SUPPLEMENTAL MATERIAL: The Visual Language of Fabrics

VALENTIN DESCHAINTRE*, Adobe Research, UK
JULIA GUERRERO-VIU*, Universidad de Zaragoza - I3A, Spain
DIEGO GUTIERREZ, Universidad de Zaragoza - I3A, Spain
TAMY BOUBEKEUR, Adobe Research, France
BELEN MASIA, Universidad de Zaragoza - I3A, Spain

The supplemental material of this paper includes:

- Our full text2fabric dataset with a web-based browser to explore it: https://valentin.deschaintre.fr/text2fabric
- Additional results of the following tasks demonstrated in the main document (Section 5):
  - Image-Based Search (with real photographs as input)
  - Text-Based Fine-Grained Retrieval (with user-provided queries as input)
  - Caption Generation (both from synthetic and real images)
  - Latent Space Invariance to Geometry (from synthetic images)
- This pdf document, offering additional information and details on the following topics:
  - (S1) User Study Details
  - (S2) Additional Details: Rendered Images
  - (S3) Additional Details: Text Post-Processing
  - (S4) Classifying the Lexicon into Attributes
  - (S5) Additional Results from Dataset Analysis
  - (S6) Quantitative Evaluation of Negative Queries

S1 USER STUDY DETAILS

In this section we include additional information of the crowdsourcing user study to build our text2fabric dataset, including the instructions and interface given to describers, and supplemental details of our data verification protocol.

S1.1 Interface and Guidelines

In order to collect our natural language descriptions of fabric materials, we ran a crowdsourcing user study, using a web-based interface. Before launching our large-scale study, we ran several iterations of a pilot study using a small subset of the dataset in order to refine the interface and explanations given to the describers, as well as to select a collection of high-quality description examples to use as guidelines. In the final study, describers were shown one 4K resolution image of a fabric material at a time, rendered on our *baseline* geometry and illumination. As we required the descriptions to capture fine details of the fabric appearance, we further showed three zoomed-in areas of the image. These close-ups were tested in the pilot study and shown to help describers better appreciate fine details such as the stitching or the weave of the fabric. We include in Figure 1 a set of examples of the stimuli shown in our interface.

Before taking part in the study, describers who met the inclusion criteria were required to conduct a short training, including reading through a set of guidelines and passing a qualification test. The instruction guidelines were as follows:

- How to describe the fabric:
  - Imagine you are describing the fabric to someone *who can not see* the fabric themselves.
  - Your answer must cover **all descriptive aspects** of the fabric in your own words (including, for example, color, touch, the weave of the fabric, distinctive patterns, yarns...)
  - You can use: adjectives, comparisons, references to materials you know (cotton, silk, wool...), cultural textile design references (Dos: "this is a Japanese looking fabric with [...]", Don'ts: "this is a fabric I would find in my parents place"..), etc.
  - The description must be made up of complete sentences in English.
  - Check that you have not made grammar or spelling mistakes.
  - You should not look at the drape of the fabric (how it physically folds, e.g., the wrinkles that are formed) but rather at its appearance.
- Take a closer look at the fabric:
  - Use the Zoom In function to better observe the fabric, stitching and colors close up. To do this, click the magnifier icon.
- Things to avoid:
  - Do not describe the fabric by creating lists of the features, for example: "Purple, rough, wool, striped, embroidered". Instead, use complete sentences and precise descriptions, e.g.: "It is a rough fabric, perhaps wool. The fabric has two shades of purple, lighter and darker, which are interwoven by horizontal and vertical lines creating a checkered pattern".
  - Do not provide cultural textile design references of the fabric in a form of: "this is a fabric I would find in my grandparents place". Instead use "this is a boldly colored tartan with criss-crossing pattern, most closely identified with Scottish garb".
  - Do not submit your description without checking your grammar first.
  - Do not describe the fabric without zooming in to get a closer look of the fabric, stitching and colors.

[*] Joint first authors

Authors' addresses: Valentin Deschaintre*, Adobe Research, UK, deschain@adobe.com; Julia Guerrero-Viu*, Universidad de Zaragoza - I3A, Spain, juliagviu@unizar.es; Diego Gutierrez, Universidad de Zaragoza - I3A, Spain, diegog@unizar.es; Tamy Boubekeur, Adobe Research, France, boubek@adobe.com; Belen Masia, Universidad de Zaragoza - I3A, Spain, bmasia@unizar.es.

