The multi-modal universe of fast-fashion: the Visuelle 2.0 benchmark

Geri Skenderi[1] Christian Joppi[2] Matteo Denitto[2] Berniero Scarpa[3] Marco Cristani[1,2]
[1] Università degli Studi di Verona [2] Humatics s.r.l. [3] Nuna Lie s.r.l.

Abstract

We present Visuelle 2.0, the first dataset useful for facing diverse prediction problems that a fast-fashion company has to manage routinely. Furthermore, we demonstrate how the use of computer vision is substantial in this scenario. Visuelle 2.0 contains data for 6 seasons / 5355 clothing products of Nuna Lie[1]*, a famous Italian company with hundreds of shops located in different areas within the country. In particular, we focus on a specific prediction problem, namely* short-observation new product sale forecasting *(SO-fore). SO-fore assumes that the season has started and a set of new products is on the shelves of the different stores. The goal is to forecast the sales for a particular horizon, given a short, available past (few weeks), since no earlier statistics are available. To be successful, SO-fore approaches should capture this short past and exploit other modalities or exogenous data. To these aims, Visuelle 2.0 is equipped with disaggregated data at the item-shop level and multi-modal information for each clothing item, allowing computer vision approaches to come into play. The main message that we deliver is that the use of image data with deep networks boosts performances obtained when using the time series in long-term forecasting scenarios, ameliorating the WAPE by 8.2% and the MAE by 7.7%. The dataset is available at :* https://humaticslab.github.io/forecasting/visuelle.

Figure 1. **Why are images important for fashion forecasting?**. A visual excerpt of Visuelle 2.0, organized per category, which shows the crucial role of images: in the *patterned shirt category* two products have exactly the same textual attributes (floreal), so the image becomes necessary for discriminative tasks. Similar considerations hold for *short sleeves* and *doll dress*.

1. Introduction

Fashion forecasting has traditionally been studied in scientific sectors other than computer vision, such as operational research and logistics, with the primary aim of predicting trends [12, 16], sales [15, 17, 18], and performing demand forecasting [9, 23]. In recent years this trend has faded, showing an increasing cross-fertilization with computer vision [1, 5, 7, 19, 21].

In this paper, we present Visuelle 2.0, which contains real data for 5355 clothing products of a retail fast-fashion Italian company, *Nuna Lie*. For the first time ever, a retail fast-fashion company has decided to share part of its data to provide a genuine benchmark for research and innovation purposes. Specifically, Visuelle 2.0 provides data from 6 fashion seasons (partitioned in Autumn-Winter and Spring-Summer) from 2017-2019, right before the Covid-19 pandemic[2]. Each product in our dataset is accompanied by an HD image, textual tags and more. The time series data are disaggregated at the shop level, and include the sales, inventory stock, max-normalized prices[3] and discounts. This

[1] http://www.nunalie.it

[2] The pandemic represented an unicum in the dynamics of the fast fashion companies, so it has not been included. The market effectively restarted in AW 21-22 which is an ongoing season at the time of writing.

[3] Prices have been normalized for the sake of confidentiality.

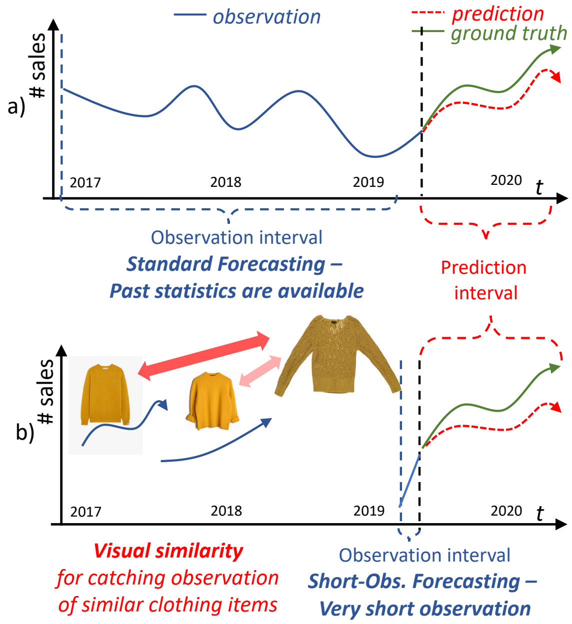
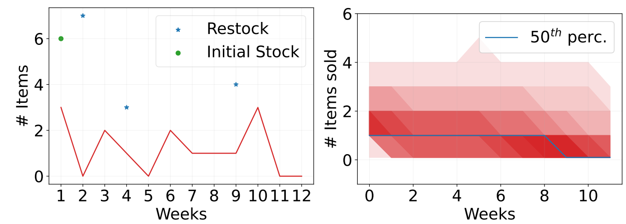


