A PICTURE IS WORTH MORE THAN A THOUSAND PURCHASES: DESIGNING AN IMAGE-BASED FASHION CURATION SYSTEM

Research in Progress

Griebel, Matthias, University of Würzburg, Würzburg, Germany, matthias.griebel@uni-wuerzburg.de

Welsch, Giacomo, University of Würzburg, Würzburg, Germany, giacomo.welsch@uni-wuerzburg.de

Greif, Toni, University of Würzburg, Würzburg, Germany, toni.greif@uni-wuerzburg.de

Flath, Christoph, University of Würzburg, Würzburg, Germany, christoph.flath@uni-wuerzburg.de

Abstract

Online retailing has been experiencing explosive growth for years and is dramatically reshaping the way people shop. Given the lack of personal interactions fashion retailers have to establish compelling service and information offerings to sustain this growth trajectory. A recent manifestation of this is the emergence of shopping curation as a service. For this purpose, experts manually craft individual outfits based on customer information from questionnaires. For the retailers as well as for the customers, this process entails severe weaknesses, particularly with regard to immediateness, scalability, and perceived financial risks. To overcome these limitations, we present an artificial fashion curation system for individual outfit recommendations that leverages deep learning techniques and unstructured data from social media and fashion blogs. Here, we lay out the artifact design and provide a comprehensive evaluation strategy to assess the system's utility.

Keywords: Curated Retailing, Fashion, Image Recognition, Deep Learning.

1 Introduction

E-commerce has dramatically changed the retailing landscape and online retail continues exhibiting explosive growth in recent years (Doherty and Ellis-Chadwick, 2010). This applies in particular to the fashion industry, where online sales are currently growing at an annual rate of ten percent (Amed et al., 2017). Vis-a-vis stationary retail, online fashion retail must overcome deficits with respect to product presentation as well as ancillary service offerings. Addressing these issues researchers have highlighted the importance of fostering the online shopping experience through social integration (Kim, Suh, and Lee, 2013), improved visualization (Won Jeong et al., 2009) as well as optimized search (Mathwick and Rigdon, 2004) and recommendations (Z. Lin, 2014). Amed et al. (2017) highlights the emergence of a new service with increasing importance to customers – *curation*.

Sebald and Jacob (2018) refer to such service as "[c]urated retailing [which] combines convenient online shopping with personal consultation service to provide a more personalized online experience through curated product selections, orientation and decision aids, and tailor-made solutions based on the customer's preferences". The increasing popularity of start-ups with curated retail logic (Modomoto or Outfittery) and the market entry of major players such as Zalando (Zalon) underline the potential of this service. In

addition, the next generation of customers is particularly open to new shopping models such as curated shopping (Heinemann, 2017).

Clearly, these curation offerings critically depend on human stylists evaluating customer looks and proposing suitable outfits. Ultimately, such solutions cannot properly scale in an e-commerce environment. A naïve solution to address this scalability challenge is to replace the curator with a conventional recommendation system. One can distinguish between content-based methods and collaborative-filtering approaches (Adomavicius and Tuzhilin, 2005). The former exploit similarities between item features, whereas the latter generate product suggestions based on purchase behavior of users with similar preferences or frequently bought item pairs. Consequently, these algorithms offer a very limited form of personalization compared to curated shopping as they do not understand or even consider the style of the customer. Similarly, classic recommendation engines typically cannot incorporate cues from outside sources which are particularly relevant in the fashion domain, e.g., social media and influencers (Amed et al., 2017).

Computational understanding of fashion and clothing is fundamentally a challenge of computer vision requiring extensive analysis of unstructured image data. Not long ago, execution of such tasks would have required an individually designed solution and extensive computational resources. However, the recent artificial intelligence (AI) revolution has brought forward comprehensive deep learning frameworks such as TensorFlow (Abadi, Barham, et al., 2016) and PyTorch (Paszke et al., 2017) as well as powerful cloud-based computing platforms. Together they facilitate fast experimentation and prototyping through user-friendliness, modularity, and extensibility.

Several authors propose such AI-based solutions for fashion outfit recommendation (Han, Wu, Jiang, et al., 2017; Vasileva et al., 2018; W. Wang et al., 2018) or trend forecasting (Al-Halah, Stiefelhagen, and Grauman, 2017; Matzen, Bala, and Snavely, 2017). However, these approaches do not explore the end-to-end automation of the curation process. Against this backdrop, our research is concerned with building and evaluating an AI-based curation system. To this end, we leverage deep learning techniques and follow up on the call for embracing the value of unstructured data in the design of analytical information system put forward by Müller et al. (2016).

Our research seeks to answer the following research questions:

RQ1: How to employ AI components to instantiate an automated curation system?

