






Article

Unsupervised Clustering Pipeline to Obtain Diversified Light Spectra for Subject Studies and Correlation Analyses

Stefan Klir ^{1,*} , Reda Fathia ¹ , Sebastian Babilon ^{1,2} , Simon Benkner ¹  and Tran Quoc Khanh ¹ 

¹ Laboratory of Lighting Technology, Technical University of Darmstadt, Hochschulstr. 4a, 64289 Darmstadt, Germany; reda.fathi@gmail.com (R.F.); babilon@lichttechnik.tu-darmstadt.de (S.B.); benkner@lichttechnik.tu-darmstadt.de (S.B.); kxanh@lichttechnik.tu-darmstadt.de (T.Q.K.)

² Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, Light and Health Research Center, One Gustave L. Levy Place, New York, NY 10029, USA

* Correspondence: klir@lichttechnik.tu-darmstadt.de; Tel.: +49-6151-16-22878

Featured Application: Selection of most diverse light spectra from a larger set of possible candidates to be used in subject studies or for machine learning to find correlations between photometric and other parameters such as psychological, physiological, or preference-based outcome measures.

Abstract: Current subject studies and data-driven approaches in lighting research often use manually selected light spectra, which usually exhibit a large bias due to the applied selection criteria. This paper, therefore, presents a novel approach to minimize this bias by using a data-driven framework for selecting the most diverse candidates from a given larger set of possible light spectra. The spectral information per wavelength is first reduced by applying a convolutional autoencoder. The relevant features are then selected based on Laplacian Scores and transformed to a two-dimensional embedded space for subsequent clustering. The low dimensional embedding, from which the required diversity follows, is done with respect to the locality of the features. In a second step, photometric parameters are considered and a second clustering is performed. As a result of this algorithmic pipeline, the most diverse selection of light spectra complying with a given set of relevant photometric parameters can be extracted and used for further experiments or applications.

Keywords: light clustering; diversified light spectra; spectral embedding; light selection; spectral feature selection



Citation: Klir, S.; Fathia, R.; Babilon, S.; Benkner, S.; Khanh, T.Q. Unsupervised Clustering Pipeline to Obtain Diversified Light Spectra for Subject Studies and Correlation Analyses. *Appl. Sci.* **2021**, *11*, 9062. <https://doi.org/10.3390/app11199062>

Academic Editor: José A. F. O. Correia

Received: 8 July 2021

Accepted: 21 September 2021

Published: 28 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Modern indoor LED-based lighting solutions make use of several different LEDs with different spectral emission characteristics to allow for a dynamic adjustment of correlated color temperature, color rendition and other more physiologically relevant parameters [1]. Depending on their purpose of application, multi-channel LED systems can for example be tailored to achieve more preferred light conditions, improve perceived naturalness, or increase the users' state of health and well-being [2–5].

In contrast to many traditional light sources, such as incandescent or compact fluorescent light bulbs, LEDs are driven by direct current (DC) and, therefore, require suitable dimming strategies. Commonly used LED driver control methods that have successfully been applied in the smart-lighting context include analog 0–10 V dimming [6–8] as well as digital protocols such as DALI [9–11] or ZigBee [12–14] standards. In the former case, a DC voltage between 0 and 10 V is used as a simple control signal to be sent to the actual LED driver [6], which eventually adjusts the LED load current and, thus, the light output in the range from minimum (typically between 1 and 10% depending on the driver design) to 100%. The corresponding electrical specifications are defined by the IEC standard 60929, Annex E [15] for both DC voltage and pulse-width-modulated (PWM) LED control. Instead of using analog control signals, the various digital protocols on the other hand make use of

encoded data packets to be distributed within the lighting device network to send dedicated commands to the individual LED drivers. The DALI standard, for example, which is regulated by the IEC standard 62386 [16], enables digital information to be transferred between the DALI controller and all network devices by means of an asynchronous, half-duplex, serial protocol over a two-wire differential bus at a fixed rate of 1200 bit s^{-1} [11]: The DALI controller, serving as the master device, sends 16-bit Manchester-encoded commands to any addressable slave device on the network (forward direction), which, in turn, may respond with 8-bit Manchester-encoded messages (backward direction) that include, e.g., information on the status of the corresponding luminaire and LED drivers.

Despite these fundamental differences between analog and digital dimming strategies, it should be emphasized that, regardless of the chosen method of control, the system's complexity as well as the number of theoretically composable light spectra increase disproportionately fast as the number of LED channels that are built into modern luminaires increases [17]. In current lighting research, light spectra are usually selected based on some pre-defined lighting parameters for which the luminaires' emission will be optimized [18–21]. However, it should be noted that spectra with an equal chromatic appearance can exhibit quite different spectral compositions, which is known as metamerism [17,22,23]. This commonly leads to a bias in the optimization-based selection of spectra for lighting research as one can never be sure that a particular selection is most representative in terms of generalizability. The use of a different selection of metameric spectra might alter the study results and, as a consequence, enforce potentially distinct study conclusions.

To profoundly explore the more fundamental relationships and interactions between lighting conditions and other, in particular, dynamic output variables, such as for example lighting preference as a function of the observer's current psychological or emotional state, it is necessary to have an unbiased selection of light spectra that are as diverse to each other as possible. Here, diversity means that the differences in the features representing the spectra must be maximized so that the span of the corresponding spectral space is maximized, too [24]. This optimized space can thus be explored for certain hypothesized relationships and parameter correlations in a completely unbiased way.

In this work, a novel machine-learning-based approach will be presented which makes use of this diversifying strategy to select the most distinct light spectra from a given larger set of possible candidates, while fulfilling certain photometric and colorimetric requirements. For simplicity, one can think of this approach as a two-step process: First, a latent space representation of the complete spectral information is used to cluster similar light spectra based on their relevant (decomposed) features. As light spectra belonging to different clusters are most diverse in nature, selecting a spectrum from each of these clusters automatically results in a maximal span of the spectral (sub-)space. Before that, however, a second processing step is needed. Based on a pre-defined selection of lighting parameters, each of the main clusters is further divided into characteristic sub-clusters. Subsequent filtering eventually ensures that the finally selected light spectra—most diverse in nature—comply with the pre-determined photometric requirements.

A detailed discussion on the implementation of the algorithmic pipeline as well as an example of clustering results obtained for a six-channel LED luminaire will be presented in the following.

2. Materials and Methods

The spectral composition of a light source is usually defined in the range from 380 nm to 780 nm showing a resolution of 1 nm, which basically results in a total number of 401 different parameters that encode the light source's (relative) intensity distribution per unit wavelength. Under this aspect, each set of arbitrary light spectra can be described as vectors in a high dimensional space of $d = 401$ dimensions. Dedicated dimension reduction algorithms can thus be applied to map this high dimensional space to a latent representation of the light spectra of lower dimension. This latent space representation of the spectral infor-

mation is subsequently used to cluster similar light spectra based on their most relevant features. A trend to use such low dimensional representations of features describing higher dimensional data can be seen in current research of representation learning. Corresponding literature clearly shows the progression from manually selected features [25–27] towards modeling the features in a latent space for further processing [28–30].

As entities of different clusters are by definition most diverse in terms of their features, picking a light spectrum out of each cluster eventually results in the most distinct selection of light spectra that can be determined from a given larger set of possible candidates used as input data. In order to avoid any initial bias with regard to this diversifying strategy, feature extraction and main clustering must account for the full spectral information. If, instead, these processing steps were based on pre-selected photometric quantities or some other lighting parameters obtained from integrating the correspondingly weighted light spectra over wavelength [31,32], potentially important spectral information would get lost and a considerable bias would be introduced.

Clustering approaches, in general, provide a fundamental pillar in unsupervised machine learning to detect the hidden structures behind arbitrary data [33]. For an adequate performance, though, they usually require to manually select and optimize the input features used for cluster definition, which, depending on the specific nature of the data, can be a very time consuming and naturally biased process. However, latest progress in computational research has led to the development of various deep-learning based algorithms that can effectively be applied to automatically extract the most relevant features from almost any kind of raw data [34].

Since light spectra are highly-dimensional complex objects, eligible cluster algorithms must be non-linear in nature taking into account the local structure of the data in order to be able to identify the underlying correlations. For this reason, standard approaches, such as principal component analysis (PCA) [35], that preserve the global structure of the data and transform the input by seeking a linear transformation into a new embedded space, are not suitable for the present use-case and are, therefore, excluded from further considerations. On the contrary, more sophisticated feature reduction algorithms, like for example Autoencoders [34], have shown to perform well on non-linear transformations by learning the local structure of the input data through feature variables by encoding and subsequent decoding of the input data [36,37]. Thus, the resulting feature vector can be improved and the loss between in- and output can be minimized accordingly. Further details on the algorithmic implementation and the spectral input data will be discussed in the following sections

2.1. Test Setup and Input Data

In order to provide an application-related proof of concept, the present work makes use of spectral data obtained from a real multi-channel LED luminaire, in this case a direct/indirect floor lamp with three distinct, individually addressable LED channels found in both the direct and the indirect component. Figure 1a shows the corresponding relative spectral power distributions (SPDs). This mixture comprises two phosphor-converted white LEDs (cool and neutral white) and an additional single-peaked cyan emitter, whose chromaticities are depicted Figure 1b. While the former allow for adjusting the luminaire's overall light emission over a wide range of different correlated color temperatures (CCTs), the latter—showing a peak wavelength of 475 nm—is used to specifically trigger the intrinsically photosensitive retinal ganglion cells (ipRGCs) for an additional stimulation of the human circadian system [38]. The corresponding LED drivers are controlled via a DALI-bus [39] and can receive values from 84 to 254 in order to dim the brightness of each channel between 1% and 100%, respectively. These 171 dimming steps for each of the luminaire's six individual LED channels (three within each directional component), yield a total number of $171^6 = 25.002 \times 10^{12}$ possible light spectra that may differ in their spectral compositions and absolute light levels.