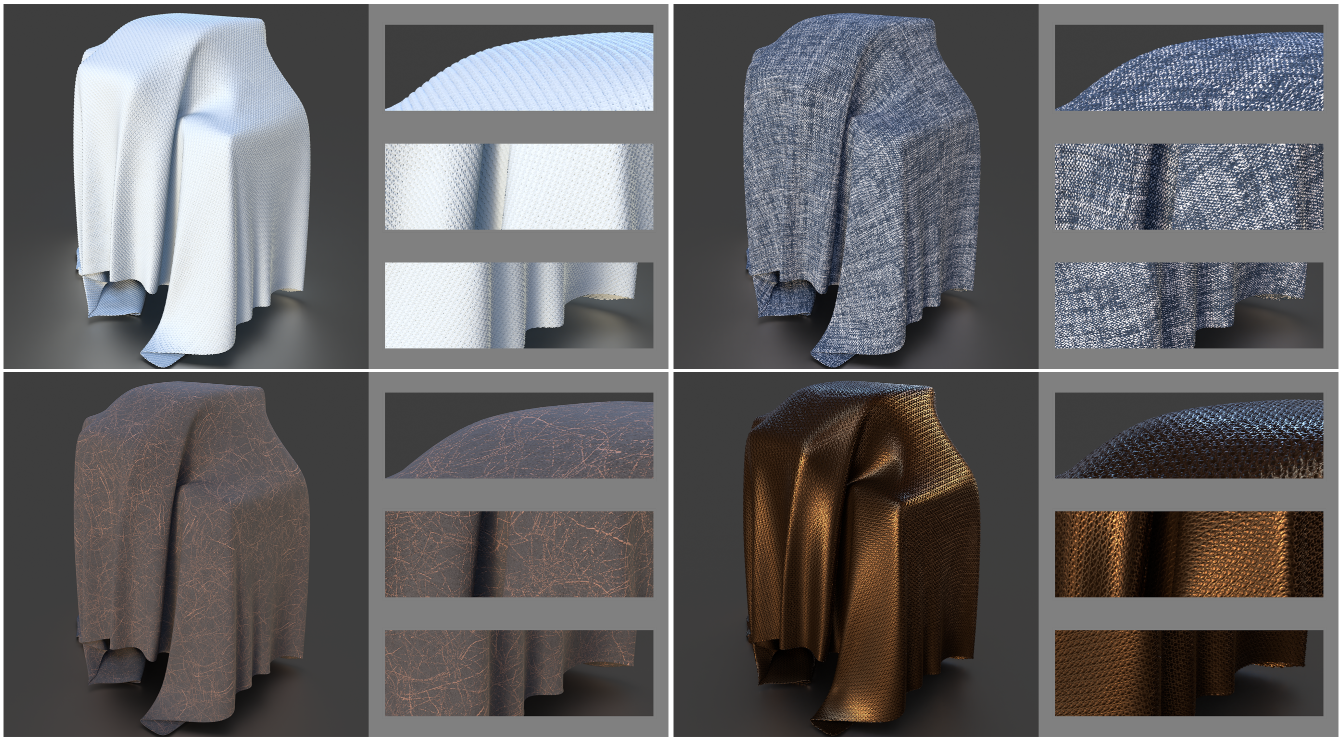
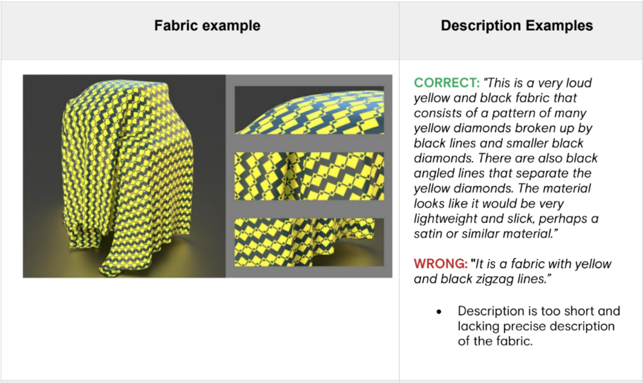


Fig. 1. Stimuli examples of the user study. We show describers the image of one fabric sample at a time, including three zoomed-in regions to better perceive the fine-grained details of the fabric appearance.



Fabric example	Description Examples
	CORRECT: *"This is a very loud yellow and black fabric that consists of a pattern of many yellow diamonds broken up by black lines and smaller black diamonds. There are also black angled lines that separate the yellow diamonds. The material looks like it would be very lightweight and slick, perhaps a satin or similar material."* WRONG: *"It is a fabric with yellow and black zigzag lines."* • Description is too short and lacking precise description of the fabric.

