Mapping and Visualizing Deep-Learning Urban Beautification

Tobias Kauer
University of Applied Science Potsdam

Sagar Joglekar
King's College London

Miriam Redi, Luca Maria Aiello, and Daniele Quercia
Nokia Bell Labs Cambridge

Information visualization has great potential to make sense of the increasing amount of data generated by complex machine-learning algorithms. We design a set of visualizations for a new deep-learning algorithm called FaceLift (goodcitylife.org/facelift). This algorithm is able to generate a beautified version of a given urban image (such as from Google Street View), and our visualizations compare pairs of original and beautified images. With those visualizations, we aim at helping practitioners understand what happened during the algorithmic beautification without requiring them to be machine-learning experts. We evaluate the effectiveness of our visualizations to do just that with a survey among practitioners. From the survey results, we derive general design guidelines on how information visualization makes complex machine-learning algorithms more understandable to a general audience.

Beautiful places make us feel better. Stendhal's motto "beauty is the promise of happiness" speaks to this and has been made use of in various studies to show that specific visual cues affect our well-being.[1,2] But what are these visual cues of beauty? Our public realm is filled with examples people perceive as beautiful. Regardless of whether beauty emerges from planning or serendipitously, we can identify its cues, and that is useful for supporting evidence-based design of the urban spaces we intuitively love.

Based on that premise, here we present a design study that visualizes data-intensive and complex results stemming from a range of deep neural networks. These generative adversarial networks (GANs) beautify a Google Street View scene according to a trained concept of beauty.[3] To validate the beautification process, we compare the original image and the beautified one in terms of the elements that have been added or removed. These elements are then mapped into urban design metrics that the urban design literature has identified to characterize great urban spaces.[4] In

Published by the IEEE Computer Society
0272-1716/18/$33.00 ©2018 IEEE

our study, we use beautification data from Boston, Massachusetts. The visualizations are designed to support different angles of insight into the process of machine-learning-based (ML-based) urban beautification. In so doing, we make three main contributions:

- We visualize the output of the deep-learning networks in the form of heat maps that identify existing spatial clusters of beauty in the city and provide an overview of beautifiable locations.
- We visualize the beautification process by showing the image before beautification (current street view) and after it (automatically beautified version of the street view). We compare the pair of images by highlighting how they differ in terms of desirable urban design metrics. These metrics make it possible to explore changes in an individual pair of street scenes, as well to identify patterns of change across the whole dataset.
- We use an online expert survey to quantitatively validate the ability of the visualizations to (a) create reproducible interpretations of the visualized data and (b) inspire urban planners to consider the approach for future application. From the survey results, we find a considerable increase both in data literacy and in facilitating data-driven decision-making in urban planning.

RELATED WORK

In recent years, several research projects used artificial intelligence in the urban-planning context. As Liu *et al.*[5] observed, ML models can produce a medium-to-good estimate of people's real urban experiences, which could be used by researchers and planners. ML was also employed by Koenig *et al.*[6] to augment existing manual urban design strategies with computational design support systems.

Other projects employed maps to present their results. Naik *et al.*,[7] for example, provided support for classical theories of urban change and illustrated the value of using computer-vision methods and street-level imagery to understand the physical dynamics of cities. Beyond the algorithmic comparison of two images, they provided an interactive map to visually compare the images and to identify how geographic census tracts changed over time. A similar approach was chosen by Seiferling *et al.*[8] who also used computer vision on Google Street View images. They quantified urban tree coverage at the street level, and the results were made available in the form of an interactive online map that uses dot-density visualization. Along small multiples of the dot-density maps, there are indices to help viewers compare different cities.

Kachkaev *et al.*[9] introduced a measure of attractiveness to a body of crowdsourced geo-located images to automatically plan beautiful routes for leisure walks through central London. This study is relevant, as it aims at identifying walkable routes based on computer-vision methods that extract the amount of green pixels in images, which relates to two out of five metrics in our own study.

To summarize the relevant literature, while past research projects did a great job in using state-of-the-art ML to gain knowledge about cities, they did not visually make this knowledge accessible to urban planners and architects. If mapping applications were part of these approaches, they were not designed to convey applicable insights, and their potential to further facilitate decision-making in urban planning was not evaluated.