RQ2: To what extent can AI-based curation systems achieve the recommendation quality and acceptance of a human curator?

2 Related Work

Recently, deep learning enabled AI solutions for fashion have attracted great attention. While earlier computer vision models (H. Chen, Gallagher, and Girod, 2012; Kiapour et al., 2014; X. Wang and T. Zhang, 2011) mostly rely on handcrafted features, modern applications are typically build on top of deep convolutional neural networks (CNNs) that automatically learn features with multiple levels of abstraction (Lecun, Bengio, and Hinton, 2015). Until today, approaches already cover fields such as clothing parsing and categorization (Yamaguchi et al., 2012, 2015a; Yang, Luo, and L. Lin, 2014), clothing attributes detection (Al-Halah, Stiefelhagen, and Grauman, 2017; Kiapour et al., 2014) and object detection via fashion landmarks (Z. Liu, Luo, et al., 2016; Z. Liu, Yan, et al., 2016; W. Wang et al., 2018), bounding boxes (Hadi Kiapour et al., 2015; J. Huang et al., 2015) or semantic segmentation (Zheng et al., 2018). Moreover, researchers tackle the challenge of fashion trend forecasting using images from online shops (Al-Halah, Stiefelhagen, and Grauman, 2017) or from social media (Gabale and Subramanian, 2018; Matzen, Bala, and Snavely, 2017). Finally, many authors focus on fashion recommendations.

Existing recommendation systems either seek to identify similar or complementary fashion items. Similar items are useful for cross domain image retrieval, i.e., matching street clothing photos in online shops (Hadi Kiapour et al., 2015; J. Huang et al., 2015; Z. Liu, Luo, et al., 2016; Shankar et al., 2017). In contrast, complementary item recommendation determines the compatibility between items. Many authors measure

the pairwise compatibility between two items, e.g. the compatibility between a shirt and a pair of pants. These approaches are based on graphs (McAuley et al., 2015), Conditional Random Fields (Simo-Serra et al., 2015), Siamese networks (Tautkute et al., 2018; Veit, Kovacs, et al., 2015), Conditional Similarity Networks (Veit, Belongie, and Karaletsos, 2017), or unsupervised models (Hsiao and Grauman, 2017). However, an outfit is typically composed of more than two fashion items (e.g., a top, a pair of pants, and shoes), which renders pairwise compatibility insufficient. Vasileva et al. (2018) address this problem by jointly learning similarity and compatibility. Han, Wu, Jiang, et al. (2017) model an outfit as a series of fashion items using a bidirectional long term short term memory (LSTM) network. This approach considers the visual compatibility of an entire outfit and generates outfits and can process input either based on user's text or image or both. Nakamura and Goto (2018) extend this approach by adding a style component that is capable of learning and controlling the styles of generated outfits.

While all studies address important topics of visual fashion understanding they do not explore the end-to-end automation of the curation process. This process is commonly structured as follows (Sebald and Jacob, 2018): First, customers register online and fill in a questionnaire about their fashion preferences. Afterwards, customers chose a curator, who is responsible for outfit selection. In case of special requests, customers optionally contact their curator via phone or chat. Then, based on the information on the customer, the curator selects and triggers the shipment of the personalized outfit. Finally, customer choose which garments to keep and which to return.

A closer inspection of the curation process reveals central weaknesses which can potentially limit its growth potentials.

- *Lack of immediateness* In current curated shopping services customers may have to wait up to two weeks for delivery.

- *Lack of scalability* Curated Shopping relies heavily on human expertise embodied by curators. Having humans in crucial positions of the process significantly limits growth potentials in times of very high employment.

- *Perceived financial risks* Customers may expect curated shopping to be more expensive than regular online shopping due to the cost of curation (Cha and You, 2018).