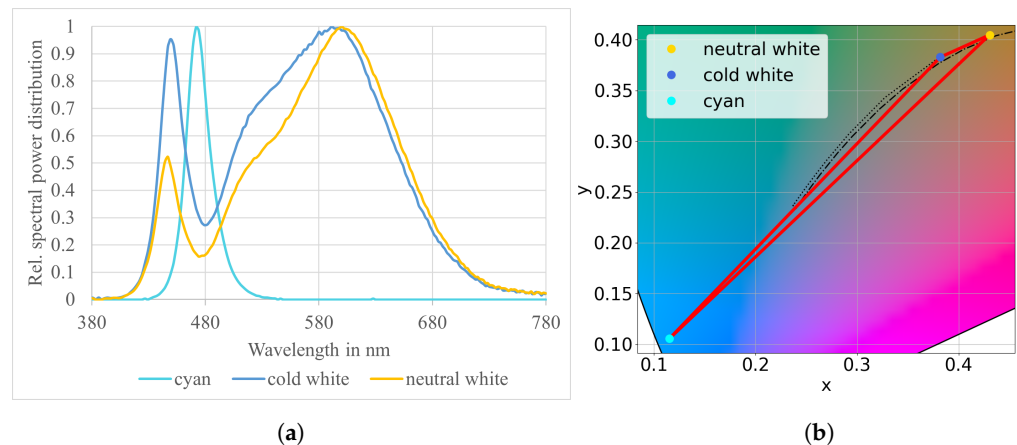


Figure 1. (a) Normalized LED spectra for the neutral white, cool white, and cyan primaries constituting the test luminaire's direct and indirect components. (b) x, y coordinates of the LED primaries illustrated in CIE 1932 chromaticity space. The corresponding gamut is shown as a red triangle.

This set of light spectra appears to be an adequate input to validate the proposed diversifying strategy on realistic data. As shown in Figure 2, which gives an overview of the computational structure, an unbiased clustering can be performed by using the complete physical information of the light spectra encoded as features in a 401 dimensional vector space. The clustered raw data are further processed by dividing each of the found main clusters into characteristic sub-clusters based on a selection of lighting parameters that are calculated for all clustered light spectra. The resulting parameterized sub-clusters can thus be used to efficiently filter the data to eventually come up with the most diverse selection of light spectra that comply with certain pre-defined user requirements. In order to avoid the bias introduced by the sub-clustering process, even though desired in most cases, the corresponding computational step can be skipped or bypassed entirely so that the final selection of diverse light spectra can be carried out on the unbiased main clusters only.

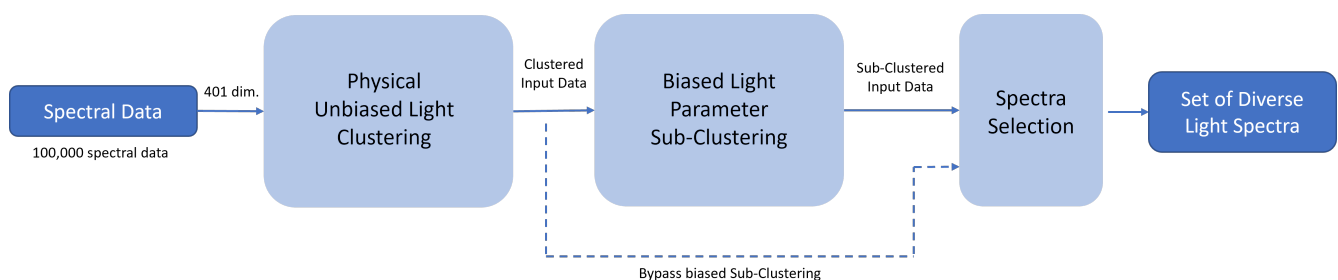


Figure 2. Overview of the computational structure of the proposed diversifying strategy. The input light spectra are clustered in an unbiased way by using their complete physical information encoded as features in a 401 dimensional vector space. Sub-clustering can eventually be performed based on pre-defined lighting parameters and may serve for a respective filtering of the main clusters. Light spectra belonging to different main clusters are maximally diverse to each other.

Due to the tremendous computational resources that would be needed to process the complete set of possible light spectra, a more practical approach had to be chosen. For this work, the size of the input data was limited in an unbiased way by randomly sampling 1×10^5 data from a uniform distribution representing the total set of 25.002×10^{12} possibilities. Following this strategy, the overall computation time could be limited to a feasible range of a few days. All calculations were performed on a linux server with 45 CPUs running at 2.6 GHz, 80 GB of RAM and a NVIDIA Quadro P4000 GPU.

2.2. Feature Reduction and Main Clustering

In order to cope with the high dimensionality of the input light spectra, suitable dimension reduction algorithms need to be applied first to find a computationally manageable representation of their most relevant features in a lower dimensional latent space that is adequate to perform the main clustering. Figure 3 summarizes the corresponding algorithmic pipeline.

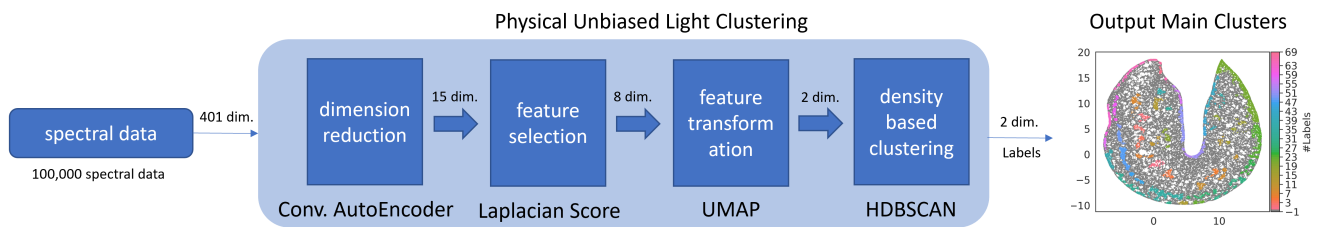


Figure 3. Algorithmic structure of the physical unbiased light clustering.

The first three steps of the depicted cluster approach take into account the local connectivity of the input data space. A convolutional autoencoder (conv. autoencoder) will reduce the dimension (dim.) to 15, which represents a trade-off between accuracy and a maximal reduction of the feature vector. The corresponding features will then be ranked by the Laplacian Score [40] and the less important upper half will be truncated. The remaining dimensions are further reduced and transformed to a two-dimensional representation using the uniform manifold approximation and projection (UMAP) [41] algorithm. As the distance between vectors within this maximally reduced space is no longer meaningful, the choice of clustering algorithms suitable for this kind of data is limited to density-based approaches that do not rely on the distance between individual data. Due to its robustness against noise, the hierarchical density-based spatial clustering of applications with noise (HDBSCAN) [42] algorithm has been chosen for this work. Further details on each of these algorithms will be given in the following sections.

2.2.1. Convolutional Autoencoder

Since their invention in the 1980s, autoencoders have successfully been applied to a huge variety of different tasks in computational research and engineering primarily focusing on dimensionality reduction and information retrieval [43–46]. They have been developed to find a feature representation (encoding) for a set of (high-dimensional) data by modeling the manifold of a dataset in a self-supervised manner while ignoring noise. In order to maintain the local structure of certain data, which is particularly important in the present case to be able to identify potential correlations between different light spectra, Masci et al. [47] proposed to extend the autoencoder approach by a convolutional architecture, where corresponding weights are shared among all locations in the input for preserving spatial locality.

To prepare the spectral data to fit the convolutional autoencoder (CAE) architecture, the corresponding 1×401 parameter vectors, one for each input light spectrum, are cropped and re-arranged in 20×20 dimensional arrays. For this purpose, the original wavelength range was truncated to 380–779 nm. The error introduced by this truncation is negligible with regard to lighting applications as the human visual system is almost completely insensitive to wavelengths larger than 750 nm.

The 20×20 dimensional arrays can then be interpreted as grey-scale pictures, where each pixel value represents the assigned wavelength's light level. By applying the CAE to the set of input data considered in this work, a significant feature reduction from 20×20 to 15 parameters could be achieved. As can be seen from Figure 4, the CAE consists of an encoding and decoding stage. During encoding, the CAE tries to find an appropriate, dimensionally reduced feature representation of the input light spectra. During decoding, this feature representation is then used to reconstruct the original 20×20 data. This allows the CAE to calculate the error between the input and output/reconstructed light spectra to optimize the internal layers with the aim of minimizing the error.