Fig. 2. Example of correct and wrong descriptions shown in the guidelines of the user study. These examples were selected from the results of a pilot study, and they helped describers to better understand the task.

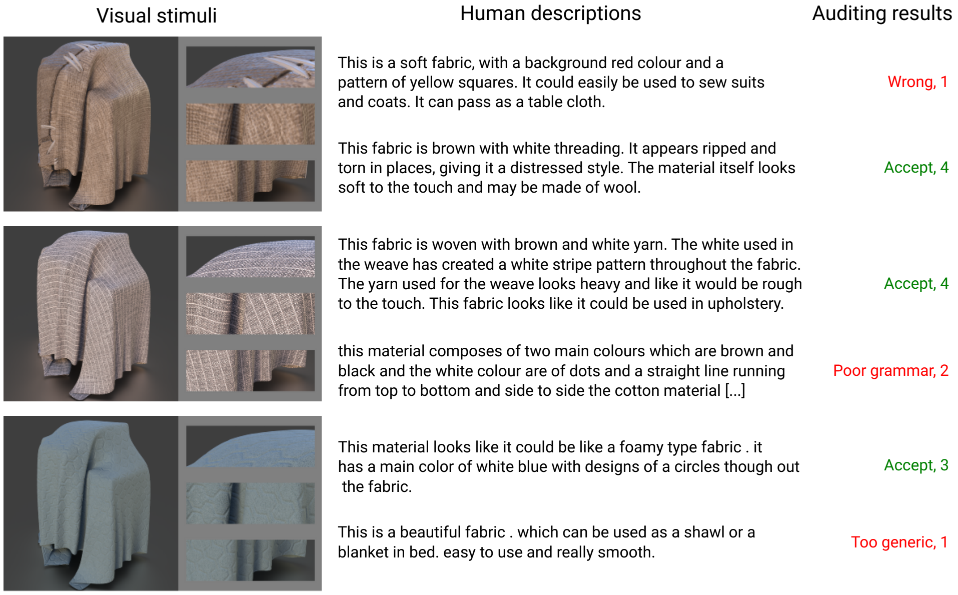
Additionally, to facilitate the understanding of the task, the guidelines included a set of ten *correct*/*wrong* example descriptions, such as the ones shown in Figure 2.

The qualification test consisted on a small task in which participants had to describe ten test images. Describers that did not follow the guidelines, provided consistently incorrect or highly generic descriptions were discarded.

S1.2 Data Verification Additional Details

We gathered a total of 19,167 free-text fabric descriptions. Verifying the quality of a large-scale dataset of natural language descriptions is a challenging task. We implemented automatic quality checks to remove descriptions duplicated or copy-pasted from the guidelines. In addition, we followed an iterative protocol to gather descriptions, using subsets of our dataset in consecutive batches of increasing size, in order to continuously check the quality of the data provided by the pool of describers. For the first initial batches, we manually audited *all* the gathered descriptions. This auditing involved labeling every description as either *accepted*, or rejected due to the description being *too generic*, being *wrong*, or using *poor grammar* to the point of hindering understandability. Additionally, it also involved giving every description a 5-point scale rating (1=totally unacceptable, 2=unacceptable, 3=acceptable, 4=very good and 5=excellent). We include several examples of descriptions and their auditing results in Figure 3. We can observe how descriptions marked as rejected with 1-2 ratings are either unintelligible, overly generic, include highly subjective/personal information or clearly do not match the appearance of the input image, which invalidates them for further use.

After full manual auditing of the first batches, we observed that the rating (and rejection rate) was highly correlated with the participant ID. The mean standard deviation of rating per participant was 0.667 ± 0.187 (median = 0.661), i.e., standard deviations within a particular describer are consistently low (describers were either



Human descriptions	Auditing results
This is a soft fabric, with a background red colour and a pattern of yellow squares. It could easily be used to sew suits and coats. It can pass as a table cloth.	Wrong, 1
This fabric is brown with white threading. It appears ripped and torn in places, giving it a distressed style. The material itself looks soft to the touch and may be made of wool.	Accept, 4
This fabric is woven with brown and white yarn. The white used in the weave has created a white stripe pattern throughout the fabric. The yarn used for the weave looks heavy and like it would be rough to the touch. This fabric looks like it could be used in upholstery.	Accept, 4
this material composes of two main colours which are brown and black and the white colour are of dots and a straight line running from top to bottom and side to side the cotton material [...]	Poor grammar, 2
This material looks like it could be like a foamy type fabric . it has a main color of white blue with designs of a circles though out the fabric.	Accept, 3
This is a beautiful fabric . which can be used as a shawl or a blanket in bed. easy to use and really smooth.	Too generic, 1