To address these weaknesses we want to design a system that is capable of detecting relevant fashion items, incorporate expert knowledge that can adjust to current fashion trends to recommend outfits and retrieves similar items from the proposed outfit from the product catalogue. Figure 1 depicts our conceptual approach.
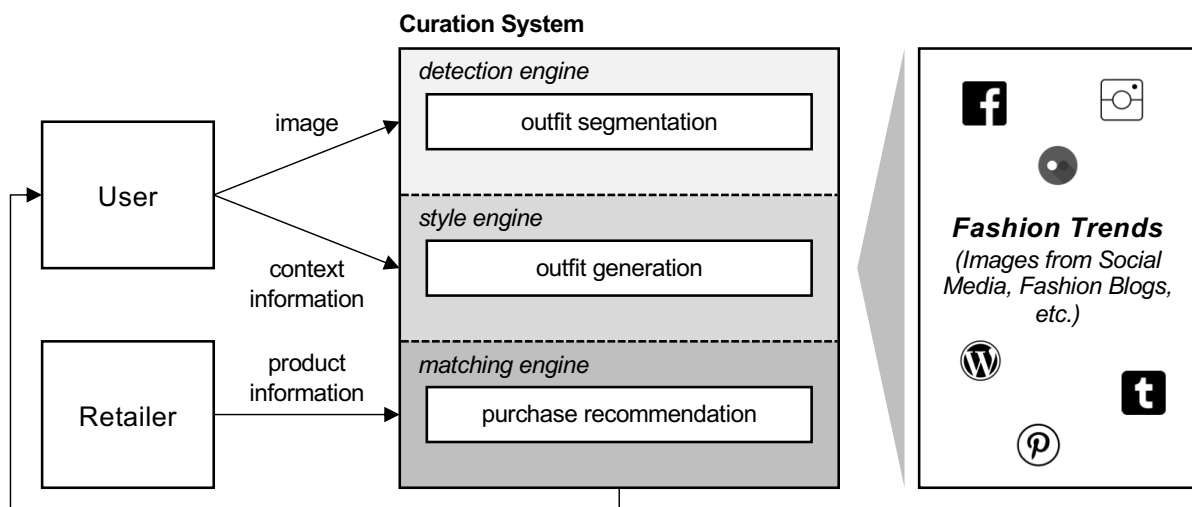

Figure 1.   Conceptual Approach

3 Methodology

As our research aims at building an artificial fashion curator, we follow the Design Science Research (DSR) paradigm which is particularly concerned with the development of useful artifacts (Baskerville et al., 2018; Hevner et al., 2004). Such artifacts can either embody (i) new solutions for known problems, (ii) known solutions extended to new problems, or (iii) new solutions for new problems (Gregor and Hevner, 2013). As we want to enrich the known domain of curated shopping with an innovative fashion curation system, we consider our artifact as a new solution for a known problem. Gregor and Hevner (2013) refer to such type of artifact as *improvement*.

3.1 Artifact overview

Figure 2 illustrates the three components of the artifact and their respective outputs (white boxes) by means of an exemplary user query. Here, the input comprises a picture of the user and a text query with contextual information (outfit style: casual, reference item in picture: pants). First, the picture passes through the detection engine that identifies four distinct fashion items: a blazer, a t-shirt, a pair of pants and a pair of high heels. Secondly, this information as well as the context information is fed into the style engine. This engine generates a casual outfit that goes with the previously detected pants based on its knowledge about (current) styles and trends. Finally, the images of the new outfit are forwarded into the matching engine that find articles in the retailers product database that are as similar as possible to the ones in the generated outfit. Subsequently, these products are recommended to the user.
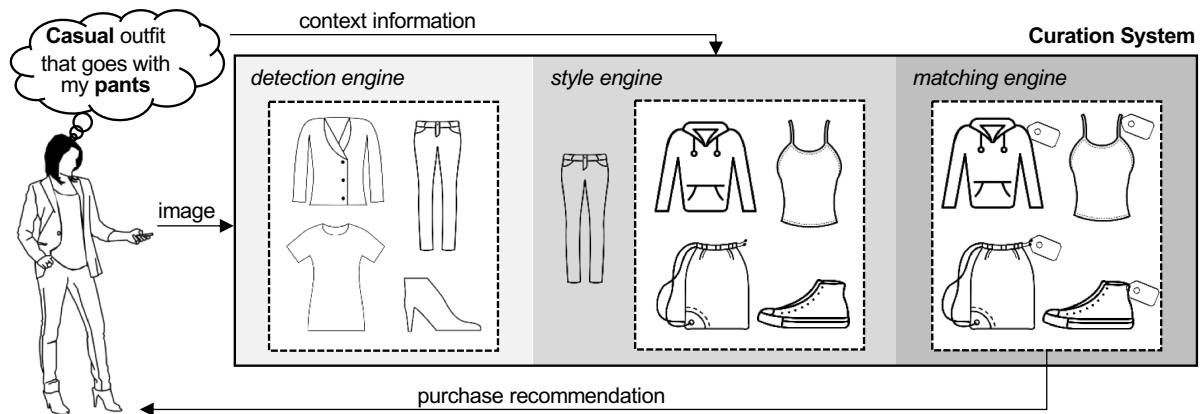

Figure 2.   Functionalities of the artifact components on an exemplary user query.

It is illustrative to describe our artifact as guidance system following the taxonomy of Morana et al. (2017) for guidance design features in information systems. The target of our curation system is to support the customer in choosing an outfit in form of a suggestive fashion recommendation. The system works in a participative mode after a user-invoked request with concurrent timing. Fashion images are the main output format of the system, however, combinations of images, animations and text are possible to improve user experience. The intention is recommending fashion items to a mostly novice audience. As the systems incorporates the knowledge fashion bloggers and influencers, it provides expert knowledge on outfit recommendations, which renders the content type as terminological. The curation system is not deliberately affecting the trust of a customer.