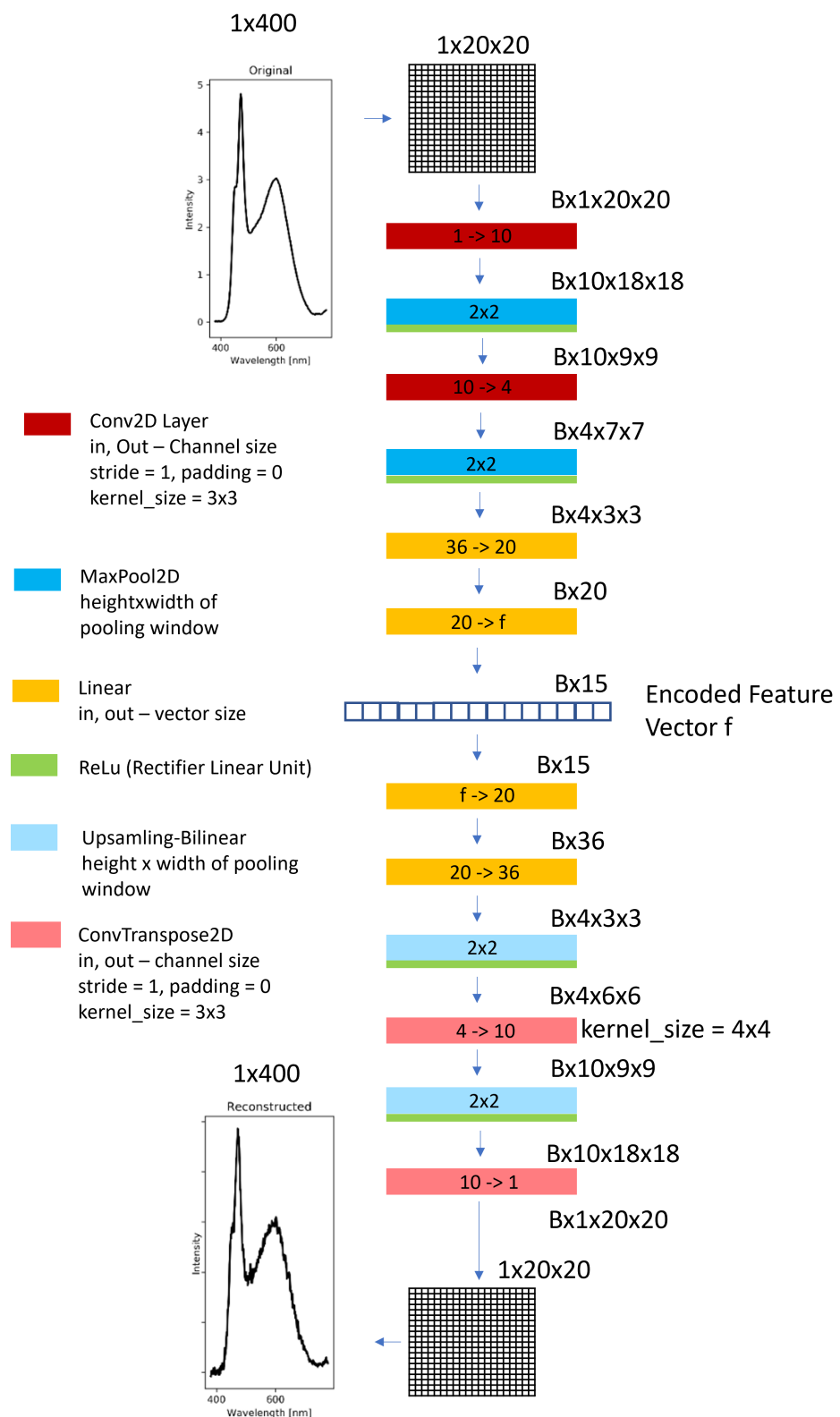


Figure 4. Illustration of the convolutional autoencoder used to process the spectral input data. Before processing, the light spectra are cropped to 380–779 nm and re-arranged as 20×20 dimensional data arrays. Subsequent encoding and decoding, while minimizing the corresponding loss function between in- and output data, resulted in a set of optimized 1×15 dimensional encoded feature vectors representing the original input light spectra.

A pre-processing step with a normalizing feature scaling between zero to one was applied to limit the range of the input data and to find the minimum (min) and maximum (max) of all possible input spectra. In this work, a kernel_size of 3 was used. The kernel_size defines the window area in which the data points are processed. With the spectral input data being represented as 20×20 arrays, the kernel windows applied to this array structure always include wavelengths from three non-contiguous regions of the spectrum. This results in the creation of a more generalized convolutional layer.

The used autoencoder basically combines two of such layers. The first one increases the dimension of the input data arrays to $10 \times 18 \times 18$, where each additional dimension represents a weighted kernel filtering of the data to learn the complexity of input spectra [43–45]. The second convolutional layer then reduces the data to $4 \times 7 \times 7$ dimensions after a 2×2 max-pooling layer [48]. This reduction of the amount of filters was performed to reduce the computational cost for the next fully connected layers. Afterwards, the fully connected linear layers are used to further reduce the dimensionality to come up with the final encoding of the light spectra in terms of 1×15 dimensional encoded feature vectors. The subsequent decoding steps are essentially given by the inverse of the encoding stage and used to reconstruct the original 20×20 dimensional data arrays representing the input light spectra from their reduced feature vectors. The in- and output data can be compared and the layers of the neural network and, thus, the feature representation can be optimized by minimizing the corresponding mean squared error (MSE) loss function.

2.2.2. Laplacian Score

After autoencoding, the complete set of input light spectra represented by their 1×15 dimensional feature vectors is post-processed by calculating the corresponding Laplacian score [40,49]. This allows for a further reduction of the dimensionality by keeping only the most important features. The Laplacian score algorithm evaluates the importance of a feature by means of its power of locality preserving. In order to model the local geometric structure of the data, a k -nearest neighbor graph is constructed, where the weight assigned to each edge of connected data points describes their similarity in terms of an Euclidean distance metric. Making use of locality preserving projections [50], the weight matrix is used to calculate the Laplacian score of the features based on Laplacian eigenmaps [51]. Features with smaller score values are in principle more important than features with higher values. In this work, the score median has been calculated and all features with a larger value have been discarded. The feature vectors representing the input light spectra can thus be reduced to 1×8 dimensions.

2.2.3. UMAP

The remaining dimensions are further reduced and transformed to a two-dimensional representation using the uniform manifold approximation and projection (UMAP) [41] method. Besides reducing computational effort with regard to a subsequent clustering, such a two-dimensional representation also allows for visualizing the corresponding results. In this context, UMAP is an efficient, non-linear manifold algorithm that preserves the local structure of the data, while trying to maintain their global structure as well. Alternative approaches, such as the widely used t-SNE [52,53] or the LargeVis [54] algorithm, that are also capable of preserving locality, often fail to retain global information. In addition, they perform rather poorly on high-dimensional data, like the set of light spectra considered in this work, in terms of run time and scalability. For these reasons, UMAP was selected as the method of choice. Its most important parts, in particular those that ensure diversity of feature representations with regard to a subsequent clustering, will be presented in the following. For further details on the background and implementation of the UMAP algorithm, the interested reader is referred to the original paper of McInnes et al. [41].

The cost function of the UMAP approach is given by

$$C_{\text{UMAP}} = \sum_{i \neq j} \underbrace{v_{ij} \log \left(\frac{v_{ij}}{\omega_{ij}} \right)}_{\text{preserve locality}} + \underbrace{(1 - v_{ij}) \log \left(\frac{1 - v_{ij}}{1 - \omega_{ij}} \right)}_{\text{preserve globality}}, \quad (1)$$

where v_{ij} denotes a metric for assessing the similarity between two light spectra i and j in the high dimensional space X given by

$$v_{ij} = (v_{j|i} + v_{i|j}) - v_{j|i}v_{i|j}, \quad (2)$$

with

$$v_{j|i} = e^{\frac{-d(x_i, x_j) - \rho_i}{\sigma_i}}, \quad (3)$$

and $d(p, q) = \|q - p\|_2$ being the Euclidean distance, whereas ω_{ij} evaluates their similarity in the lower dimensional embedded space Y and reads

$$\omega_{ij} = (1 + a \|y_i - y_j\|_2^{2b})^{-1}. \quad (4)$$

Here, the first term of Equation (1), the so-called cross-entropy, preserves the local connectivity of the original data and brings similar light spectra closer together. The second term of the cost function, on the other hand, is intended to preserve the global structure of the data and moves dissimilar spectra further away from each other. This specific behavior of the UMAP cost function basically ensures that more diverse light spectra exhibit larger differences in the embedded space, even though, as stated previously, their absolute distance can no longer be used as a meaningful measure of true difference.

2.2.4. HDBSCAN

As the distance between vector representations of the input light spectra within the two-dimensional, maximally reduced embedded space are no longer meaningful after the application of the UMAP algorithm, the choice of clustering algorithms suitable for this kind of data is limited to density-based approaches that do not rely on the distance between individual data points. Due to its robustness against noise, the hierarchical density-based spatial clustering of applications with noise (HDBSCAN) [42] algorithm has been chosen for this work. In contrast to other density-based cluster approaches, such as the quite popular DBSCAN [55] algorithm, it performs much better on detecting variable density clusters without increasing computational run-time. Compared to DBSCAN, it further eliminates the need for manually selecting and fine-tuning distance scale parameters. As can be seen from Figure 3, HDBSCAN clusters the dense regions of the embedded space from UMAP and leaves the rest as noise. In preparation of the following sub-clustering process, all light spectra independent of whether they were assigned to a cluster or identified as noise were labelled accordingly.

2.3. Sub-Clustering Based on Lighting Parameters

The clustered raw data can be further processed by dividing each of the found main clusters (including the noise as a separate cluster) into representative sub-clusters based on a pre-defined selection of lighting parameters. If no further partitioning of the clusters is required on basis of the light parameters, this step can optionally be skipped. Otherwise, Table 1 gives an overview of the respective parameters considered in this work. However, it should be noted that any other parameter describing certain aspects of the properties of a light spectrum could also be used here. Thus, by including arbitrary lighting parameters, the individual sub-clusters can be arbitrarily biased, e.g., in such a way that they adhere to certain pre-defined photometric requirements.