DATA

Our underlying framework uses GAN[10] to generate images of beautiful urban scenes. It is based on a model that learned from a dataset of images of urban spaces annotated according to their aesthetic appeal. The data we extract from the framework is composed of three elements: (1) geo-located images (Google Street Views) with associated beauty scores, (2) beautified versions of these images, and (3) selected urban evaluation metrics and attributes explaining the process. The selected urban metrics are walkability, visual complexity, openness, green spaces, and landmarks.

Urban Images with Beauty Scores

To train the algorithm, we need a large set of annotated data. Despite the highly subjective nature of intangible properties like beauty, there is a previous study that curated a dataset called StreetScore in which a large and diverse "crowd" of Internet users annotated images (Google Street Views) on different subjective scales, including beauty.[11] We use this data for training the network. The StreetScore dataset was built by showing participants pairs of urban images and asking them to choose the more beautiful one, resulting in votes for roughly a million images. We transform these votes into ordinal ranks using a Bayesian algorithm called TrueSkill.[12] The images are transformed on an ordinal scale of TrueSkill scores, which range from 0 to 40.

Beautified Images

Using these beauty scores,[6] we train a deep-learning network to classify urban images into one of two classes: beautiful and ugly.

Once the classifier learns the concept of beauty, the framework uses GAN to generate template images of beautified urban scenes.[10] This is done by modifying the pipeline by Nguyen *et al*.[13] to maximize beauty. This process takes in a preferably ugly image and generates images that maximize beauty. The resulting generated images are called template images, which are then used as retrieval templates. The retrieval process matches the template image with an existing Google Street View image using a similarity function. In our visualizations, we transform each image below a TrueSkill score of 15 using the pipeline shown in Joglekar *et al*.[3] and retain the top five most similar natural images.

Compositional Scores and Attributes

To understand what changes are being done during the transformation, there needs to be some insight about the composition of the Google Street View image before and after the transformation.

Attributes

Deep-learning frameworks like PlacesNet[14] and SegNet[15] help us understand the compositional aspects. PlacesNet is a convolutional deep-learning framework that uses a deep-learning-based classifier to classify an urban image. The framework has been proven to have a very high accuracy for at least the first five labels it outputs.[14] For the sake of visualization, we represent each urban image by these top five matching labels among the 205 possible labels. Concurrently, it is also important to visualize the changes in individual objects in an image as it is transformed. To that end, we use SegNet, which is another convolutional deep-learning-based framework that segments the pixels of the urban image into a set of 12 possible object segment types. The segment types represent common categories of objects seen in urban images, such as buildings, roads, pavement, pedestrians, vehicles, signage, fences, and trees.

Scores

Based on the features extracted by SegNet and PlacesNet, the framework combines these features to compute five composite scores,[3] each representing a distinct urban design metric derived from the literature.[4] These five scores have been found to characterize great urban spaces:

- *Walkability*. According to the urban design literature, walkable places have a higher chance of being perceived as beautiful and habitable. In our particular dataset of images, concepts such as presence of pavement as found by SegNet and presence of walkable scene types like parks, beaches, and gardens as found by PlacesNet are considered to be contributing to walkability. We calculate a simple term frequency–inverse document frequency (TF-IDF) of all the possible walkable scene types from PlacesNet and normalize the resulting score on a 1-to-5 scale.

- *Green cover*. The ratio of total pixels classified as trees according to SegNet to the overall area in terms of pixels was used as a proxy for green cover. The ratio is normalized on a 1-to-5 scale.
- *Landmarks*. Landmarks are icons that facilitate wayfinding and space legibility. Presence of landmarks was approximately measured by counting the number of iconic scene types according to PlacesNet. This is a gross approximation, albeit a useful one for crudely capturing an important metric such as the presence of landmarks.
- *Openness*. Open spaces are good for leisure but do not necessarily give a sense of security, according to the literature. To capture this metric, we calculate the amount of sky visible in a particular image as a proportion of the entire image. The more sky, the less inhibited the view, and the higher its openness.
- *Visual complexity*. Visual complexity is the diversity of visible objects. Previous studies have used entropy as a measure of visual complexity.[16] In a similar way, we compute the entropy of the ratio of pixels occupied by each urban element (extracted by SegNet) as a proxy for visual complexity.