Fig. 3. Examples of gathered descriptions, and their auditing results. During auditing, descriptions are labeled as either *accepted* (green), or rejected (red) due to the description being *too generic* or subjective, being clearly *wrong* because it does not correspond to the shown image, or using *poor grammar* or spelling to the point of hindering understandability. Additionally, each description is given a rating in a 5-point scale, from 1=totally unacceptable to 5=excellent.

consistently good or consistently bad at the task). Based on this observation, for the rest of the dataset we established a semi-automatic per-describer data verification protocol, as follows:

- We audited manually a minimum of 40 randomly chosen descriptions per describer.
- For describers with a rejection rate < 10% and an average rating ≥ 3.15, we accepted all their remaining, non-audited descriptions.
- For describers with a rejection rate > 35%, we rejected all their remaining, non-audited descriptions.
- For describers between these bounds, we continued the individual auditing of their descriptions until they fell into one of the previous two categories, or we audited all their descriptions.

The specific thresholds for the categories above were established to find a trade-off between having a good quality dataset of descriptions and a tractable manual auditing process.

After this exhaustive data verification, we manually audited 6,614 descriptions, and 12,553 were automatically classified (34.5% manual auditing rate), having a final dataset of $15,461$ valid descriptions and $3,706$ invalid ones (19.3% rejection rate). A total of 122 describers took part in our user study, contributing with a mean of 157.11 descriptions per person.

S2 ADDITIONAL DETAILS: RENDERED IMAGES

Our dataset includes 3,000 fabric materials rendered on five different geometries (shown in the main document) and three different illuminations; we show in Figure 4 the environment maps used.

S2.1 Additional Image Statistics

We include additional statistics of our text2fabric images, compared to general-purpose datasets like ImageNet [Deng et al. 2009] and LAION [Schuhmann et al. 2021] in Figure 5. We show histograms of luminance and gradients distribution, in addition to the GLCM entropy already included in the main document. As expected, the image statistics of our data differ from those of general-purpose datasets, reflecting the specific nature of our images.

S3 ADDITIONAL DETAILS: TEXT POST-PROCESSING

To post-process our natural language descriptions, we first filter stop words as follows. We first include a general-purpose stop words list [sto 2023] with an initial set of 851 words, that contains standard prepositions, linking words, pronouns, and non-meaningful verbs and adverbs. Additionally, we manually extend this list to our context with 334 extra words, including very high-level or highly-subjective concepts (e.g., 'beautiful'), and specific context words (e.g., 'fabric', 'material', 'image'). This creates a full list of 1,185 stop words that do not represent meaningful concepts in the vocabulary of fabrics descriptions.

To correct potential spelling mistakes present in the descriptions, we use the Autocorrect spelling corrector [Sondej 2023]. Additionally, for lemmatization and part-of-speech tagging, we use the standard spaCy library [Honnibal et al. 2020].

We show examples of lemmas from our lexicon together with the list of types that correspond to each of them in Table 1. We observe how lemmas are very useful to group words that represent the same concept, allowing us to conduct a more compact and effective analysis of the textual data.