Figure 2. Short-observation new product sales forecasting (SO-fore): in a), the standard forecasting setup is reported. In b), SO-fore is sketched, focusing on a very short observation window (2 weeks) to predict sales. Here we show that relying on visual similarity to extract statistics of similar data improves forecasting.



(a) (b)

Figure 3. Time series in Visuelle 2.0; a) an example of a sale signal, with the restocks highlighted by the asterisks; b) the [0;95$^{\text{th}}$] percentile statistics of the sales signals related to the season SS19, considering all the stores and all the products.

permits to perform SO-fore considering each single store. Exogenous time series data is also provided, in the form of Google Trends based on the textual tags and multivariate weather conditions of the stores' locations. Finally, we also provide purchase data for 667K customers whose identity has been anonymized, to capture personal preferences. With these data, Visuelle 2.0 allows to cope with several problems which characterize the activity of a fast fashion company: *new product demand forecasting* [9, 23], *short-observation new product sales forecasting* [15, 17, 18] and *product recommendation* [4].

In this paper, we focus on one of these problems: short-observation new product forecasting (SO-fore). SO-fore aims at predicting the future sales in the short term, having a past statistic given by the early sales of a given product (Fig. 2b provides a visual comparison with standard forecasting in Fig. 2a). In practice, after a few weeks from the delivery on market, one has to sense how well a clothing item has performed and forecast its behavior in the coming weeks. This is crucial to improve *restocking policies* [20]: a clothing item with a rapid selling rate should be restocked to avoid stockouts. Two particular cases of the SO-fore problem will be taken into account: SO-fore$_{2-10}$, in which the observed window is 2 weeks long and the forecasting horizon is 10 weeks long, required when a company wants to implement few restocks [10]; SO-fore$_{2-1}$, where the forecasting horizon changes to a single week, and is instead required when a company wants to take decisions on a weekly basis, as in the *ultra-fast fashion* supply chain [6, 27].

Our findings show that the usage of image data is crucial in absence of long term statistics which characterize the sales, because the pictorial content of a fast fashion product can be used to inherit long term statistics *from visual similarity* (see Fig. 2b): the images allow to refer with high precision to past data that are akin to the product of analysis, providing informative priors.

Visuelle 2.0 is a substantial extension of the unpublished Visuelle dataset [25], used only for *new product demand forecasting*, where less data (no weather conditions, no customer data) was furnished, aggregated per-product over all the retail stores (no geographical/store dimension).

2. The Dataset

Visuelle 2.0 describes the sales between Nov. 2016 and Dec. 2019 of 5355 different products across 110 different shops. For each product, multi-modal information is available, as described in the following.

Time series data. Given a product i of size s at a retail store r, we refer to its *product sale* signal as $S(i,s,r,t)$ where t refers to the t-th week of market delivery, with $i = 1, ..., N$, $s = 1, ..., M$, $r = 1, ..., L$ and $t = 1, ..., K$. We also define the *inventory position* signal $I(i,s,r,t)$ indicating the inventory on-hand for that quadruplet (i,s,r,t). Combining these data, we can individuate all those *legit* sales signals that do not involve a stockout until the K-th week. Formally, a legit sale signal is $S(i,s,r,1), ..., S(i,s,r,t_{legit})$ where $t_{legit}+1$ indicates the first week with a stockout $I(i,s,r,t_{safe}+1) = 0$, and $t_{safe} > K$. Hence, we guarantee that a zero-sale legit signal (i.e. $S(i,s,r,t) = 0$) is provided *only* when nobody bought i (even tough it was available at the shop) and not because of a stockout. These signals are important because they focus on the net performance of a product, independently on the inventory management. To make the signals denser, we aggregate the different sizes obtaining the final $sale(i,r,t) = \sum_{s=1}^{M} A(i,s,r,t)$. Additionally, we include the *Restock flag* signal $R(i,s,r,t)$ indicating when a restocking has