3.2 Design Science Process

To carry out our study and build the artifact, we follow the DSR methodology introduced by Peffers et al. (2007). Besides conceptual principles and practice rules, the methodology provides a process for executing and presenting a DSR project. Figure 3 depicts the current status of our study within the process. The following sections explore the three components in more detail.
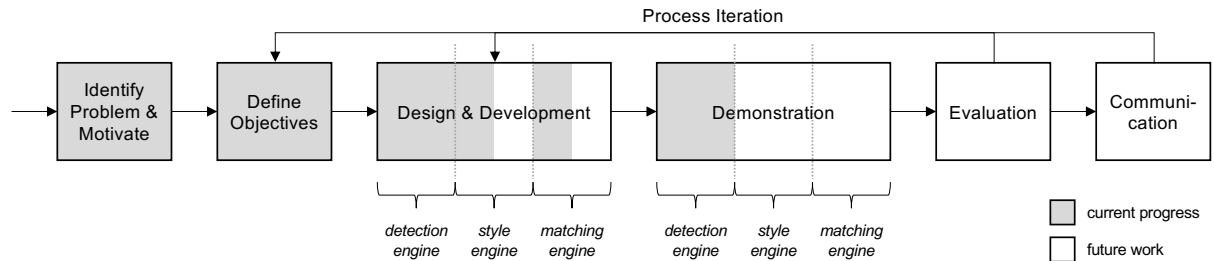
Figure 3. *Project status mapped to Peffers et al. (2007) design science research process.*

4 Detection Engine

We design the detection engine for automated and reliable identification of fashion items and their exact position within images. To this end, we leverage state-of-art image recognition techniques based on supervised deep learning. Supervised learning requires a dataset of labeled images (Lecun, Bengio, and Hinton, 2015), e.g., fashion images with attributes considering shape, color, or pattern. There are plenty of fashion datasets addressing various applications such as cross-scenario clothing retrieval, attributes recognition, clothing parsing, image retrieval and aesthetic evaluation (Zou, Wong, and Mo, 2018). To implement our detection engine, we utilize the brand-new street fashion dataset ModaNet (Zheng et al., 2018). It is built on top of the Paperdoll dataset (Yamaguchi et al., 2015b) and adds large-scale polygon-based fashion product annotations for 52,377 training images and a 2,799 validation images. These annotations render ModaNet the only publicly available dataset that enables semantic image segmentation[1]. Figure 4 depicts some training images and the corresponding segmentation masks.

Figure 4. *Training images and corresponding segmentation masks of ModaNet (Zheng et al., 2018).*

Zheng et al. (2018) benchmark several deep learning approaches for semantic image segmentation on the Modanet dataset for which DeepLabv3+ (L.-C. Chen, Zhu, et al., 2018) yields the best results. As the ModaNet dataset only consists of images showing a single person (and a single outfit respectively), the choice of semantic image segmentation is appropriate and also applicable for detecting our user

---

[1] Semantic segmentation means understanding an image at pixel level. For instance, a pixel is assigned to the class *dress* or *background*

input images. However, we expect our fashion trend data to be more diversified, e.g., containing multiple persons in one image. In order to reuse our detection module for processing the trend data, it is necessary to combine the concepts of instance detection (e.g., differentiation between two persons in an image) and semantic segmentation. Several approaches such as Mask R-CNN (K. He, Gkioxari, et al., 2017), MaskLab (L.-C. Chen, Hermans, et al., 2017) or PANet (S. Liu et al., 2018) have been proposed for this task. These instance segmentation methods detect objects in an image while simultaneously generating a high-quality segmentation mask for each instance.

Our detection engine rests upon the popular TensorFlow (Abadi, Agarwal, et al., 2016) implementation of Mask-RCNN (Abdulla, 2017), this setup provides deployment into production systems via tensorflow-serving. It is based on a Feature Pyramid Network (T. Lin et al., 2017), ResNet101 backbone (K. He, X. Zhang, et al., 2016) and uses pre-trained weights from the COCO dataset (T.-Y. Lin et al., 2014). To adopt our model for fashion purposes, we used the ModaNet dataset for further training. This enables our model to distinguish between 13 meta fashion categories (*bag, belt, boots, footwear, outer, dress, sunglasses, pants, top, shorts, skirt, headwear, scarf/tie*). We demonstrate the detection engine in the section below.