The corresponding algorithmic pipeline is shown in Figure 5. All lighting parameters are calculated using the full spectral information of the input data so that small losses

in spectral accuracy observed during the dimensional reduction process are not further propagated to the sub-clusters. Due to the much lower dimensionality of the input data structure, no initial autoencoding step is required. Other than that, the algorithmic pipeline is exactly the same as before and results in a variable number of sub-clusters for each main cluster.

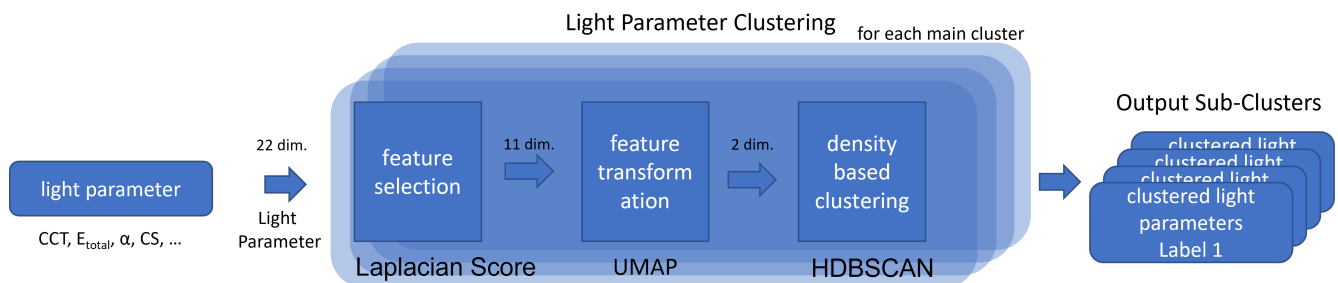


Figure 5. Algorithmic pipeline of the sub-clustering process based on a pre-defined selection of different lighting parameters.

Table 1. Selection of lighting parameters used for dividing each main cluster into respective sub-clusters.

Lighting Parameter	Description
CCT	correlated color temperature
E_{total}	total illuminance
D_{uv}	distance from the planckian locus for a given CCT in CIE 1960
E_{ind}	uniform chromaticity space
Q_a [56]	illuminance of the indirect part of the luminaire
ΔC	general color quality scale (NIST CQS 9.0)
CS [57]	average CIELAB chroma difference to reference illuminant
CL_A [57]	calculated for the 15 NIST CQS 9.0 color samples
α [58]	circadian stimulus
α_{opic} [31]	circadian light
E_e	E_{ind}/E_{total}
R_f [31]	integrated L-, M-, S-cone, rod, and ipRGC values
R_g [31]	irradiance
R_t [31]	color fidelity index
FCI [59]	color gamut Index
MCRI [60]	general metameric uncertainty index
CPI [61]	feeling of contrast index
	memory color rendition index
	color preference index

3. Results

3.1. Autoencoding and Main Clustering of the Embedded Space Representation

The training of the autoencoder was performed on the available GPU system. 60 million data were sampled uniformly from all possible light spectra of the test luminaire (see Section 2.1). All of these data were used for training. Afterwards, 1×10^5 spectra were randomly picked for evaluating the autoencoder performance. Taking approximately 1.5 days, the training of the autoencoder was completed after 400 epochs with a batch size of 1000, a learning rate of 1×10^{-6} , and a weight decay of 1×10^{-5} and resulted in an average mean squared error loss of 5.503×10^{-5} calculated on the evaluation data.

For illustrative purposes, Figure 6 correspondingly shows the autoencoder results for two light spectra arbitrarily chosen from the evaluation set, one of low and the other of high intensity. In both cases, the black solid line depicts the ground truth, while the blue solid line represents the reconstructed spectrum obtained from the decoding stage. The red graphs on the right hand side show the respective loss between the original spectrum and the reconstructed spectrum per unit wavelength. As can be seen, the absolute loss

is approximately the same for both low- and high-intensity spectra, which results in a considerably larger relative loss for the former compared to latter. This is reflected in pronounced ripples that can be observed on the blue curve in the upper left graph. Nevertheless, even for low intensities, it can be stated that the autoencoder is able to properly reconstruct the general trend in the data and, thus, provides sufficient accuracy for defining the main clusters. With regard to the density-based approach, it is sufficient that the reconstructed spectrum is close but not necessarily identical to the original input.

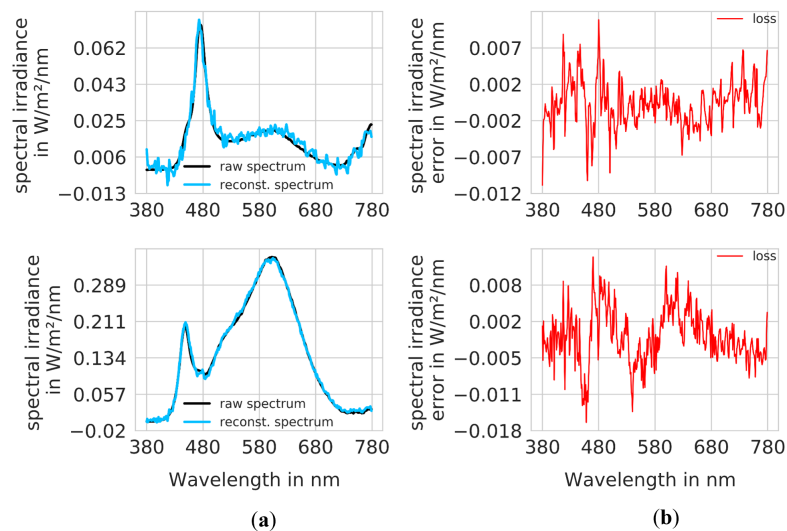


Figure 6. (a) Convolutional autoencoder input (black solid line) and reconstructed (blue solid line) light spectra after encoding and decoding the corresponding 1×15 dimensional feature vectors. (b) Loss between input and reconstructed spectra.

After the autoencoder step, the evaluation set, which basically comprised a representative, unbiased selection of 1×10^5 input test spectra, was further processed to provide a proof of concept of the algorithmic pipeline proposed in this work. Figure 7 thus shows the corresponding results of the UMAP embedding after autoencoding and Laplacian score based feature reduction. The light spectra are represented as data points in the two-dimensional embedded space, where, for illustrative purposes, the color and transparency coding correspond to their respective CCT and illuminance values. Here, the `min_distance` parameter was set to 0.0 in order to preserve the locality of the data and, by bringing similar data points closer together, create dense regions for the subsequent application of the density-based clustering algorithm. At this processing step, a total of 90 neighbors was checked for each data point and compared for similarity to also take into account the global structure of the data. Note that as the number of neighbors that are checked increases, more weight is put on the latter. As a downside, this however also increases computational time.

From Figure 7, it can be seen that the proposed algorithmic pipeline, starting from purely unbiased spectral data, is capable of arranging their embedded space representation in such a way that similar light spectra, here in terms of CCT and illuminance, are closer together. Spectra of lower CCT can for example be found close to the outer-contour of the embedded body structure constituted by the data points, while towards the more central parts the CCT increases. It should again be noted that the absolute distance between data points is no longer a meaningful measure of similarity or difference.

This example confirms that the proposed algorithmic pipeline as discussed in Section 2.2 gives meaningful and interpretable results. It shows that 8 features (after autoencoding and Laplacian score filtering) are sufficient for encoding the most relevant information of the input light spectra. Using more input data, the shape of the embedded body structure will essentially remain the same and only the gaps will be filled. If another luminaire with different LED primaries is used, a different shape of the resulting embedded body structure

can be expected. With the proposed pipeline, an arbitrary number of LED primaries can be processed. Combinations of other light sources can be included as well.

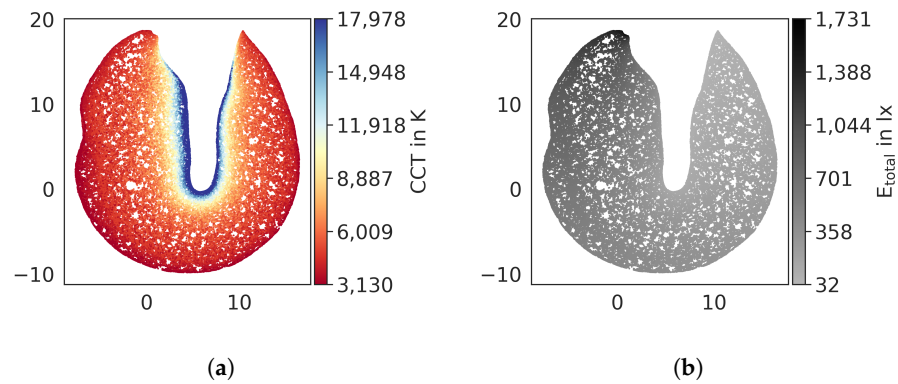


Figure 7. Results of the UMAP algorithm applied to the test data set of 1×10^5 randomly selected input spectra of the test luminaire discussed in Section 2.1 after autoencoding and Laplacian score feature reduction. The color coding denotes different CCTs (a) and the grayscale in (b) indicates the corresponding illuminance E_{total} .