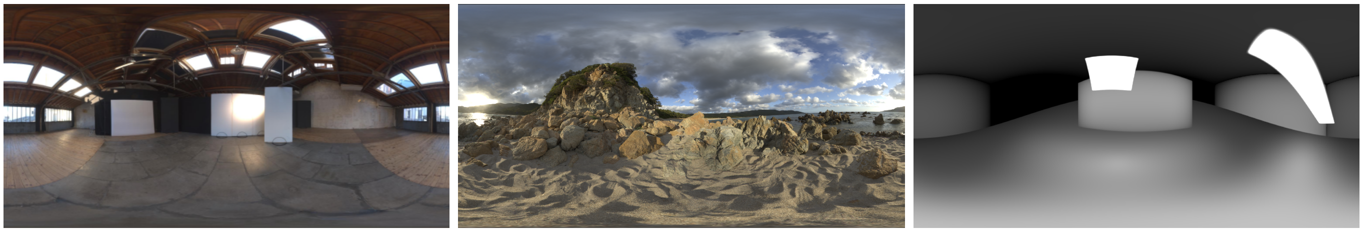
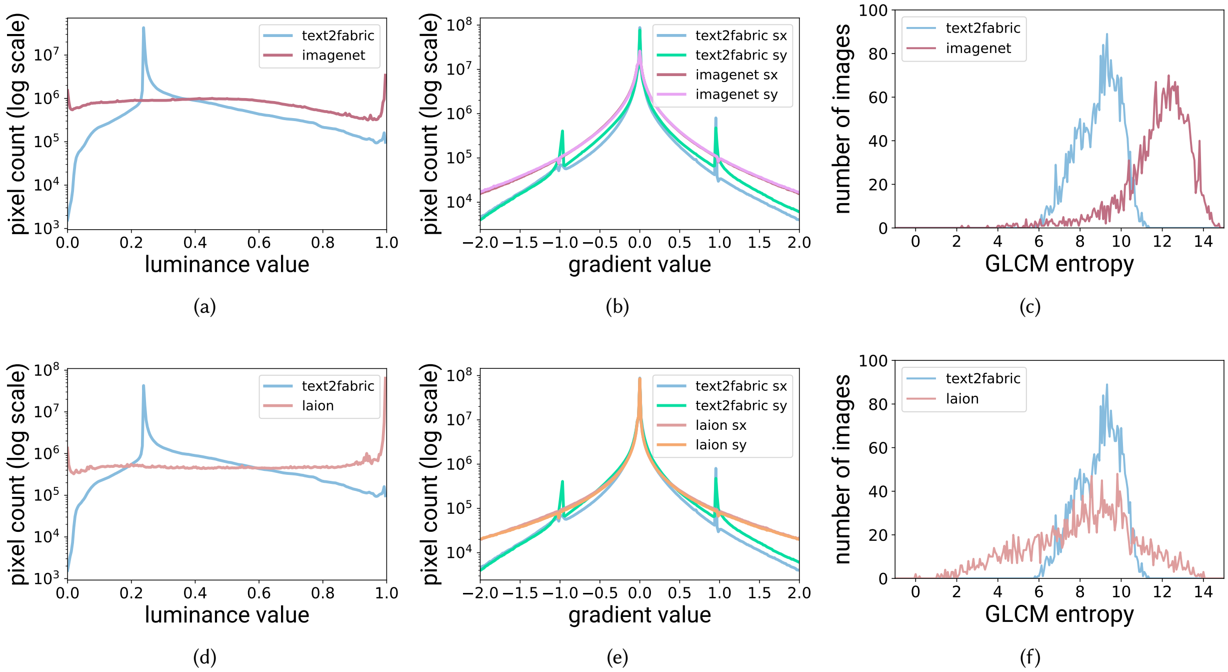


Fig. 4. Environment maps used in our text2fabric dataset. *Left:* Interior Atelier Soft Daylight (our *baseline* illumination). *Center:* Corsica Beach (*outdoor*). *Right:* Studio 6 (*studio*).



(a)

(b)

(c)

(d)

(e)

(f)

Fig. 5. Image statistics of the text2fabric dataset compared to general-purpose large-scale datasets. We show, *from left to right*, histograms of: luminance, gradients and GLCM entropy. *Top row:* Comparison to a random subset of 3,000 ImageNet images. *Bottom row:* Comparison to a random subset of 3,000 LAION images. In the luminance histograms, the grey background in our images yields a clear peak in the corresponding luminance value, similar to the one on the white luminance value for the LAION images, which often feature a white uniform background, whereas ImageNet data do not show such prominent peak.

Before post-processing, our text2fabric dataset included 543,708 tokens and 9,379 types. After post-processing, it includes 191,783 tokens, 3,539 types and 2,762 lemmas.

Our fabric-specific lexicon contains 524 lemmas that cover 95% of the descriptions (see Section 4.1 of the main document). This lexicon includes very common terms that apply to a wide variety of fabrics (e.g., ‘pattern’, ‘soft’, ‘cotton’) and more specific concepts (e.g., ‘hexagonal’, ‘denim’, ‘embroider’, ‘jacket’). On the other hand, our lexicon does not include uncommon lemmas that appear in the descriptions but are rarely used (i.e., have low *arf* value), such as: ‘china’, ‘tiger’, ‘citrus’, ‘shelf’ or ‘strawberry’.