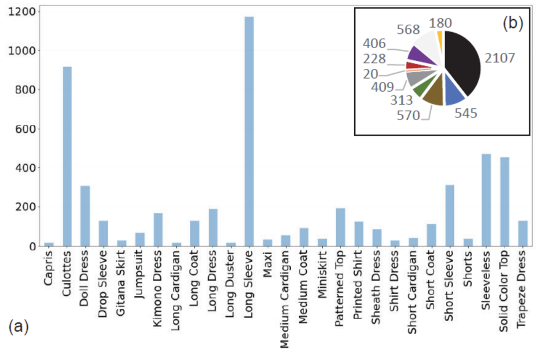


Figure 4. Cardinality of products in Visuelle 2.0 by (a) categories, (b) colors

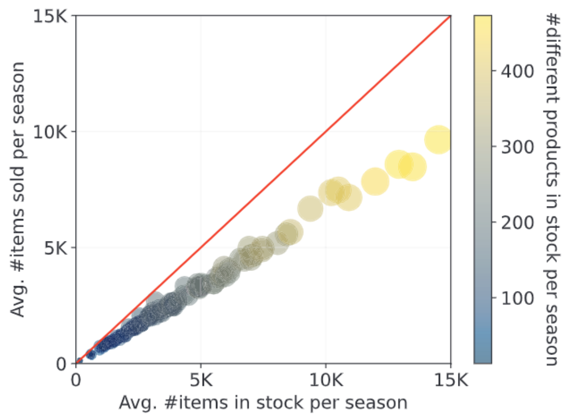


Figure 5. Retail stores sales statistics; each blob indicates the number of items received in their inventory (x-axis) VS number of items sold. The blob size indicates the number of *different* products in the inventory. The red line represents an ideal performance (sales=inventory). The figure says small shops tend to have better performances.

been carried out. Fig. 3a reports one example of a legit sale signal. The initial stock is represented by the initial amount of items available in the inventory. Fig. 3b depicts a log-density plot of the legit sales of all the products, averaged over categories and retail stores during the SS19 season. Fig. 5 shows sales statistics for the 110 available shops.

Image data Each product is associated with an RGB image that has a resolution which varies from 256x256 to 1193x1172, with a median size of 575x722 (WxH). Each image portrays the clothing item on a white background, with no person wearing it. Some examples of these images are provided in Fig. 1.

Text data Multiple textual tags related to each product's visual attributes are available. These tags have been extracted with diverse procedures or chosen by hand, carefully validated by the Nuna Lie team. The first tag is the *category*, taken from a vocabulary of 27 elements, visualized in Fig. 4a; the cardinality of the products shows large variability among categories overall, due to the fact that some categories (e.g. long sleeves) cost less and ensure higher earnings. The *color* tag (Fig. 4b) represents the most dominant color chosen among 10 manually detected colors. The *fabric* tag comes directly from the technical sheet of the products, chosen from a vocabulary of 59 elements. Finally, the release date for each product-shop pair is recorded as a textual string.

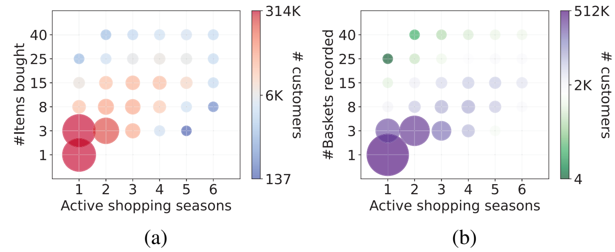


Figure 6. **Customer shopping history statistics**: a) the blob size indicates how many customers have a shopping history of n seasons long (even non consecutive seasons, at least one item bought per "active" season) on the x-axis, consisting of m items in total on the y-axis; b) Number of baskets associated to customers over shopping seasons.

