.pdf.

[31] S. Zhu R. Urtasun S. Fidler D. Lin C. C. Loy. “Be Your Own Prada: Fashion Synthesis with Structural Coherence”. In: (2017).

[32] Yuying Ge Ruimao Zhang Xiaogang Wang Xiaoou Tang Ping Luo. “DeepFashion2: A Versatile Benchmark for Detection, Pose Estimation, Segmentation and Re-Identification of Clothing Images”. In: (2020).

[33] Y.Shih K.Chang H.Lin M.Sun. “Compatibility Family Learning for Item Recommendation and Generation”. In: (2018).

[34] X. Song F. Feng J. Liu Z. Li L. Nie J. Ma. “Neurostylist: Neural Compatibility Modeling for Clothing Matching”. In: (2017).

[35] W. Yang M. Toyoura X. Mao. “Hairstyle Suggestion using Statistical Learning”. In: (2012).

[36] C. Corbiere H. Ben-Younes A. Ramé C. Ollion. “Leveraging Weakly Annotated Data for Fashion Image Retrieval and Label Prediction”. In: (2017).

[37] K. Yamaguchi T. L. Berg L. E. Ortiz. “Chic or Social: Visual Popularity Analysis in Online Fashion Networks”. In: (2014).

[38] S. Zheng F. Yang M. H. Kiapour R. Piramuthu. “ModaNet: A Large-scale Street Fashion Dataset with Polygon Annotations”. In: (2018).

[39] J. Feng C. Domokos H. Xu J. Huang Z. Hu S. Yan S. Liu. “Fashion Parsing with Weak Color-category Labels”. In: (2014).

[40] T. Iwata S. Watanabe H. Sawada. “Fashion Coordinates Recommender System using Photographs from Fashion Magazines”. In: (2011).

[41] W. Wang Z. Zhang S. Qi J. Shen Y. Pang L. Shao. “Learning Compositional Neural Information Fusion for Human Parsing”. In: (2020).

[42] S. C. Hidayati K.-L. Hua W.-H. Cheng S.-W. Sun. “What are the Fashion Trends in New York?” In: (2014).

[43] Aditya Ramesh Mikhail Pavlov Gabriel Goh Scott Gray Chelsea Vos Alec Radford Mark Chen Ilya Sutskevern. “Zero-Shot Text-to-Image Generation”. In: (2021).

[44] Z. Liu S. Yan P. Luo X. Wang X. Tang. “Fashion Landmark Detection in the Wild”. In: (2016).

[45] K. Yamaguchi T. Okatani K. Sudo K. Murasaki Y. Taniguchi. “Mix and Match: Joint Model for Clothing and Attribute Recognition”. In: (2015).

[46] K. E. Ak A. A. Kassim J. H. Lim J. Y. Tham. “Learning Attribute Representations with Localization for Flexible Fashion Search”. In: (2018).

[47] Y. Ma J. Jia S. Zhou J. Fu Y. Liu Z. Tong. “Towards Better Understanding the Clothing Fashion Styles: A Multimodal Deep Learning Approach”. In: (2017).

[48] S.Yan Z.Liu P.Luo S.Qiu X.Wang X.Tang. “Unconstrained Fashion Landmark Detection via Hierarchical Recurrent Transformer Networks”. In: (2017).

[49] S. Vittayakorn T. Umeda K. Murasaki K. Sudo T. Okatani K. Yamaguchi. “Automatic Attribute Discovery with Neural Activations”. In: (2016).

[50] J. Dong Q. Chen W. Xia Z. Huang S. Yan. “A Deformable Mixture Parsing Model with Parselets”. In: (2013).

[51] J. Huang R. S. Feris Q. Chen S. Yan. “Cross-domain Image Retrieval with a Dual Attribute-aware Ranking Network”. In: (2015).

[52] L. Liu J. Xing S. Liu H. Xu X. Zhou S. Yan. “Wow! You are so Beautiful Today!g”. In: (2014).

[53] S. Liu Z. Song G. Liu C. Xu H. Lu S. Yan. “Street-to-shop: Cross-scenario Clothing Retrieval via Parts Alignment and Auxiliary Set”. In: (2012).

[54] X. Liang S. Liu X. Shen J. Yang L. Liu J. Dong L. Lin S. Yan. “Deep Human Parsing with Active Template Regression”. In: (2015).

[55] S. Qiu X. Wang X. Tang Z. Liu P. Luo. “DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations”. In: (2016).

[56] Z. Kuang Y. Gao G. Li P. Luo Y. Chen L. Lin W. Q. Zhang. “Fashion Retrieval via Graph Reasoning Networks on a Similarity Pyramid”. In: (2019).

[57] W. Wang Y. Xu J. Shen S. Zhu. “Attentive Fashion Grammar Network for Fashion Landmark Detection and Clothing Category Classification”. In: (2018).