Figure 8 finally shows the output of the density-based clustering of the embedded UMAP space using HDBSCAN. Here, a minimum_cluster_size of 100 was chosen to also consider smaller groups of similar spectra. In order to prevent, at the same time, the formation of too many of such small clusters, an epsilon of 0.1, was used for counterbalancing purposes. The epsilon parameter ensures that neighboring clusters are not partitioned any further if the distance is less or equal to epsilon, resulting in bigger clusters in case that a sufficient degree of similarity is achieved. In addition, the clustering was performed with a minimum_samples parameter of 90, which results in a more conservative clustering with a considerably larger number of data points being identified as noise than when choosing a smaller value.

In this example, 69 different main clusters have been found. As discussed before, these clusters are most diverse to each other in terms of their spectral features. As can be seen, they primarily cover the contour of the embedded body structure. Towards the more central parts, the density of neighboring data points in UMAP space decreases so that only smaller clusters could be identified in these regions.

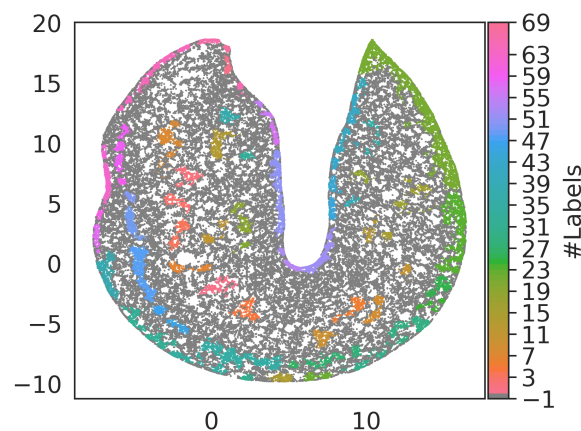


Figure 8. Results of the density-based clustering process of the 1×10^5 test spectra using the HDBSCAN algorithm. The different colors indicate the respective cluster assignment, while gray data points represent identified noise.

Figure 9 illustrates how the physical diversity of the different clusters with regard to the spectral features of the input light spectra transmits to the corresponding lighting parameters. Note that for the sake of brevity, only a smaller selection of the lighting parameters discussed in Section 2.3 is shown here. As can be seen, there are some parameters, like the illuminance E_{total} and the circadian stimulus CS, that show a large diversity between different clusters, while for others, such as the direct/indirect parameter α , a considerable overlap between the clusters still persists. Thus, it can be concluded that the unbiased physical clustering of light spectra as proposed in this work yields a superior diversifying strategy that is not confounded or affected by some potentially limiting pre-selection of lighting parameters. Instead, the choice for the most diverse light spectra is solely based on their spectral main features.

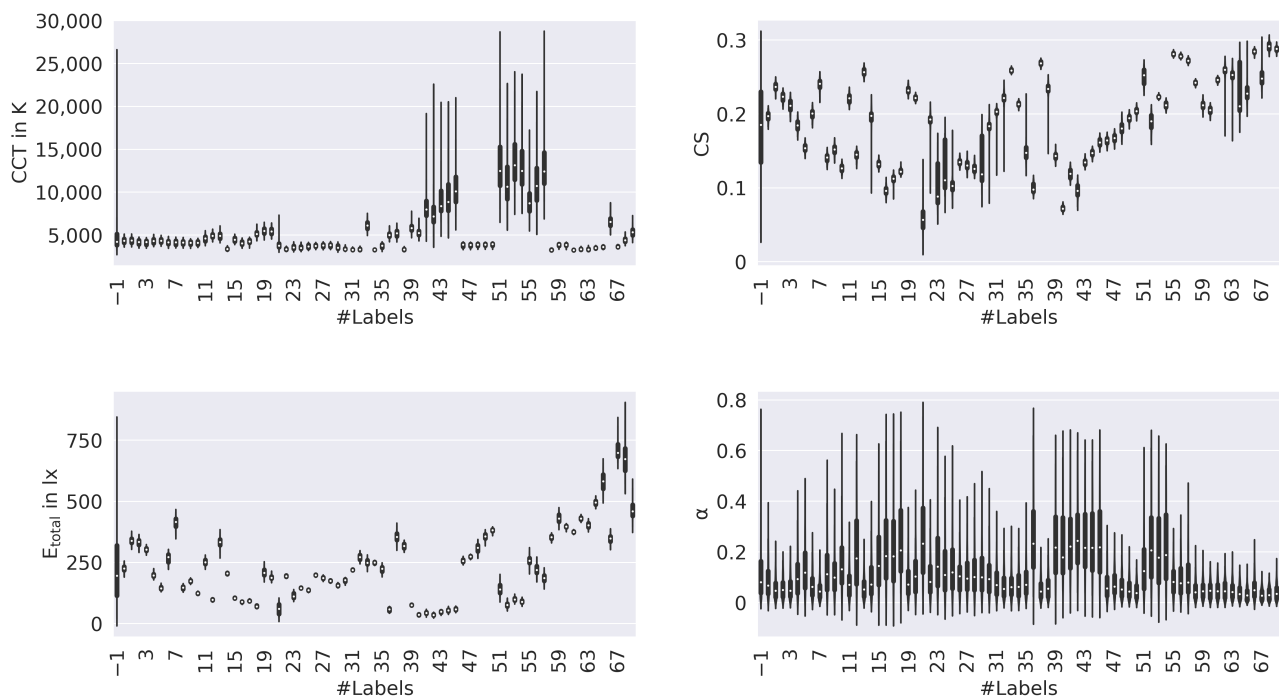


Figure 9. Result of the physical clustering of the 1×10^5 input test spectra for a selection of different lighting parameters. The labels on the x -axis denote the different main clusters.

3.2. Sub-Clustering and Data Filtering

Figure 10 eventually illustrates the partitioning of an arbitrary main cluster into its respective sub-clusters based on the pre-selection of different lighting parameters discussed in Section 2.3. After Laplacian feature reduction, only the 11 most relevant lighting parameters were used for UMAP and subsequent HDBSCAN processing (see Figure 5). On this basis, the algorithmic pipeline is capable of finding within each main cluster distinct groups of light spectra that are more similar to each other in terms of their lighting parameters. In the example shown here, again, only for a subset of lighting parameters, the main cluster was divided into eight sub-clusters labelled from 0 to 7 and one noise cluster labelled by -1 . Note that the number of sub-clusters may vary between different main clusters.

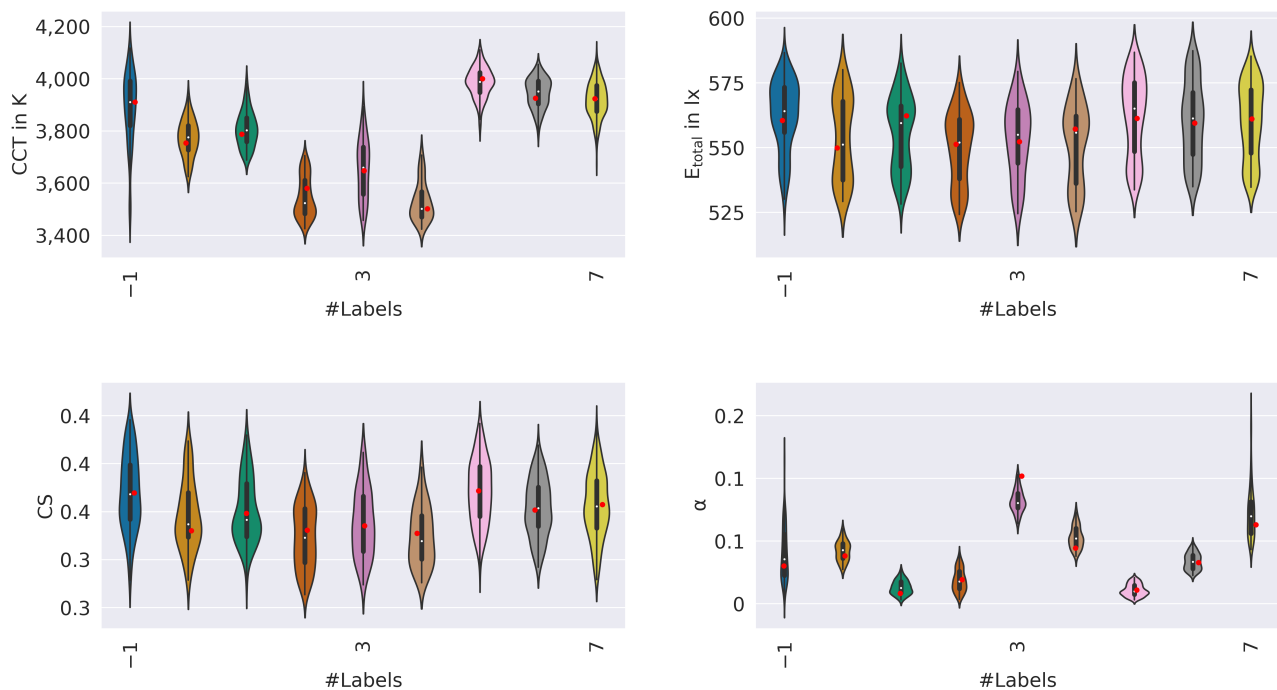


Figure 10. Partitioning of an arbitrary main cluster obtained from the first processing stage into its respective sub-clusters. The plots show the results for a subset of lighting parameters. The labels from 0 to 7 represent the different sub-clusters, whereas a label of -1 denotes the noise cluster.

As can be seen, the sub-clusters of the present example mainly differ in the direct/indirect α parameter of the assigned light spectra. The other lighting parameters depicted here show a similar range and a large overlap across the different sub-clusters. This basically emphasizes the large degree of similarity of the light spectra assigned to the same main cluster. A further differentiation between the spectra can usually be achieved only with regard to very few lighting parameters, where the importance of these parameters may differ between different main clusters.