Customer purchase data Visuelle 2.0 contains anonymized data for 667086 customers, who have requested a fidelity card thanks to which it is possible to extract the history of their purchases and the baskets of products they bought. These data consists of: ID of the purchased product, date-time of purchase, retail store ID and quantity. Fig. 6 gives a glimpse on the distribution of these data, where it is possible to note that there are around 6k users which have bought continuously a total of 25 products over 4 seasons (Fig. 6a). More than 2K users have bought continuously 15 baskets of products over 4 seasons. Customer data are useful to test recommendation approaches, whose goal is to recommend available products that the user will eventually buy, possibly exploiting image data to capture personal aesthetic preferences. In this paper, we do not tackle this problem.

Exogenous time series *Google Trends* time series for each product are provided, based on the product's three associated attributes: $color, category, fabric$. The trends are downloaded starting from 52 weeks before the product's release date, essentially providing a popularity curve for each of the attributes. Google Trends' efficacy for predictions on fast fashion problems has been demonstrated before in [24, 25]. *Weather reports* downloaded from IlMeteo[4] are also supplied, containing the real weather conditions on a daily basis at the municipality level. The efficacy of weather reports for forecasting in fast fashion has been demonstrated in [3, 26].

[4] https://www.ilmeteo.it/portale/archivio-meteo/

3. Experiments

Here we show how Visuelle 2.0 is a genuine benchmark for two types of SO-fore: i) SO-fore$_{2-10}$ and ii) SO-fore$_{2-1}$.

SO-fore$_{2-10}$ allows the company to customize the restocking operations for each product on the basis of the early sales, minimizing the number of such operations. Firstly, the sales series are split into an observation window and a horizon window (i.e. the known past and the future to forecast), set here to 2 and 10 weeks respectively, covering the 12-week fast fashion life-cycle [29]. Formally, the goal is to perform a multi-step forecast of the sales of an item i in a store r $(sale(i, r, 3), ..., sale(i, r, 12))$, given the first two time-steps. Two weeks is a standard period to sufficiently understand the whereabouts of the fashion market and take decisions for the future [28, 30].

SO-fore$_{2-1}$ serves in other contexts where a weekly restocking schedule is adopted [6, 27]. The idea is to estimate the time-point $sale(i, r, t)$ given the previous two time-steps $sale(i, r, t-1)$, $sale(i, r, t-2)$. Similarly to before, we set the initial observation window to 2, but use a sliding window approach to perform an autoregressive forecast.

In both cases we split the data into train and test sets, where the test set contains the 10% most recent item-shop pairs, such that the items that are seen in training will have always been released before the ones we test on. We use the following three approaches as baselines:

- Classical forecasting methods, namely the Naive method [13] (using the last observed ground truth value as the forecast) and Simple Exponential Smoothing (SES) [13, 14];
- kNN [8, 22], which produces forecasts based on product similarity. This is done by finding k-Nearest Neighbors from the past that are similar to the input product and performing a weighted average of their sales. We set $k = 11$ and we compute the similarity between products using the known time series or image features extracted by a pretrained ResNet [11];
- An autoregressive, attention-based RNN architecture [8], where the different data modalities are first processed separately and then merged together through several additive attention modules [2].

Results are displayed in Table 1 and Table 2, where all the listed approaches only use sales time series data as input, unless specified otherwise. Classical forecasting approaches tend to give poor performances due to the small number of observations [13]. The kNN-based methods show an improvement over the statistical forecasting baselines, demonstrating that inter-product similarity is important when predicting future sales. This is tied with the notion that new products will sell comparably to older, similar products. Trivially utilizing the images with kNN lowers the performances, but their contribution is highlighted when utilising an expressive neural network architecture. Cross-Attention RNN [8] outperforms the others by a noticeable margin, because the model is able to learn non-linear, inter-product dependencies throughout the whole training set and also advanced temporal dynamics. It is worth noting that we reach the best performances in SO-fore$_{2-10}$ by pairing each product's time series input with its respective image. This shows that visual representations allow the model to better understand long term forecasting patterns. Another important takeaway is that the results for time series only methods are much better on SO-fore$_{2-1}$, due to the localized temporal information and shorter forecasting horizon, while the images become less important. Additionally, we tested Cross-Attention RNN in the *Demand forecasting* task, i.e., predicting the full sales series without having access to any previous observations, but only to the product image. Obviously, results are definitely inferior to the SO-fore variants, but comparable to the kNN approaches.