S4 CLASSIFYING THE LEXICON INTO ATTRIBUTES

In this section we include a step-by-step description of our methodology to classify the lexicon of 524 lemmas into eleven identified attributes. Our objective is to find the common attributes present within the fabric descriptions lexicon, which we pose as a clustering problem. To solve this problem, we start by taking the 250 most prominent lemmas from our lexicon (~90% coverage) and computing their ConceptNet Numberbatch embeddings [Speer et al. 2017]. We then build a matrix with the pair-wise cosine similarities of these embeddings, and use it to run the affinity propagation algorithm [Frey and Dueck 2007] and cluster the lemmas. Unlike other typical clustering algorithms such as k-means, affinity propagation does not require to estimate the number of clusters beforehand, and is able to find a representative word per cluster, named ‘exemplar’.

Table 1. Example lemmas from our text2fabric descriptions and the list of types associated to them.

Lemma	Types			
color	colors	color	coloring	colored
pattern	pattern	patterns	patterning	patterned
brown	browns	brown		
light	light	lighter	lighting	lightest
soft	softer	soft		
design	designs	design	designed	designing
dark	darker	dark		
touch	touch	touching	touched	touches
weave	weave	weaving	woven	weaves
line	lines	lined	line	lining
texture	texture	textured	textures	
stitch	stitching	stitched	stitch	stitches
thin	thinner	thin		
camouflage	camouflage	camouflaged	camouflaging	
shape	shapes	shaping	shape	shaped

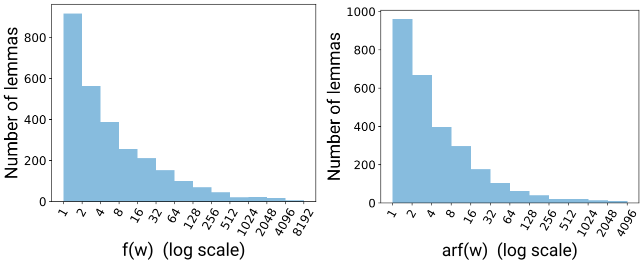
Therefore, the algorithm is not biased towards a fixed number of attributes, and the resulting clusters are more intuitive to understand. The resulting 15 exemplars were the following: ‘soft’, ‘blue’, ‘heavy’, ‘knit’, ‘bright’, ‘horizontal’, ‘gold’, ‘nylon’, ‘linen’, ‘blotch’, ‘zigzag’, ‘floral’, ‘rectangle’, ‘jacket’ and ‘military’.

We then merge together clusters that represent the same high-level notion or attribute in terms of (fabric) appearance: we group the clusters ‘linen’ and ‘nylon’ into a common cluster named *fabric_type*, and the clusters ‘horizontal’, ‘zigzag’, ‘floral’ and ‘rectangle’ into a more general cluster named *pattern*. The other nine clusters are kept as is, constituting the eleven attributes. Finally, we give each cluster (attribute) a representative name; we do not use the exemplar, which is just the lemma closest to the centroid, but rather a higher-level concept (e.g., ‘blue’ is the exemplar of the cluster ‘color’).

Since what we are seeking is a reliable initial classification into attributes that can then be extended to the full lexicon, once the attributes are established we manually re-classify some of this subset of 250 lemmas. Using this initial classification, we automatically classify the rest of the lemmas from our lexicon. For every lemma, we select the most common cluster (attribute) within the nearest neighbours of its embedding, filtered by a threshold similarity ($th_s = 0.80$). This classification is further checked manually. In the end, the automatic classification obtains a top-1 accuracy of 63.87% and a top-3 accuracy of 87.23% within the remaining lemmas of the lexicon. While the whole process to classify the lexicon into attributes has certain manual steps, it should be noted that we do not aim to propose a clustering algorithm; rather, our focus is on evaluating whether a proper classification into meaningful attributes can be found for the lexicon used in natural language descriptions of fabric appearance. Figure 7 shows the embeddings of the most prominent 250 lemmas from our lexicon (reducing their dimensionality from 300D to 2D with t-SNE) and their classification into attributes. We can observe how some attributes are more clearly clustered in the space (e.g., *color*, *lightness*), while others with a more general meaning are widely spread (e.g., *pattern*, *use*).

S5 ADDITIONAL RESULTS FROM DATASET ANALYSIS

We include here additional details and results of our analysis of the text2fabric dataset included in Section 4 of the main document.



(a)

(b)

Fig. 6. Histograms of frequency of lemmas. *(a)* Absolute frequency of lemmas ($f(w)$). *(b)* Average reduced frequency of lemmas ($arf(w)$).