Nevertheless, the application of UMAP and HDBSCAN to each main cluster again yields selections of light spectra assigned to different sub-clusters that by definition are most diverse to each other. This assignment, as it is based on lighting parameters, can now be used to filter out light spectra that, on top of their diversity, are also in a certain parameter range or match a specific target definition. Examples for such target definitions could be a certain illuminance range or a specific correlated color temperature defined as a prerequisite for running subsequent subject studies. For illustrative purposes, the target value of the present work was to extract one spectrum out of each sub-cluster that best fits the corresponding medians for each considered lighting parameter in terms of an overall accumulating distance metric. Note that this specific example serves only as a proof of concept. Any other selection criterion could also be used here.

Figure 11, finally, shows the accordingly filtered light spectra, one for each sub-cluster, as obtained for two different main clusters. As can be seen, the shape and intensities of the SPDs are quite similar for the same main cluster but vary significantly between different main clusters, confirming the suitability and performance of the data-based diversifying strategy proposed in this work. In addition, Figure 12 depicts the parameter representations of all likewise selected light spectra categorized according to their main cluster assignment. Again only a subset of lighting parameters is shown. In most cases, parameter deviations between the selected light spectra within a given main cluster are small compared to the

between-cluster differences. Overall, the main clusters cover a wide range of lighting parameter values, representing the diversity between them.

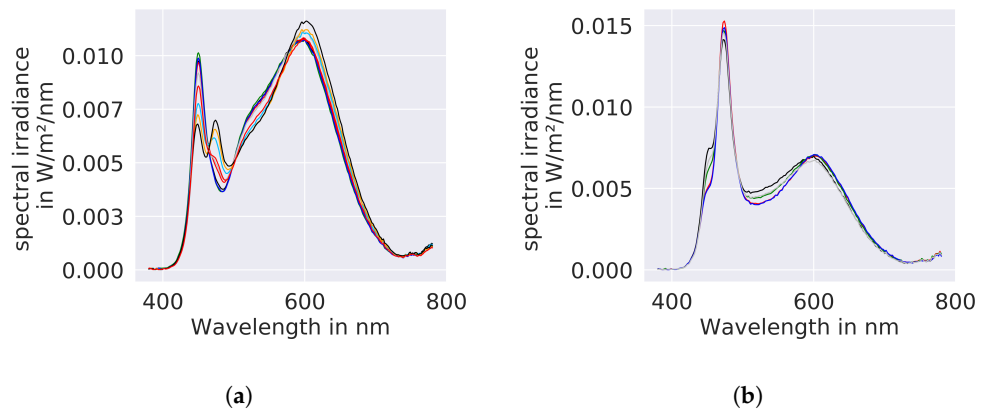


Figure 11. Light spectra selected from each sub-cluster of two different main clusters (a) and (b) by applying a pre-defined median-based filter metric for proof of concept.

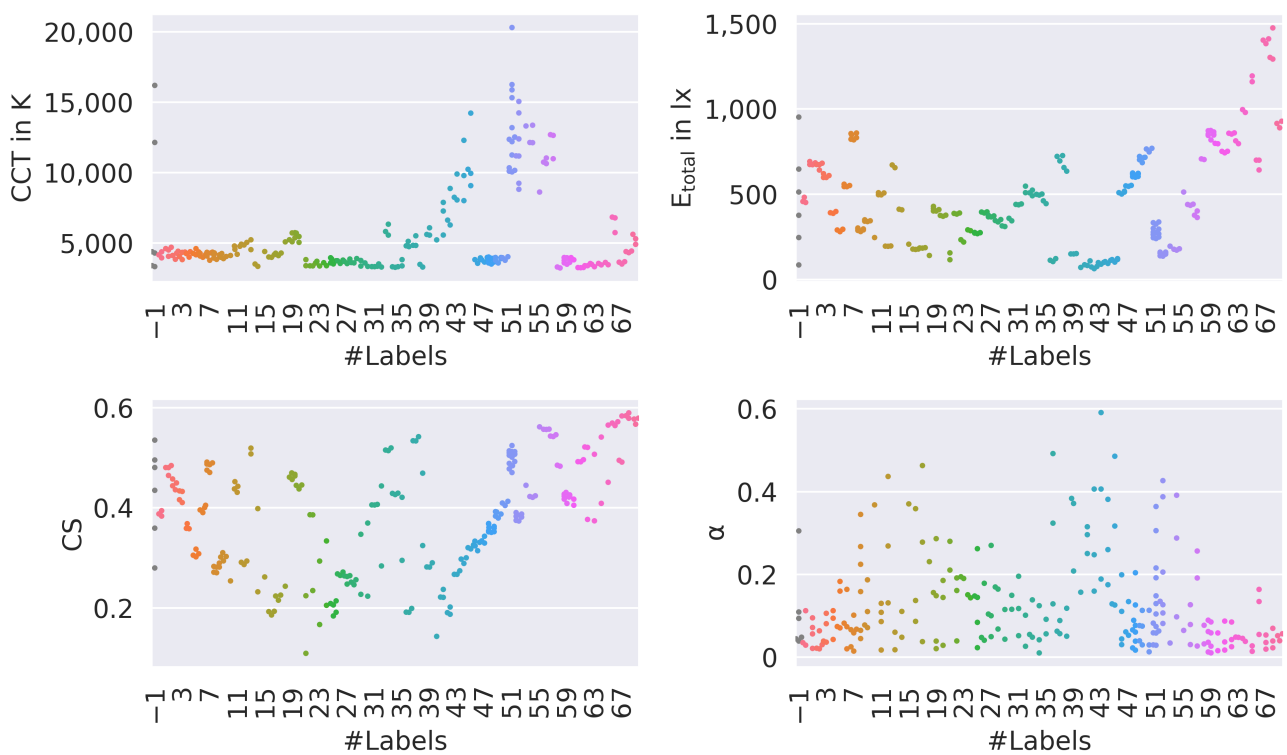


Figure 12. Parameter representations of the light spectra filtered out of each sub-cluster for a subset of lighting parameters. Labels on the x -axis represent the corresponding main cluster assignment.

4. Discussion

In this work, a novel unsupervised clustering pipeline has been proposed for lighting research applications. It basically provides a generalized, data-driven framework of sampling diversified light spectra composed of a certain number of primaries to be used for lighting preference studies and correlation analyses. To the authors' best knowledge, this is the first approach ever that, in contrast to the criteria-based manual selection of light spectra known from existing literature, makes use of state-of-the-art feature reduction and clustering algorithms in order to select the most distinct light spectra from a given larger

set of possible candidates, while still having the option of adhering to certain pre-defined photometric and colorimetric requirements. Summarizing the individual processing steps discussed in detail in the previous sections, Figure 13 eventually gives an overview of the complete algorithmic pipeline of the proposed approach. Starting from unbiased, purely physical features, this orchestrated assembly of algorithms can be used to find representative clusters of most diverse light spectra from an arbitrary set of spectral input data. Optionally, the respective processing pipeline additionally allows for sub-clustering and data filtering based on some pre-selected lighting parameters identified as relevant for the specific research question at hand. Furthermore, the two-dimensional cluster representations provided by applying the UMAP and HDBSCAN algorithms may help to visualize and detect local correlations in the data.

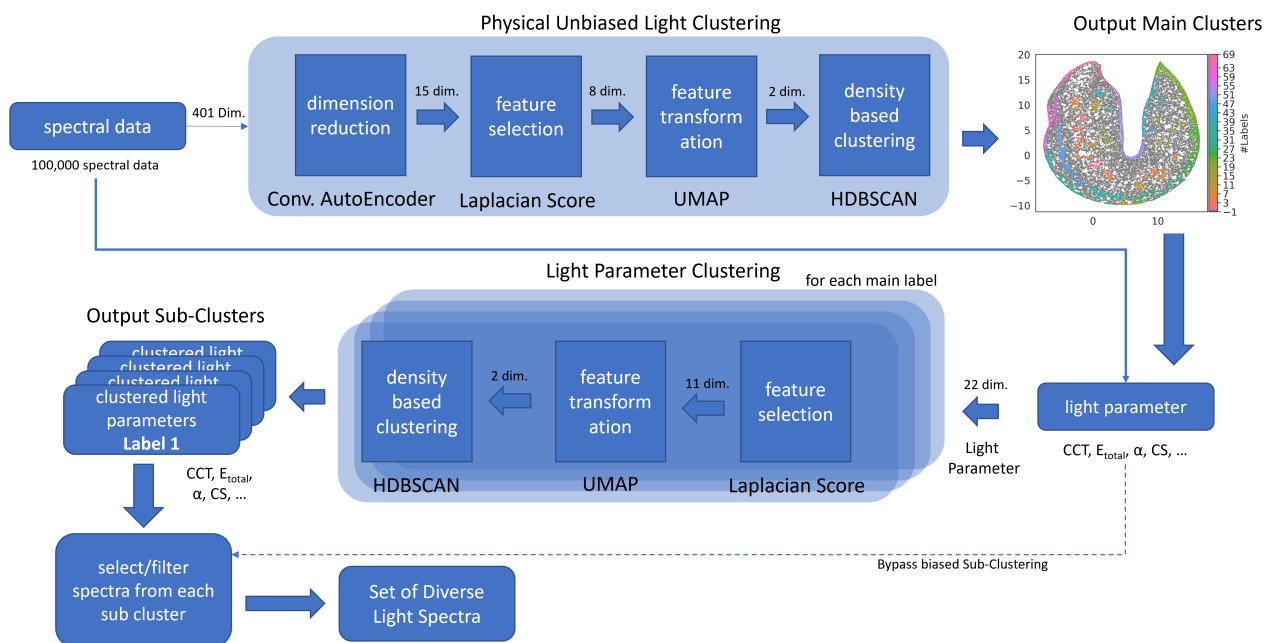


Figure 13. Complete algorithmic pipeline of the proposed light-spectra-diversifying strategy.