Method	WAPE	MAE
***Demand* CrossAttnRNN w/ image**	**83.33**	**0.97**
Naive	118.176	1.31
SES	111.265	1.23
kNN	91.13	0.98
kNN + image	97.97	1.06
CrossAttnRNN	35.13	0.39
CrossAttnRNN w/ image	**32.25**	**0.36**

Table 1. Results for SO-fore$_{2-10}$, showing the Weighted Average Percentage Error (WAPE) and Mean Absolute Error (MAE) [25] for the different baselines. The lower the better for both metrics. In the first row we also report results for the demand forecasting of new products without past sales, demonstrating how much the knowledge of the initial sale dynamics improves forecasting.

Method	WAPE	MAE
Naive	101.922	1.13
SES	97.85	1.08
kNN	87.11	0.94
kNN + image	88.97	0.96
CrossAttnRNN	**23.20**	**0.26**
CrossAttnRNN w/ images	23.70	**0.26**

Table 2. Results for SO-fore$_{2-1}$, showing the Weighted Average Percentage Error (WAPE) and Mean Absolute Error (MAE) [25] for the different baselines. The lower the better for both metrics.

Our dataset and experiments provide a general overview of how problems in the fashion realm can be tackled and how the use of computer vision and multi-modal approaches is key to providing better solutions. The dataset page will contain further information regarding other, possible challenges and tasks for Visuelle 2.0.

Acknowledgements This work was partially supported by the Italian MIUR through PRIN 2017 - Project Grant 20172BH297: "I-MALL - improving the customer experience in stores by intelligent computer vision" and the MIUR project "Dipartimenti di Eccellenza 2018-2022".

References

[1] Ziad Al-Halah, Rainer Stiefelhagen, and Kristen Grauman. Fashion forward: Forecasting visual style in fashion. In *Proceedings of the IEEE international conference on computer vision*, pages 388–397, 2017. 1

[2] Dzmitry Bahdanau, Kyung Hyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. Jan. 2015. 3rd International Conference on Learning Representations, ICLR 2015 ; Conference date: 07-05-2015 Through 09-05-2015. 4

[3] Samaneh Beheshti-Kashi, Hamid Reza Karimi, Klaus-Dieter Thoben, Michael Lütjen, and Michael Teucke. A survey on retail sales forecasting and prediction in fashion markets. *Systems Science & Control Engineering*, 3(1):154–161, 2015. 3

[4] Samit Chakraborty, Md Hoque, Naimur Rahman Jeem, Manik Chandra Biswas, Deepayan Bardhan, Edgar Lobaton, et al. Fashion recommendation systems, models and methods: A review. In *Informatics*, volume 8, page 49. Multidisciplinary Digital Publishing Institute, 2021. 2

[5] Audrey Aurelia Chang, Jihad Fahri Ramadhan, Zainina Kyla Shafira Adnan, Bayu Kanigoro, Edy Irwansyah, et al. Fashion trend forecasting using machine learning techniques: A review. In *Proceedings of the Computational Methods in Systems and Software*, pages 34–44. Springer, 2021. 1

[6] Tsan-Ming Choi, Chi-Leung Hui, Na Liu, Sau-Fun Ng, and Yong Yu. Fast fashion sales forecasting with limited data and time. *Decision Support Systems*, 59:84–92, 2014. 2, 4

[7] Yujuan Ding, Yunshan Ma, Lizi Liao, Waikeung Wong, and Tat-Seng Chua. Leveraging multiple relations for fashion trend forecasting based on social media. *IEEE Transactions on Multimedia*, 2021. 1

[8] Vijay Ekambaram, Kushagra Manglik, Sumanta Mukherjee, Surya Shravan Kumar Sajja, Satyam Dwivedi, and Vikas Raykar. Attention based multi-modal new product sales time-series forecasting. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 3110–3118, 2020. 4

[9] Javad Feizabadi. Machine learning demand forecasting and supply chain performance. *International Journal of Logistics Research and Applications*, 25(2):119–142, 2022. 1, 2

