

# OPPOSITION INFORMATION

The author of the thesis: Lukas Frösslund

The title of the thesis: Learning Style Compatibility on Fashion Data

The name of the opponent: Simon Jäger

The date of the opposition: 24/06/2021

## EVALUATION

### **Was it easy to understand the underlying purpose of the project?**

The thesis initially provides an understanding of the fashion industry's economic opportunities and sustainability challenges, helpful in understanding its purpose. The given context describes a need for systems based on reuse instead of waste, where the resale of used fashion items promotes that idea. Subsequently, a societal purpose for the thesis is apparent. Furthermore, the author provides a user-oriented understanding of the challenges to browse and filter many unique second-hand fashion items according to their preferences, as faced by the host company. Finally, the thesis describes its purpose to explore unconstrained outfit matching in fashion compatibility modeling using state-of-the-art methods and a fashion data set with both visual and categorical data to identify compatible fashion items to create stylistic collections for end-users. Given the aforementioned context, the underlying purpose of the thesis was easy to comprehend.

### **Do you consider that the report title justly reflects the contents of the report?**

The title of the thesis adequately describes the research area of the project. The experiment evaluates the baseline model and two state-of-the-art models with the unconstrained outfit matching task, considered fashion style compatibility. The data sets used for the experiment are of fashion items.

### **How did the author describe the project background?**

The author describes the theoretical background in three sections. The first section outlines and justifies convolutional neural networks with the convolution and pooling operations. Next, deep residual learning is presented with skip connections and residual blocks, building on the concept of convolutional neural networks. The second section describes the transfer learning method, where the final layers of an already trained model are fine-tuned for another domain. Finally, the third and most extensive section outlines deep metric learning as the method of extracting embedded representations of data, siamese networks that share weights and compare the output of subnetworks, triplet loss with the use of an anchor sample, negative sample, and possible sample for loss computation, and lastly, the alternative outfit ranking loss suitable for the outfit matching task.

**Was there an introduction and general survey of this area?**

The thesis explores the research area by first contrasting fashion recommendation and fashion modeling compatibility, which have attained additional interest in recent years. Next, the evolution of a pre-trained convolutional neural network by McAuley et al. to estimate pairwise compatibility using a siamese neural network to compute pairwise distances by Veit et al. is surveyed. Their ineptness for use with fashion item collections is mentioned. Then, a proposed solution for fashion modeling compatibility for fashion item collections by Han et al. is presented. The research considers fashion items as an ordered sequence and applies a bidirectional LSTM on fashion item images in addition to semantic textual information. However, the sequential requirement of the method renders it inflexible and unsuitable for the problem description of the thesis. Another study proposed by Vasileva et al. is then presented. It attempts to solve the shortcomings of the aforementioned solutions by using predefined embedding subspaces for each pair of fashion item types. The need for predefined subspaces was resolved in research by Tan et al. with the introduction of the Similarity Condition Embedding Network (SCE-Net), which learns the subspaces. Finally, the further improvement on the SCE-Net by Lin et al. with the Category-based Subspace Attention Network (CSA-Net) is presented. The solution introduces the use of category vectors along with the fashion item image for the model.

Conclusively, the thesis presents a survey of the area that explores incremental improvements leading up to recent and state-of-the-art research. Furthermore, it provides the motivations for the improvements of the mentioned work along with a summary.

**To what degree did the author justify his/her choice of method of tackling the problem?**

The author provides a high-level justification for the method anchored in state-of-the-art methods, standards, and related work. The state-of-the-art models are surveyed through the literature study, and the necessity of an industrial fashion item data set in the project to evaluate the models beyond the standard Polyvore Outfits data set is justified. Moreover, the model selection (SCE-Net and CSA-Net) with the baseline model and evaluation methods are justified. The selection of state-of-the-art models and evaluation methods are clearly connected to the fashion research area.

Many choices are grounded in the related research, the thesis may benefit from additional justification for how choices are relevant for the project. Moreover, the thesis could mention choices that were not included in the project and why. To convince the author that the research area was holistically surveyed.

**Did the author discuss the extent to which the prerequisites for the application of such a method are fulfilled?**

The author discusses the need for an industrial fashion item data set in addition to the Polyvore Outfits data set to adequately evaluate the reimplemented models. To fulfil the requirement, the provided raw product data from the host company is presented. The raw product data requires preprocessing, but is sufficient to craft the industrial fashion item data set to conduct the

experiments. Additionally, the author discusses the necessity and absence at the host company of an established baseline model for the evaluation of the state-of-the-art models. A baseline model is implemented as part of the experiment.