S5.1 Average Reduced Frequency Computation Details

To analyze natural language data at word-level, one of the most typical statistics to compute is the absolute frequency per word [Brezina 2018]. More precisely, as we are analyzing the data based on lemmas (see Section 3.2.3 of the main document), we can define the absolute frequency per lemma $f(w)$ as the total count of occurrences (i.e., the total number of tokens) that belong to a particular lemma within our corpus (i.e., the full text2fabric dataset). We show the distribution of $f(w)$ in our dataset in Figure 6a.

However, the prominence or importance of a term within a corpus is not only determined by the number of times it occurs, but also by its dispersion. Therefore, we compute a metric named *average reduced frequency* [Brezina 2018; Savickỳ and Hlavácová 2002], which combines both the absolute frequency of a given term with its dispersion in a corpus: the more frequent and evenly distributed it is, the more prominent it is considered to be. Average reduced frequency of a lemma w present in our corpus is computed as follows: the corpus is subdivided into x parts of the same size, where $x = f(w)$, and a measure of *reduced frequency*, $rf(w)$, is obtained as the number of those parts that include at least one occurrence of the lemma, i.e., $rf(w) \in [1, x]$. The intuition is that, if a lemma occurs several times, but all its occurrences are very close to each other in the corpus, the number of parts including it will be smaller than its absolute frequency, hence the name *reduced*. In our case, we sort our full corpus of descriptions by describer and then we compute the average reduced frequency per lemma ($arf(w)$). We refer the reader to related literature [Brezina 2018; Savickỳ and Hlavácová 2002] for further details about the mathematical definition of the metric and its implementation. Results are shown in Figure 6b. We can observe how the distributions of $arf(w)$ and $f(w)$ are similar, suggesting the existence of a common vocabulary for fabrics descriptions as most of the lemmas with a high frequency in our dataset are also evenly distributed among describers.

S5.2 Structure of Descriptions: Additional Details

We study the structure of descriptions by analyzing the order of appearance of our attributes, computing their rank products, as explained in the main document (Section 4.3). Figure 8 shows complete rank histograms per attribute, ordered from lowest to highest rank product. These histograms show, for each attribute, the frequency of occurrence of lemmas of such attribute in each rank order within the descriptions.

S6 QUANTITATIVE EVALUATION OF NEGATIVE QUERIES

Without ground-truth data for negative queries, a systematic quantitative evaluation becomes difficult. We conduct a preliminary evaluation by extracting sentences including negative bigrams and trigrams from our test set (e.g., “not shiny”, “without any pattern”), and using them as negative queries for text-based retrieval. Compared to native CLIP, our model achieves 5x better top-5 recall, with at least a 3.95x improvement for all top-K recall results. We think that detailed descriptions do not help with negative modifiers per se, but rather that our dataset contains examples of them which are learnt by our model. As suggested by Figure 16 in the main document, our model may not learn the notion of negatives in general, but rather specific negatives that are typically used to describe fabrics appearance.

REFERENCES

2023. English Stop Words List. https://countwordsfree.com/stopwords.

Vaclav Brezina. 2018. *Statistics in Corpus Linguistics: A Practical Guide.* Cambridge University Press.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A Large-scale Hierarchical Image Database. In *IEEE Conference on Computer Vision and Pattern Recognition 2009*. 248–255.

Brendan J Frey and Delbert Dueck. 2007. Clustering by Passing Messages between Data Points. *Science* 315, 5814 (2007), 972–976.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python. (2020). https://doi.org/10.5281/zenodo.1212303

Petr Savickỳ and Jaroslava Hlavácová. 2002. Measures of Word Commonness. *Journal of Quantitative Linguistics* 9, 3 (2002), 215–231.

Christoph Schuhmann, Richard Vencu, Romain Beaumont, Robert Kaczmarczyk, Clayton Mullis, Aarush Katta, Theo Coombes, Jenia Jitsev, and Aran Komatsuzaki. 2021. LAION-400M: Open Dataset of CLIP-Filtered 400 Million Image-Text Pairs. *arXiv preprint arXiv:2111.02114* (2021).

Filip Sondej. 2023. Autocorrect - Spelling Corrector in Python. https://github.com/filyp/autocorrect.

Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An Open Multilingual Graph of General Knowledge. In *The 31st AAAI Conf. on Artificial Intelligence*.