[10] Marshall Fisher, Kumar Rajaram, and Ananth Raman. Optimizing inventory replenishment of retail fashion products. *Manufacturing & service operations management*, 3(3):230–241, 2001. 2

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 4

[12] Gwyneth Holland and Rae Jones. *Fashion trend forecasting*. Laurence King, 2017. 1

[13] Robin John Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, Australia, 2nd edition, 2018. 4

[14] Rob Hyndman, Anne B Koehler, J Keith Ord, and Ralph D Snyder. *Forecasting with exponential smoothing: the state space approach*. Springer Science & Business Media, 2008. 4

[15] Majd Kharfan, Vicky Wing Kei Chan, and Tugba Firdolas Efendigil. A data-driven forecasting approach for newly launched seasonal products by leveraging machine-learning approaches. *Annals of Operations Research*, 303(1):159–174, 2021. 1, 2

[16] Eundeok Kim, Ann Marie Fiore, Alice Payne, and Hyejeong Kim. *Fashion trends: Analysis and forecasting*. Bloomsbury Publishing, 2021. 1

[17] Ana LD Loureiro, Vera L Miguéis, and Lucas FM da Silva. Exploring the use of deep neural networks for sales forecasting in fashion retail. *Decision Support Systems*, 114:81–93, 2018. 1, 2

[18] Shaohui Ma and Robert Fildes. Retail sales forecasting with meta-learning. *European Journal of Operational Research*, 288(1):111–128, 2021. 1, 2

[19] Yunshan Ma, Yujuan Ding, Xun Yang, Lizi Liao, Wai Keung Wong, and Tat-Seng Chua. Knowledge enhanced neural fashion trend forecasting. In *Proceedings of the 2020 International Conference on Multimedia Retrieval*, pages 82–90, 2020. 1

[20] Dennis Maaß, Marco Spruit, and Peter de Waal. Improving short-term demand forecasting for short-lifecycle consumer products with data mining techniques. *Decision analytics*, 1(1):1–17, 2014. 2

[21] Utkarsh Mall, Kevin Matzen, Bharath Hariharan, Noah Snavely, and Kavita Bala. Geostyle: Discovering fashion trends and events. In *ICCV*, 2019. 1

[22] Francisco Martínez, M.P. Frías, María Pérez, and Antonio Rivera Rivas. A methodology for applying k-nearest neighbor to time series forecasting. *Artificial Intelligence Review*, 52, 10 2019. 4

[23] Shravan Sajja, Nupur Aggarwal, Sumanta Mukherjee, Kushagra Manglik, Satyam Dwivedi, and Vikas Raykar. Explainable ai based interventions for pre-season decision making in fashion retail. In *8th ACM IKDD CODS and 26th COMAD*, pages 281–289. 2021. 1, 2

[24] Emmanuel Sirimal Silva, Hossein Hassani, Dag Øivind Madsen, and Liz Gee. Googling fashion: forecasting fashion consumer behaviour using google trends. *Social Sciences*, 8(4):111, 2019. 3

[25] Geri Skenderi, Christian Joppi, Matteo Denitto, and Marco Cristani. Well googled is half done: Multimodal forecasting of new fashion product sales with image-based google trends. *arXiv preprint arXiv:2109.09824*, 2021. 2, 3, 4

[26] Donald Sull and Stefano Turconi. Fast fashion lessons. *Business Strategy Review*, 19(2):4–11, 2008. 3

[27] Ian Malcolm Taplin. Global commodity chains and fast fashion: How the apparel industry continues to re-invent itself. *Competition & Change*, 18(3):246–264, 2014. 2, 4

[28] Sébastien Thomassey. Sales forecasting in apparel and fashion industry: A review. *Intelligent fashion forecasting systems: Models and applications*, pages 9–27, 2014. 4

[29] S Thomassey et al. Intelligent demand forecasting systems for fast fashion. In *Information systems for the fashion and apparel industry*, pages 145–161. Elsevier, 2016. 4

[30] Yong Yu, Tsan-Ming Choi, Kin-Fan Au, and Chui-Yan Kwan. A web-based system for fashion sales forecasting. *Research Journal of Textile and Apparel*, 2008. 4