4.1 Demonstration

Figure 5 highlights the detection functionality and limitations of our detection engine by means of two example images. In the upper picture all five garments are detected and masked accurately. The person in the lower picture wears six different kinds of garments and accessories partly hidden from each other. Notwithstanding the high difficulties, the detection engine recognizes all the garments. However, the mask of the coat (class *outer*) only classifies the left part correctly and adds the bag instead of the coats right part. Such misclassification errors occasionally occur with photographs that contain multi-layered outfits or are captured under difficult light conditions. Assuming that customers use photos on which the relevant clothing is easily recognizable, we consider the detection engine to be fully functional for our purposes. In addition, we expect that further training, improved algorithms and more training data will improve the model performance in the future.

Figure 5. Demonstration of the detection engine (example images taken from zalando.com). The detected garment class names and probabilities are shown in the white box.

5 Style Engine

Our style engine creates fashionable outfits based on image and/or textual inputs. Therefore, we follow the approach of Han, Wu, Jiang, et al. (2017) and Nakamura and Goto (2018) for outfit generation and style extraction. The deep learning approach comprises multiple convolutional neural networks, a visual-semantic embedding space (VSE) (Han, Wu, P. X. Huang, et al., 2017), a bidirectional LSTM and a style extraction autoencoder (SE) R. He et al. (2017). The CNNs are used to extract features of the input images. These features are combined with contextual information (i.e., text input) within the VSE. Simultaneously, the SE module extract the style of the images. Finally, the LSTM combines all information to model an outfit as a series of fashion items.

For training, we feed complete outfits (i.e., series of single fashion items) into the model. We utlize the Polyvore dataset, which comprises 21,889 outfits from polyvore.com, a former fashion website where users could create and upload outfit data (Han, Wu, Jiang, et al., 2017). In order to adopt the model to current fashion trends, we are constantly scraping images from relevant social media platforms, fashion blogs, and magazines. These images are preprocessed by our detection engine (having a series of single fashion items as output) and put into the the style engine for further training. Figure 6 depicts the training procedure of the style engine.
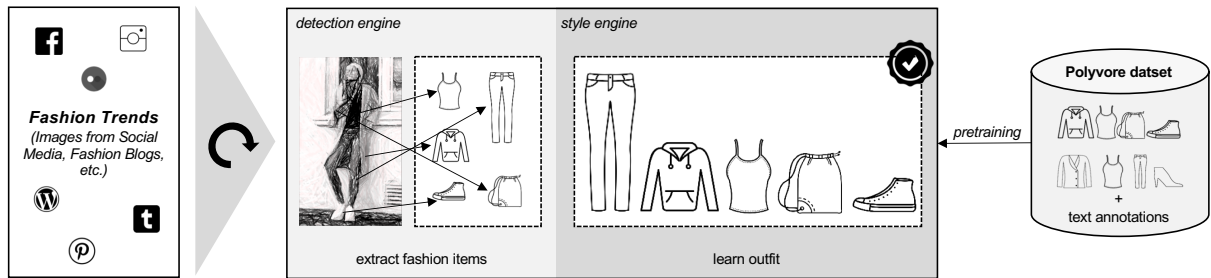

Figure 6.   Training procedure of the style engine.

During inference, one or more fashion items and/or contextual information serve as input. Subsequently, the model calculates the remaining items and returns the corresponding images. This approach also enables the evaluation of a complete outfit (Han, Wu, Jiang, et al., 2017).

As fashion photographs often contain multi-layered outfits that are partly hiding garments of the lower layers (see Figure 5), we need to carefully evaluate the style engine performance on such fashion items. A potential strategy to avoid malfunction of the style engine is to learn an additional model that replaces the partly visible item with a picture of a fully visible, similar item using generative models.

6 Matching Engine

As the user is interested in buying the generated outfit, the matching engine has to find similar products in the retailers product database. Simple similarity measures such as pixel-wise comparison often fail this task given the variety of clothes. Our matching engine needs to learn abstract high-level concepts as well as low level details.

To this end, we adapted the approach by Shankar et al. (2017) who designed a specific CNN based on the VGG-16 architecture (Simonyan and Zisserman, 2014). Using the triplet-based deep ranking paradigm (J. Wang et al., 2014), this method is capable of ranking the images concerning their similarity. In this context, triplets are sets of three images containing a street photo and two shop photos. Street photos are real-world photos of people wearing fashion items, captured in everyday uncontrolled settings, whereas shop photos are captured by professionals in more controlled settings. One of the shop photos matches the street photo (positive) while the other one is different (negative).

Similar to Shankar et al. (2017), we train the model with two types of triplets. On the one hand, there are out-of-class triplets containing negatives that differ significantly from our street photo. On the other hand, there are in-class triplets containing negative images which are very similar to the street photo. While the out-of-class triplets train the model to learn coarse differences, the in-class triplets let the model pay attention on minor differences as well.