### **Is the method adequately described?**

The author describes the method in multiple sections. First, the details of the required data, collection, and preprocessing are given. The addition of Figure 4.2 is helpful to explain the sequence of the data preprocessing steps. After which, the triplet data, style ranking data, and evaluation data are outlined. Then, the justification and implementation of the baseline, SCE-Net, and CSA-Net models are outlined. Followed by the quantitative and qualitative evaluation methods; Fill-in-the-blank (FITB), item style retrieval with *recall @ top k*, and t-SNE visualization. Finally, the method provides the necessary description of the software implementation details.

While the thesis adequately describes the method on a high level, some choices would benefit from better justification. For example, in the data processing, the normalization step, the data split ratios, the definition of poor image quality, are presented but not justified. As a reader, these are intriguing choices to learn about. See annotated report for details.

### **Did the author present his/her results clearly and concisely?**

The author presents the results of the thesis concisely. In addition, the relevant information is provided, and tables are easy to comprehend with accompanying descriptions in-text. The descriptions outline patterns or unusual details of the results. The use of figures further enhances the understanding of the experiments, an excellent addition. Only section 5.3 t-SNE has some unclarity, where the figures are too small to observe and with the absence of accompanying figure descriptions. Moreover, some relevant figures are not referenced in in-text.

### **Do you consider the author's conclusions to be credible?**

The author based the conclusions of the thesis on the findings of the results. At the same time, the author carefully offsets the conclusions with remarks from the discussion. The use of the unseen fashion item data set makes the generalizability of the results convincing. Using a baseline model and two state-of-the-art models adds to the credibility of the suggested increased difficulty for fashion style compatibility modeling when the number of constraints is decreased. Likewise, it provides credibility to the concluded increase in accuracy with higher dimensionality of the embedded representations.

### **What is your opinion of the bibliography? What types of literature are included? Do you feel they are relevant?**

The bibliography of the thesis includes both relevant and recent references. Many of the references are published books and conference papers presented at various IEEE conferences. Most references include pertinent details, while a few could be improved. For example, the book Fashion

Recommender Systems by Nima Dokoochaki has many additional attributes available that should be included. The KTH Library helps obtain additional attributes: [Fashion recommender systems](#). All of the references provided in the bibliography are used and relevant in the report.

### Which sections of the report were difficult to understand?

The below-mentioned sections of the report indicate where a certain unclarity resides. It does not indicate that the entire section is difficult to understand unless explicitly said, but there is an area of improvement within the section.

1. **Section 1.3 Purpose & Goals:** The aim of evaluating the robustness of the state-of-the-art models for the outfit matching task is mentioned twice. It is unclear what exactly robustness means in this context. The thesis may be improved if robustness is defined or a different term is used.
2. **Chapter 2 Background:** In the thesis, the intended reader is not clear. A short introduction to machine learning and deep learning may provide the computer scientist with a more solid context to understand the project. Otherwise, a more advanced intended reader is useful to declare. Concepts such as a tensor, one-hot encoding, embedded representations, vanishing and exploding gradients, should be explained. See annotated report for details.
3. **Section 2.1 Convolutional Neural Network:** A suggested, but minor addition to the section would be a mathematical expression of the convolution and pooling operators.
4. **Figure 2.3:** The mathematical symbols should be explained. The flow of the arrows are understandable from the text, although the meaning of the symbols is not. For example,  $G_W(\dots)$  and  $E_W$ .
5. **Chapter 3 Method:** As for the background chapter, some concepts are not defined or clear. A computer scientist with a limited understanding of machine learning may need to look these up whereas they should be provided in the background. See annotated report for details.
6. **Section 4.2 Models:** It is not clear how the initial hyperparameters were chosen for the models. Are they default values from the implementation?
7. **Section 4.3.2 Style Retrieval:** Consider formally explaining *recall @ top k* to improve the clarity of the measure.
8. **Section 5.1.2 Sellpy Data:** What were the 52 of 300 collections and 365 fine-grained item categories (style ranking data)? Is this confidential, or can it be provided, perhaps in an appendix?
9. **Section 6.1 Summary of Results & Findings:** The non-linear phenomenon of the highest accuracy for the CSA-Net with ten subspaces and the lowest with seven subspaces is intriguing. Consider adding some discussion as to why this may be the case. Moreover, the SCE-Net had the best accuracy for the lowest number of subspaces, whereas the CSA-Net had the best accuracy for the highest number of subspaces. If plausible, appending some sentences about why this inverse behavior happens may be interesting for the reader.

### **Other comments on the report and its structure?**

1. **Figure 2.1, 2.3, 2.4:** Ensure that the copyright owner of the figures has permitted the use of the figures.
2. **Section 4.4 Experiments:** Consider moving the section earlier in the chapter as a way to set the context for the rest of its structure. In addition, it provides a good understanding of how the various sections tie together before. Otherwise, the order of sections is very good.
3. Standardize on either American English or British English. For example, the author uses the spelling “modelling” commonly used in British English and the spelling “analyzing” commonly used in American English. Standardize terms, such as “2-D” or “two-dimensional”.
4. Refrain from using references as nouns. Author names may be used as done in the related work chapter if needed.
5. Allow for more hyphenation to minimize text-overflow.
6. Figures are not referenced in the text. Otherwise, the placements of figures are good.
7. Number the equations and reference them in-text.
8. The tense used in the chapter descriptions would be better if set in the present. The use of future tense (e.g., “will be”) makes the work sound incomplete.