Throughout this work, it has been shown that, by applying a convolutional autocencoder to pre-process the spectral input data, the physical information contained within the 400 wavelength bins between 380 nm and 799 nm can efficiently be captured with sufficient accuracy in a 15-dimensional reduced space, where the corresponding MSE loss was 5.5×10^{-5} . Subsequent feature selection and transformation lead to a further dimensional reduction and an embedded space representation of the relevant spectral information taking into account both the local similarity and the global structure of the data. As an example, it has been shown the resulting embedded space can meaningfully be interpreted with respect to the CCT and illuminance distributions of the input light spectra. A density-based clustering approach is able to identify similar light spectra and assign them to corresponding clusters, which by definition are most diverse to each other and, thus, can be used to sample light spectra of maximal diversity. Thus, a sampling of diversified light spectra can be based on selecting candidates from each of these distinct main clusters. Optional sub-clustering can be applied to achieve a further diversification of light spectra within each of the main clusters. A minimal working example of the algorithmic implementation has been prepared and made available for download (<https://github.com/KlirS/MDPI-Diversified-Light-Spectra>, accessed on 24 September 2021).

It should be noted that, although a DALI-controlled lighting device was chosen to be used for validation purposes, the proposed methodology constitutes a universal approach in such a way that it is solely based on physical (i.e., spectral) input data, which basically makes it independent of any chosen electrical dimming strategy. Moreover, it is not limited by the complexity of the spectral input data. A higher number of LED primaries

only increases the number of theoretically composable light spectra but other than that has no direct impact on the performance of the proposed algorithmic pipeline. Yet, it is certainly true that a large amount of input data ($>1 \times 10^6$) may increase the required computational time tremendously. On the given system a calculation was measured and took for 1×10^5 data approximately 31 min. For 1×10^6 data, the time is approximately 384 min. This corresponds to an increase in time by a factor of 12 for an increase in data by a factor of 10. Further optimizations with parallelized calculations can be implemented in future work to overcome the trade-off between the number of spectra considered and the calculation time. Even though such considerations are beyond the scope of the present paper, it should, therefore, be emphasized that future work must focus on the development of expedient methods to pre-filter the spectral input data for relevance before entering the diversifying pipeline, especially as the number of LED primaries used in modern lighting systems and, consequently, the amount of possible light settings per luminaire steadily increases. A further downside of the pipeline's current implementation is that the hyperparameters of the UMAP and HDBSCAN algorithms still need to be manually adjusted to the input data for an optimal performance. Hence, future research intentions will focus on the development and exploration of new methods to estimate an optimal set of these hyperparameters from the underlying or expected structure of the spectral input data. Finally, it should be noted that the proposed diversifying strategy, although tested on a DALI-controlled lighting device only, is universal in nature and, therefore, independent of the specific protocol used for luminaire control.

Author Contributions: Conceptualization, S.K. and R.F.; Data curation, S.K. and R.F.; Formal analysis, S.K. and R.F.; Methodology, S.K., S.B. (Sebastian Babilon) and S.B. (Simon Benkner); Software, S.K. and R.F.; Supervision, T.Q.K.; Validation, S.K., S.B. (Sebastian Babilon) and S.B. (Simon Benkner); Visualization, S.K.; Writing—original draft, S.K.; Writing—review and editing, S.K., S.B. (Sebastian Babilon), S.B. (Simon Benkner) and T.Q.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under grant No. 445336968 and by the Open Access Publishing Fund of the Technical University of Darmstadt. S.B. (Sebastian Babilon & Simon Benkner) received personal funding by the Ernst Ludwig Mobility Grant of the Technical University of Darmstadt.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data generated or analyzed to support the findings of the present study are included in this article. A minimal example to generate the presented data can be found on <https://github.com/KlirS/MDPI-Diversified-Light-Spectra>, accessed on 24 September 2021.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Pust, P.; Schmidt, P.J.; Schnick, W. A revolution in lighting. *Nat. Mater.* **2015**, *14*, 454–458. [[CrossRef](#)] [[PubMed](#)]
2. Khanh, T.Q.; Bodrogi, P.; Guo, X.; Vinh, Q.T.; Fischer, S. Colour Preference, Naturalness, Vividness and Colour Quality Metrics, Part 5: A Colour Preference Experiment at 2000 lx in a Real Room. *Light. Res. Technol.* **2019**, *51*, 262–279. [[CrossRef](#)]
3. Bodrogi, P.; Guo, X.; Stojanovic, D.; Fischer, S.; Khanh, T.Q. Observer preference for perceived illumination chromaticity. *Color Res. Appl.* **2018**, *43*, 506–516. [[CrossRef](#)]
4. Tähkämö, L.; Partonen, T.; Pesonen, A.K. Systematic review of light exposure impact on human circadian rhythm. *Chronobiol. Int.* **2019**, *36*, 151–170. [[CrossRef](#)] [[PubMed](#)]
5. Babilon, S.; Lenz, J.; Beck, S.; Myland, P.; Klages, J.; Klir, S.; Khanh, T.Q. Task-related Luminance Distributions for Office Lighting Scenarios. *Light Eng.* **2021**, *29*, 115–128. [[CrossRef](#)]
6. Yan, L.; Chen, Y.; Chen, B. Integrated analog dimming controller for 0–10 V dimming system. In Proceedings of the 10th China International Forum on Solid State Lighting (ChinaSSL), Beijing, China, 10–12 November 2013; Institute of Electrical and Electronics Engineers (IEEE): Beijing, China, 2013; pp. 147–149. [[CrossRef](#)]

7. Gagliardi, G.; Casavola, A.; Lupia, M.; Cario, G.; Tedesco, F.; Lo Scudo, F.; Cicchello Gaccio, F.; Augimeri, A. A smart city adaptive lighting system. In Proceedings of the Third International Conference on Fog and Mobile Edge Computing (FMEC), Barcelona, Spain, 23–26 April 2018; Institute of Electrical and Electronics Engineers (IEEE): Barcelona, Spain, 2018; pp. 258–263. [CrossRef]
8. Gagliardi, G.; Lupia, M.; Cario, G.; Tedesco, F.; Cicchello Gaccio, F.; Lo Scudo, F.; Casavola, A. Advanced adaptive street lighting systems for smart cities. *Smart Cities* **2020**, *3*, 1495–1512. [CrossRef]
9. Sinha, A.; Sharma, S.; Goswami, P.; Verma, V.K.; Manas, M. Design of an energy efficient IoT enabled smart system based on DALI network over MQTT protocol. In Proceedings of the 3rd International Conference on Computational Intelligence & Communication Technology (CICT), Ghaziabad, India, 9–10 February 2017; Institute of Electrical and Electronics Engineers (IEEE): Ghaziabad, India, 2017. [CrossRef]
10. Sikder, A.K.; Acar, A.; Aksu, H.; Uluagac, A.S.; Akkaya, K.; Conti, M. IoT-enabled smart lighting systems for smart cities. In Proceedings of the 8th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 8–10 January 2018; Institute of Electrical and Electronics Engineers (IEEE): Las Vegas, NV, USA, 2018; pp. 639–645. [CrossRef]
11. Adam, G.K. DALI LED driver control system for lighting operations based on Raspberry Pi and kernel modules. *Electronics* **2019**, *8*, 1021. [CrossRef]
12. Kaleem, Z.; Ahmad, I.; Lee, C. Smart and energy efficient LED street light control system using ZigBee network. In Proceedings of the 12th International Conference on Frontiers of Information Technology, Islamabad, Pakistan, 17–19 December 2014; Institute of Electrical and Electronics Engineers (IEEE): Islamabad, Pakistan, 2014; pp. 361–365. [CrossRef]
13. Varghese, S.G.; Kurian, C.P.; George, V.I.; John, A.; Nayak, V.; Upadhyay, A. Comparative study of ZigBee topologies for IoT-based lighting automation. *IET Wirel. Sens. Syst.* **2019**, *9*, 201–207. [CrossRef]
14. Wei, W.; Min, W. The design of ZigBee routing algorithm in smart lighting system. *Ferroelectrics* **2019**, *549*, 254–265. [CrossRef]
15. IEC International Electrotechnical Commission. IEC 60929:2011—AC and/or DC-Supplied Electronic Control Gear for Tubular Fluorescent Lamps: Performance Requirements. 2011. Available online: <https://webstore.iec.ch/publication/3926> (accessed on Sept 24th, 2021).
16. IEC International Electrotechnical Commission. IEC 62386-207:2018—Digital Addressable Lighting Interface—Part 207: Particular Requirements for Control Gear—LED Modules (Device Type 6). 2018. Available online: <https://webstore.iec.ch/publication/30618> (accessed on 24 September 2021).
17. Zandi, B.; Eissfeldt, A.; Herzog, A.; Khanh, T.Q. Melanopic limits of metamer spectral optimisation in multi-channel smart lighting systems. *Energies* **2021**, *14*, 527. [CrossRef]
18. Schweitzer, S.; Schinagl, C.; Djuras, G.; Frühwirth, M.; Hoschopf, H.; Wagner, F.; Schulz, B.; Nemitz, W.; Grote, V.; Reidl, S.; et al. Investigation of gender- and age-related preferences of men and women regarding lighting conditions for activation and relaxation. In Proceedings of the Fifteenth International Conference on Solid State Lighting and LED-based Illumination Systems, San Diego, CA, USA, 28 August–1 September 2016; Volume 9954, p. 99540. [CrossRef]
19. Despenic, M.; Chraibi, S.; Lashina, T.; Rosemann, A. Lighting preference profiles of users in an open office environment. *Build. Environ.* **2017**, *116*, 89–107. [CrossRef]
20. Chraibi, S.; Lashina, T.; Shrubsole, P.; Aries, M.; van Loenen, E.; Rosemann, A. Satisfying light conditions: A field study on perception of consensus light in Dutch open office environments. *Build. Environ.* **2016**, *105*, 116–127. [CrossRef]
21. Zhu, Y.; Yang, M.; Yao, Y.; Xiong, X.; Li, X.; Zhou, G.; Ma, N. Effects of Illuminance and Correlated Color Temperature on Daytime Cognitive Performance, Subjective Mood, and Alertness in Healthy Adults. *Environ. Behav.* **2019**, *51*, 199–230. [CrossRef]
22. Finlayson, G.; Mackiewicz, M.; Hurlbert, A.; Pearce, B.; Crichton, S. On calculating metamer sets for spectrally tunable LED illuminators. *J. Opt. Soc. Am. A* **2014**, *31*, 1577. [CrossRef]
23. Allen, A.E.; Hazelhoff, E.M.; Martial, F.P.; Cajochen, C.; Lucas, R.J. Exploiting metamerism to regulate the impact of a visual display on alertness and melatonin suppression independent of visual appearance. *Sleep* **2018**, *41*. [CrossRef]
24. Jain, A.K.; Murty, M.N.; Flynn, P.J. Data clustering: A review. *ACM Comput. Surv.* **1999**, *31*, 264–323. [CrossRef]
25. Huang, H.C.; Chuang, Y.Y.; Chen, C.S. Affinity aggregation for spectral clustering. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 773–780. [CrossRef]
26. Singh, S.; Gupta, A.; Efros, A.A. Unsupervised discovery of mid-level discriminative patches. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Proceedings of the 12th European Conference on Computer Vision, Florence, Italy, 7–13 October 2012*; Springer: Berlin/Heidelberg, 2012; Volume 7573 LNCS, pp. 73–86. [CrossRef]
27. Hariharan, B.; Malik, J.; Ramanan, D. Discriminative decorrelation for clustering and classification. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Proceedings of the 12th European Conference on Computer Vision, Florence, Italy, 7–13 October 2012*; Springer: Berlin/Heidelberg, 2012; Volume 7575 LNCS, pp. 459–472. [CrossRef]
28. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
29. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
30. Sermanet, P.; Eigen, D.; Zhang, X.; Mathieu, M.; Fergus, R.; LeCun, Y. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv* **2013**, arXiv:1312.6229.