The Exact Street2Shop dataset created by Hadi Kiapour et al. (2015) is the most popular dataset containing triplets. With only 39,479 exact matching street-to-shop pairs, this dataset is comparatively small and we need to pretrain the model. We use street-to-shop pairs (only street photo and positive) from the DeepFashion dataset (Z. Liu, Luo, et al., 2016) for this task. The missing negative photos for the triplets are sampled based on a set of basic image similarity scoring techniques (Shankar et al., 2017).

During inference, our style engine matches the street photo to the article in the retailers product database by finding nearest neighbour in the embedding space of the model.

7 Expected Contributions and Future Work

In the present study, we sketch an artificial curation system that leverages deep learning techniques and unstructured data. Addressing *RQ1*, the system design consists of three components: the detection, style and matching engine. So far, we have demonstrated the functionality of the detection component and going forward will seek to do the same for the remaining engines. To evaluate the extent to which an AI-based curation systems achieves the recommendation quality and acceptance of a human curator (*RQ2*), we envision three studies:

- We will ask a group of stylists (i.e., curators) to create several outfits based on different "input images" and context information. For comparison, our AI curation system generates outfits based on the same information. Hereafter, the stylists evaluate outfits (except for their own creations) without knowing the source.

- In the second study, we will conduct an extensive curation A/B testing. Here, one part of the customers is directed to the expert-based curation system website (A) and its traditional curation process described in Section 1. The other part of the customers is directed to the website with our AI curation system (B). This enables the comparison of key indicators such as turnover, return rate, repurchases, and more.

- The final study comprises a survey among the participants. It aims to measure satisfaction, trust, perceived usefulness, and perceived ease of use based on Xiao and Benbasat (2007).

To pursue this ambitious evaluation program, we partner with a big German fashion retailer. We will iteratively incorporate new findings from these studies in the design of our artifact as depicted in Figure 3. We plan to communicate the contributions of our research in peer-reviewed scientific publications.

For future research, we will also improve the engines with respect to current developments in AI research. We plan to test our system beyond the curation scenario. Potential use cases are style assessment for the user, recommendations in online shops, or curated design of new collections, i.e., supporting the designer for inspiration or for outfit compatibility evaluation. Furthermore, the rise of the Internet of Things (IoT) presents the opportunity to deploy our system in omni-channel environments, for instance in-store recommendations in brick and mortar stores as described by Hanke et al. (2018).

References

Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng (2016). "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems." *GPU Computing Gems Emerald Edition*, 277–291.

Abadi, M., P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al. (2016). "Tensorflow: A system for large-scale machine learning." In: *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pp. 265–283.

Abdulla, W. (2017). *Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow.* `https://github.com/matterport/Mask_RCNN`.

Adomavicius, G. and A. Tuzhilin (2005). "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions." *IEEE Transactions on Knowledge & Data Engineering* (6), 734–749.

Amed, I., A. Berg, S. Kappelmark, S. Hedrich, J. Andersson, M. Drageset, and R. Young (2017). *The State of Fashion 2018.* URL: `https://cdn.businessoffashion.com/reports/The_State_of_Fashion_2018_v2.pdf` (visited on 11/25/2018).

Baskerville, R., A. Baiyere, S. Gregor, A. Hevner, and M. Rossi (2018). "Design Science Research Contributions: Finding a Balance between Artifact and Theory." *Journal of the Association for Information Systems* 19 (5), 358–376.

Cha, H. S. and S. You (2018). "The Value and Risk of Curated Shopping: Online Consumer's Choice." In: *Proceedings of the 51st Hawaii International Conference on System Sciences*.

Chen, H., A. Gallagher, and B. Girod (2012). "Describing clothing by semantic attributes." In: *European Conference on Computer Vision*. Springer, pp. 609–623.

Chen, L.-C., A. Hermans, G. Papandreou, F. Schroff, P. Wang, and H. Adam (2017). "MaskLab: Instance Segmentation by Refining Object Detection with Semantic and Direction Features." arXiv: `1712.04837`.

Chen, L.-C., Y. Zhu, G. Papandreou, F. Schroff, and H. Adam (2018). "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation." arXiv: `1802.02611`.

Doherty, N. F. and F. Ellis-Chadwick (2010). "Internet retailing: the past, the present and the future." *International Journal of Retail & Distribution Management* 38 (11/12), 943–965.

Gabale, V. and A. P. Subramanian (2018). "How To Extract Fashion Trends From Social Media? A Robust Object Detector With Support For Unsupervised Learning." *arXiv preprint arXiv:1806.10787*.