### **What are the strong aspects of the work/report?**

1. The thesis surveys and implements state-of-the-art methods for its experiments. Moreover, it is based on recent, relevant, and published research. Subsequently, the author gave a sound basis for any findings to contribute to the research field and be newsworthy.
2. Much of the applied methodology uses methods relevant to the fashion research area from the general survey. Subsequently, it makes the work both compelling but also credible.
3. The thesis distinctly relates to the societal impact of the research topic and opportunities for sustainable fashion consumption. It contrasts the opportunity of the findings with a recognition of the computer power and energy requirements of deep learning and the outfit ranking loss function. The author rightfully weighs the possible good contribution of the work with the potential negative contribution due to its application in fast fashion instead of second-hand fashion. Furthermore, it recognizes and discusses the risk of bias in the data and its potential discriminatory effect. The thesis is well-anchored in ethics, sustainability, and its societal impact.

### **What are the weak aspects of the work/report?**

1. Several concepts used in the thesis need an explanation or clarification, preferably formal. There need not be extensive descriptions of everything, but a few inline sentences would be helpful, and some concepts may be considered known with the declaration of an intended reader. Undefined concepts or words create distractions. For example, an activation function need not be explained, but the concept of ablation studies should. See annotated report for details.

2. Missing references in many statements and paragraphs of the thesis. The references certainly are present in the bibliography, but there is limited use in-text. Many sentences that provide details not from the author do not declare from where the information originates. Place relevant references **after** a sentence or paragraph, unless it is perfectly understandable where the information comes from.

**What is your estimation of the news value of the work?**

Some of the research findings are noteworthy and likely contribute to new knowledge given the use of state-of-the-art methods. In addition, the increased difficulty in the fashion style compatibility task compared to the outfit matching task and the generalizability to unseen fashion item data set provides further clarity in the research area. Moreover, the higher dimensionality of the embedding representations improved the performance of the state-of-the-art models. However, the performance results for the state-of-the-art models on the fashion style compatibility task are likely not large enough to be newsworthy.

**Summarize the work in a few lines.**

The thesis implements a new industrial fashion item data set and implements two state-of-the-art models, SCE-Net and CSA-Net, for fashion style compatibility by adjusting the constraints of outfit matching. The state-of-the-art models were evaluated using the Fill-in-the-blank FITB method, item style retrieval with *recall @ top k*, and t-SNE visualizations. The results showed that fashion style compatibility is more challenging than outfit matching and that higher dimensionality for the embedding representations increased the performance. Moreover, both the state-of-the-art models generalized well onto an unseen fashion item data set. Future work should focus on the fashion item data set and ensure distinct styles in addition to texture details.

**Questions to the author that could be asked at the seminar (please list ca. six questions)**

1. The CSA-Net has the highest accuracy with a maximum of ten subspaces, whereas it has the lowest accuracy with seven subspaces, even less than three, four, or five subspaces. Any particular reason for this non-linear behavior?
  - a. Moreover, the SCE-Net had the best accuracy for the lowest number of subspaces, whereas the CSA-Net had the best accuracy for the highest number of subspaces. If plausible, appending some sentences about why this inverse behavior happens may be interesting for the reader.
2. How did you come up with the unconstrained outfit matching task? It is an interesting take for fashion style compatibility, but it is not clear how this came to be envisioned. Would be interesting to know the intuition behind it.
3. For the description of Fill-in-the-blank (FITB), four options are mentioned to be used in each answer set. How was the number four chosen?
  - a. Is it a standard?
  - b. What do you think the impact would be for a larger amount of options?

4. Apart from the chosen state-of-the-art models SCE-Net and CSA-Net, were there any other models that could have been used as well that were not? If so, why weren't they considered? Similarly, were there any other loss functions other than the triplet loss and outfit ranking loss that could have been considered?
5. How well do the findings generalize to an industrial fashion item data set created by another company other than the host company?
6. As briefly mentioned in the limitations of the study, how much does the subjectivity of what is a stylistic fashion collection impact the results of the study? Especially considering the personal opinion of what is considered compatible or not in fashion.

**What is your concluding assessment of this Master's project?**

The thesis concerns a fascinating research topic of great societal matter and a creative extension of the outfit matching task for fashion style compatibility. Furthermore, the methodology is both state-of-the-art and anchored in the fashion research field. The results are concisely presented and later critically discussed. The thesis has a good structure, and the author writes well. It would make an excellent read for anyone interested in fashion style compatibility modeling with a bit of polish and clarifications.