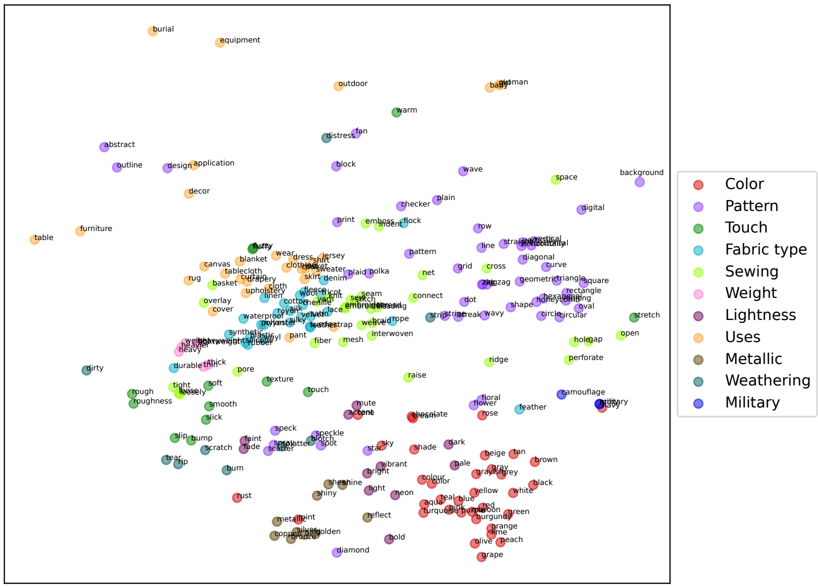


Fig. 7. Visualization of the embeddings space for the most prominent 250 lemmas and their clustering into attributes. We show every lemma as a point in 2D space using t-SNE dimensionality reduction (300D to 2D), indicating its associated attribute by the color of the point.

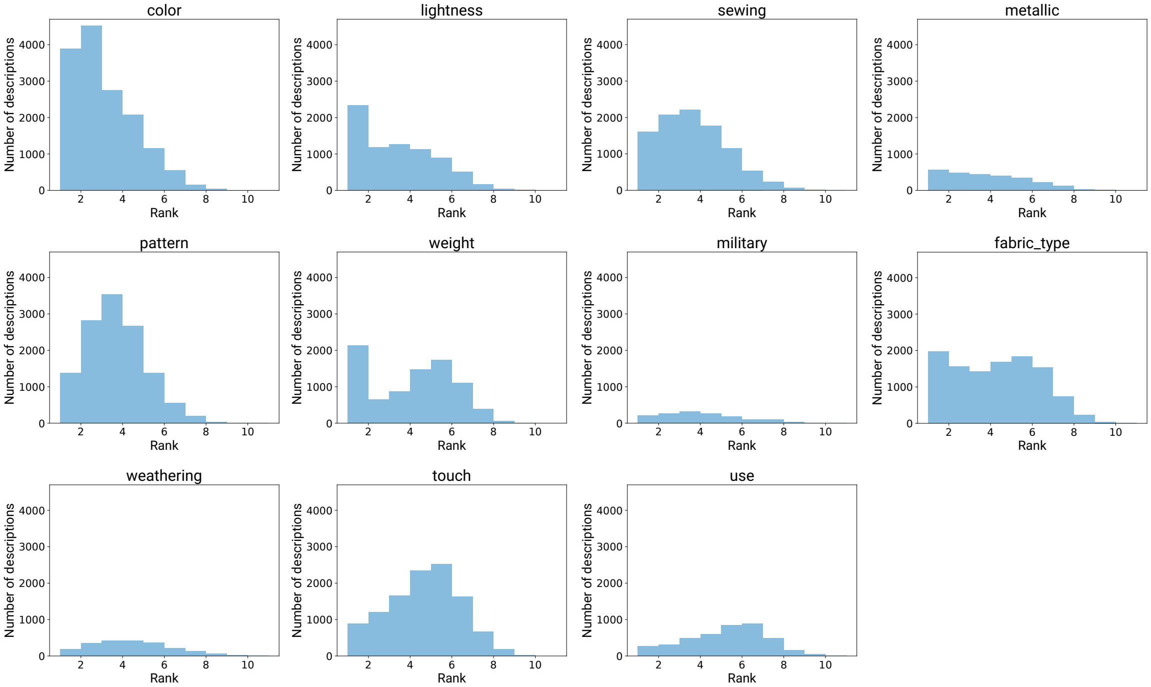


Fig. 8. Rank distribution of each attribute within the descriptions.