31. IES-TM-30-15. *Method for Evaluating Light Source Color Rendition*; Illuminating Engineering Society of North America: New York, NY, USA, 2015.
32. Lucas, R.J.; Peirson, S.N.; Berson, D.M.; Brown, T.M.; Cooper, H.M.; Czeisler, C.A.; Figueiro, M.G.; Gamlin, P.D.; Lockley, S.W.; O'Hagan, J.B.; et al. Measuring and using light in the melanopsin age. *Trends Neurosci.* **2014**, *37*, 1–9. [[CrossRef](#)]
33. Celebi, M.E.; Aydin, K. *Unsupervised Learning Algorithms*; Springer International Publishing: Cham, Switzerland, 2016. [[CrossRef](#)]
34. Kramer, M.A. Nonlinear principal component analysis using autoassociative neural networks. *AIChE J.* **1991**, *37*, 233–243. [[CrossRef](#)]
35. Wold, S.; Esbensen, K.; Geladi, P. Principal Component Analysis. *Chemometr. Intell. Lab. Syst.* **1987**, *2*, 37–52. [[CrossRef](#)]
36. McConville, R.; Santos-Rodríguez, R.; Piechocki, R.J.; Craddock, I. N2D: (Not too) deep clustering via clustering the local manifold of an autoencoded embedding. *arXiv* **2019**, arXiv:1908.05968.
37. Amarbayasgalan, T.; Jargalsaikhan, B.; Ryu, K.H. Unsupervised novelty detection using deep autoencoders with density based clustering. *Appl. Sci.* **2018**, *8*, 1468. [[CrossRef](#)]
38. Zaidi, F.H.; Hull, J.T.; Peirson, S.N.; Wulff, K.; Aeschbach, D.; Gooley, J.J.; Brainard, G.C.; Gregory-Evans, K.; Rizzo, J.F.; Czeisler, C.A.; et al. Short-Wavelength Light Sensitivity of Circadian, Pupillary, and Visual Awareness in Humans Lacking an Outer Retina. *Curr. Biol.* **2007**, *17*, 2122–2128. [[CrossRef](#)] [[PubMed](#)]
39. IEC International Electrotechnical Commission. IEC 62386-102:2009—Digital Addressable Lighting Interface—Part 102: General Requirements—Control Gear. 2009. Available online: <https://webstore.iec.ch/publication/20477> (accessed on 24 September 2021).
40. He, X.; Cai, D.; Niyogi, P. Laplacian Score for feature selection. *Adv. Neural Inf. Process. Syst.* **2005**, *18*, 507–514.
41. McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv* **2018**, arXiv:1802.03426.
42. McInnes, L.; Healy, J. Accelerated Hierarchical Density Based Clustering. In Proceedings of the IEEE International Conference on Data Mining Workshops, ICDMW, New Orleans, LA, USA, 18–21 November 2017; pp. 33–42. [[CrossRef](#)]
43. Mousavi, S.M.; Zhu, W.; Ellsworth, W.; Beroza, G. Unsupervised Clustering of Seismic Signals Using Deep Convolutional Autoencoders. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1693–1697. [[CrossRef](#)]
44. Lee, K.; Carlberg, K.T. Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders. *J. Comput. Phys.* **2020**, *404*, 108973. [[CrossRef](#)]
45. Wang, S.; Chen, H.; Wu, L.; Wang, J. A novel smart meter data compression method via stacked convolutional sparse auto-encoder. *Int. J. Electr. Power Energy Syst.* **2020**, *118*, 105761. [[CrossRef](#)]
46. Wang, Y.; Yao, H.; Zhao, S. Auto-encoder based dimensionality reduction. *Neurocomputing* **2016**, *184*, 232–242. [[CrossRef](#)]
47. Masci, J.; Meier, U.; Cireşan, D.; Schmidhuber, J. Stacked convolutional auto-encoders for hierarchical feature extraction. In *Artificial Neural Networks and Machine Learning—ICANN 2011*; Honkela, T., Duch, W., Girolami, M., Kaski, S., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 52–59. [[CrossRef](#)]
48. Gholamalinezhad, H.; Khosravi, H. Pooling Methods in Deep Neural Networks, a Review. *arXiv* **2020**, arXiv:2009.07485.
49. Li, J.; Cheng, K.; Wang, S.; Morstatter, F.; Trevino, R.P.; Tang, J.; Liu, H. Feature selection: A data perspective. *ACM Comput. Surv. (CSUR)* **2018**, *50*, 94. [[CrossRef](#)]
50. He, X.; Niyogi, P. Locality preserving projections. *Adv. Neural Inf. Process. Syst.* **2004**, *6*, 153–160.
51. Belkin, M.; Niyogi, P. Laplacian eigenmaps and spectral techniques for embedding and clustering. *Adv. Neural Inf. Process. Syst.* **2002**, *14*, 585–591. [[CrossRef](#)]
52. van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2625.
53. van der Maaten, L. Accelerating t-SNE using Tree-Based Algorithms. *J. Mach. Learn. Res.* **2014**, *15*, 3221–3245.
54. Tang, J.; Liu, J.; Zhang, M.; Mei, Q. Visualizing large-scale and high-dimensional data. In Proceedings of the 25th International Conference on World Wide Web, Montréal, QC, Canada, 11–15 April 2016; International World Wide Web Conferences Steering Committee: Geneva, Switzerland, 2016; pp. 287–297. [[CrossRef](#)]
55. Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR, USA, 2–4 August 1996; pp. 226–231.
56. Davis, W. Color quality scale. *Opt. Eng.* **2010**, *49*, 033602. [[CrossRef](#)]
57. Rea, M.S.; Figueiro, M.G. Light as a circadian stimulus for architectural lighting. *Light. Res. Technol.* **2018**, *50*, 497–510. [[CrossRef](#)]
58. Houser, K.W.; Tiller, D.K.; Bernecker, C.A.; Mistrick, R.G. The subjective response to linear fluorescent direct/indirect lighting systems. *Light. Res. Technol.* **2002**, *34*, 243–260. [[CrossRef](#)]
59. Hashimoto, K.; Yano, T.; Shimizu, M.; Nayatani, Y. New method for specifying color-rendering properties of light sources based on feeling of contrast. *Color Res. Appl.* **2007**, *32*, 361–371. [[CrossRef](#)]
60. Smet, K.A.; Ryckaert, W.R.; Pointer, M.R.; Deconinck, G.; Hanselaer, P. A memory colour quality metric for white light sources. *Energy Build.* **2012**, *49*, 216–225. [[CrossRef](#)]
61. Thornton, W.A. A validation of the color-preference index. *J. Illum. Eng. Soc.* **1974**, *4*, 48–52. [[CrossRef](#)]