Gregor, S. and A. R. Hevner (2013). "Positioning and Presenting Design Science Research for Maximum Impact." *MIS Quarterly* 37 (2), 337–356.

Hadi Kiapour, M., X. Han, S. Lazebnik, A. C. Berg, and T. L. Berg (2015). "Where to buy it: Matching street clothing photos in online shops." In: *Proceedings of the IEEE international conference on computer vision*, pp. 3343–3351.

Al-Halah, Z., R. Stiefelhagen, and K. Grauman (2017). "Fashion forward: Forecasting visual style in fashion." In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 388–397.

Han, X., Z. Wu, P. X. Huang, X. Zhang, M. Zhu, Y. Li, Y. Zhao, and L. S. Davis (2017). "Automatic Spatially-Aware Fashion Concept Discovery." *Proceedings of the IEEE International Conference on Computer Vision* 2017-October, 1472–1480.

Han, X., Z. Wu, Y.-G. Jiang, and L. S. Davis (2017). "Learning Fashion Compatibility with Bidirectional LSTMs." In: *Proceedings of the 2017 ACM on Multimedia Conference - MM '17*.

Hanke, J., M. Hauser, A. Dürr, and F. Thiesse (2018). "Redefining the Offline Retail Experience: Designing Product Recommendation Systems for Fashion Stores." In: *Proceeding of the European Conference on Information Systems 2018, Portsmouth, UK*.

He, K., G. Gkioxari, P. Dollar, and R. Girshick (2017). "Mask R-CNN." In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980–2988.

He, K., X. Zhang, S. Ren, and J. Sun (2016). "Deep Residual Learning for Image Recognition." *IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

He, R., W. S. Lee, H. T. Ng, and D. Dahlmeier (2017). "An Unsupervised Neural Attention Model for Aspect Extraction." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 388–397.

Heinemann, G. (2017). *Der neue Online-Handel: Geschäftsmodelle, Geschäftssysteme und Benchmarks im E-Commerce*. Springer-Verlag.

Hevner, A. R., S. T. March, J. Park, and S. Ram (2004). "Design Science in Information Systems Research." *MIS Quarterly* 28 (1), 75–105.

Hsiao, W.-L. and K. Grauman (2017). "Learning the Latent "Look": Unsupervised Discovery of a Style-Coherent Embedding from Fashion Images." In: *2017 IEEE International Conference on Computer Vision (ICCV)*. Vol. 2017-Octob. c. IEEE, pp. 4213–4222.

Huang, J., R. S. Feris, Q. Chen, and S. Yan (2015). "Cross-domain image retrieval with a dual attribute-aware ranking network." In: *Proceedings of the IEEE international conference on computer vision*, pp. 1062–1070.

Kiapour, M. H., K. Yamaguchi, A. C. Berg, and T. L. Berg (2014). "Hipster wars: Discovering elements of fashion styles." In: *European Conference on Computer Vision*. Springer, pp. 472–488.

Kim, H., K.-S. Suh, and U.-K. Lee (2013). "Effects of collaborative online shopping on shopping experience through social and relational perspectives." *Information & Management* 50 (4), 169–180.

Lecun, Y., Y. Bengio, and G. Hinton (2015). "Deep learning." *Nature* 521 (7553), 436–444.

Lin, T., P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie (2017). "Feature Pyramid Networks for Object Detection." In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 936–944.

Lin, T.-Y., M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick (2014). "Microsoft COCO: Common Objects in Context." In: *Computer Vision – ECCV 2014*. Ed. by D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars. Cham: Springer International Publishing, pp. 740–755.

Lin, Z. (2014). "An empirical investigation of user and system recommendations in e-commerce." *Decision Support Systems* 68, 111–124.

Liu, S., L. Qi, H. Qin, J. Shi, and J. Jia (2018). "Path Aggregation Network for Instance Segmentation." arXiv: 1803.01534.

Liu, Z., P. Luo, S. Qiu, X. Wang, and X. Tang (2016). "DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations." In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1. IEEE, pp. 1096–1104.

Liu, Z., S. Yan, P. Luo, X. Wang, and X. Tang (2016). "Fashion landmark detection in the wild." In: *European Conference on Computer Vision*. Springer, pp. 229–245.

Mathwick, C. and E. Rigdon (2004). "Play, flow, and the online search experience." *Journal of consumer research* 31 (2), 324–332.

Matzen, K., K. Bala, and N. Snavely (2017). "Streetstyle: Exploring world-wide clothing styles from millions of photos." *arXiv preprint arXiv:1706.01869*.

McAuley, J., C. Targett, Q. Shi, and A. van den Hengel (2015). "Image-Based Recommendations on Styles and Substitutes." In: *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '15*. SIGIR '15. New York, New York, USA: ACM Press, pp. 43–52.

Morana, S., S. Schacht, A. Scherp, and A. Maedche (2017). "A review of the nature and effects of guidance design features." *Decision Support Systems* 97, 31–42.

Müller, O., I. Junglas, J. V. Brocke, and S. Debortoli (2016). "Utilizing big data analytics for information systems research: Challenges, promises and guidelines." *European Journal of Information Systems* 25 (4), 289–302.

Nakamura, T. and R. Goto (2018). "Outfit Generation and Style Extraction via Bidirectional LSTM and Autoencoder." arXiv: 1807.03133.

Paszke, A., S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer (2017). "Automatic differentiation in PyTorch."

Peffers, K., T. Tuunanen, M. A. Rothenberger, and S. Chatterjee (2007). "A Design Science Research Methodology for Information Systems Research." *Journal of Management Information Systems* 24 (3), 45–77.

Sebald, A. K. and F. Jacob (2018). "Help welcome or not: Understanding consumer shopping motivation in curated fashion retailing." *Journal of Retailing and Consumer Services* 40, 188–203.

Shankar, D., S. Narumanchi, H. A. Ananya, P. Kompalli, and K. Chaudhury (2017). "Deep Learning based Large Scale Visual Recommendation and Search for E-Commerce." arXiv: 1703.02344.

Simonyan, K. and A. Zisserman (2014). "Very Deep Convolutional Networks for Large-Scale Image Recognition." arXiv: 1409.1556. URL: http://arxiv.org/abs/1409.1556.

Simo-Serra, E., S. Fidler, F. Moreno-Noguer, and R. Urtasun (2015). "Neuroaesthetics in fashion: Modeling the perception of fashionability." In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp. 869–877.

Tautkute, I., T. Trzcinski, A. Skorupa, L. Brocki, and K. Marasek (2018). "DeepStyle: Multimodal Search Engine for Fashion and Interior Design," 1–11. arXiv: 1801.03002.

Vasileva, M. I., B. A. Plummer, K. Dusad, S. Rajpal, R. Kumar, and D. Forsyth (2018). "Learning Type-Aware Embeddings for Fashion Compatibility." In: *ECCV*.

Veit, A., S. Belongie, and T. Karaletsos (2017). "Conditional Similarity Networks." In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Vol. 1. 2. IEEE, pp. 1781–1789.

Veit, A., B. Kovacs, S. Bell, J. McAuley, K. Bala, and S. Belongie (2015). "Learning Visual Clothing Style with Heterogeneous Dyadic Co-Occurrences." In: *2015 IEEE International Conference on Computer Vision (ICCV)*. Santiago, Chile: IEEE, pp. 4642–4650.

Wang, J., Y. Song, T. Leung, C. Rosenberg, J. Wang, J. Philbin, B. Chen, and Y. Wu (2014). "Learning fine-grained image similarity with deep ranking." In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1386–1393.

Wang, W., Y. Xu, J. Shen, and S.-C. Zhu (2018). "Attentive fashion grammar network for fashion landmark detection and clothing category classification." In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4271–4280.

Wang, X. and T. Zhang (2011). "Clothes search in consumer photos via color matching and attribute learning." In: *Proceedings of the 19th ACM international conference on Multimedia*. ACM, pp. 1353–1356.

Won Jeong, S., A. M. Fiore, L. S. Niehm, and F. O. Lorenz (2009). "The role of experiential value in online shopping: The impacts of product presentation on consumer responses towards an apparel web site." *Internet Research* 19 (1), 105–124.

Xiao, B. and I. Benbasat (2007). "E-commerce product recommendation agents: use, characteristics, and impact." *MIS quarterly* 31 (1), 137–209.

Yamaguchi, K., M. H. Kiapour, L. E. Ortiz, and T. L. Berg (2012). "Parsing clothing in fashion photographs." In: *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 3570–3577.

— (2015a). "Retrieving similar styles to parse clothing." *IEEE transactions on pattern analysis and machine intelligence* 37 (5), 1028–1040.

— (2015b). "Retrieving similar styles to parse clothing." *IEEE transactions on pattern analysis and machine intelligence* 37 (5), 1028–1040.

Yang, W., P. Luo, and L. Lin (2014). "Clothing co-parsing by joint image segmentation and labeling." In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3182–3189.

Zheng, S., F. Yang, M. H. Kiapour, and R. Piramuthu (2018). "ModaNet: A Large-Scale Street Fashion Dataset with Polygon Annotations." In: *ACM Multimedia*.

Zou, X., W. K. Wong, and D. Mo (2018). "Artificial Intelligence on Fashion and Textiles." In: *Proceedings of the Artificial Intelligence on Fashion and Textiles (AIFT) Conference 2018*, pp. 255–267.