VISUALIZING URBAN BEAUTIFICATION

To make the product of beautification understandable, we resort to information visualization. We use a user-centered design approach to set design goals and to iteratively design our visualizations.

User-Centered Design

We first need to identify our audience. To do that, we use an iterative user-centered design process to (a) identify our target audience, (b) define their problems, and (c) establish design goals that can solve these problems.

In a first step, we draft rough personas for four user groups: (1) activists who search for low-cost-high-reward interventions to beautify public space; (2) data scientists who share domain knowledge on the technological, but not application, side; (3) architects whose goal is to learn about the beauty of isolated buildings or compositions; and (4) urban planners who are trying to identify possible reasons for subjective judgements to beautify whole neighborhoods (but—compared to the activists—they do so in a more institutional manner).

Since the urban planner's goal is closely aligned to the nature of our underlying framework, we primarily chose them as our focus group. Secondly, we added data scientists, as they are able to provide valuable feedback on the used technology. Based on this focus group, we developed a use case of our application. This use case consists of having our visualization support the design of environments that are intuitively beautiful and that increase its dwellers' well-being.

Design Goals

First, we built a range of prototypes (see Figure 1). Those were then shown to a group of experts from our focus group during an urban-planning symposium. Afterwards, we interviewed them; a majority found the approach interesting, but criticized:

- the lack of a spatial overview;
- the oversimplification, referring to a 1D scale of beauty in the existing sketches that left no room to track changes; and
- the idea of building a prescriptive tool that acts as an authority for beauty.

We set design goals to address each of these three points:

- *Support spatial exploration*. The overall user experience should facilitate exploration of the geo-located results provided by the neural network. Hence, the user interface should provide us with a spatial representation of the data and should use low-level micro-interactions to encourage the user to spend time exploring.

- *Compare images and track changes*. For every combination of ugly and beautiful images we need to (a) juxtapose “the before beautification image” and “the after image,” (b) connect the explanatory computer-vision analysis results with the actual images, and (c) support comparative analysis.
- *Provide a comprehensive view for contingency*. Viewers should be able to draw different conclusions from the visualization, depending on context.[17] A very simple way of doing that is to create an alternative view of the same data, using a different angle.

Figure 1. One of the static prototypes first shown to experts at an urban-planning symposium.

Goal 1: Support Spatial Exploration

To achieve the first design goal, we link interactive textual labels to a dynamic geographic heat map and create an easy-to-use interface, which allows for the identification of spatial patterns of beauty (see Figure 2). By doing so, the actual examples of beautification can be localized to provide the orientation that the previous prototypes were missing. The landing page of the visualization is a set of markers overlaid on a map, together with an associated list of labels. The color scheme of the map is kept on a dark grayscale to make the markers placed on top of it pop out. The markers (dots) on the map represent 1.2 thousand locations in our dataset, 84 of which are beautifiable locations that are highlighted and expand on hover to indicate interactivity. On click, an overlay with the detail view for that location appears. The dot color indicates the perceived beauty of the locations, as determined by the TrueSkill score in the dataset.

The list of labels includes the most common scene types detected by PlacesNet.[14] To keep this list clearly laid out, we display only the 51 most frequent scenes (the scenes detected in five or more images). Scenes are ranked according to the scene-beauty score Score(l), which is the TrueSkill score of all locations l depicting scene L, weighted by the confidence in the label:

$$Score(l) = \frac{\sum_i TrueSkill(i) * Certainty(l,i)}{|L|} \text{ with } L = \{i | i \ contains \ l\} \ .$$

The color of each label corresponds to its Score(l) on a linear scale ranging from beautiful (blue) to ugly (red). This continuous color scheme was chosen to symbolize the need to act on ugly spaces with an alerting red color in contrast to the blue that exudes calmness.

Each label is connected to a set of associated points on the map, namely the locations of the images where the label was detected. The spatial density of these locations is expressed in a heat map; each location has a heat level that is based on the certainty with which a label is assigned to it.

The label list acts as an interactive trigger for the heat map; hovering on a label temporarily activates the heat map for that label. Ordering the labels by beauty provides the user with a low-threshold exploration method; by moving the cursor down the list of labels, one can playfully explore the spatial distribution of urban scenes in decreasing order of beauty. This explorative momentum is fostered by the possibility of keeping the click on a label—this permanently adds it to a list of "hot" labels.

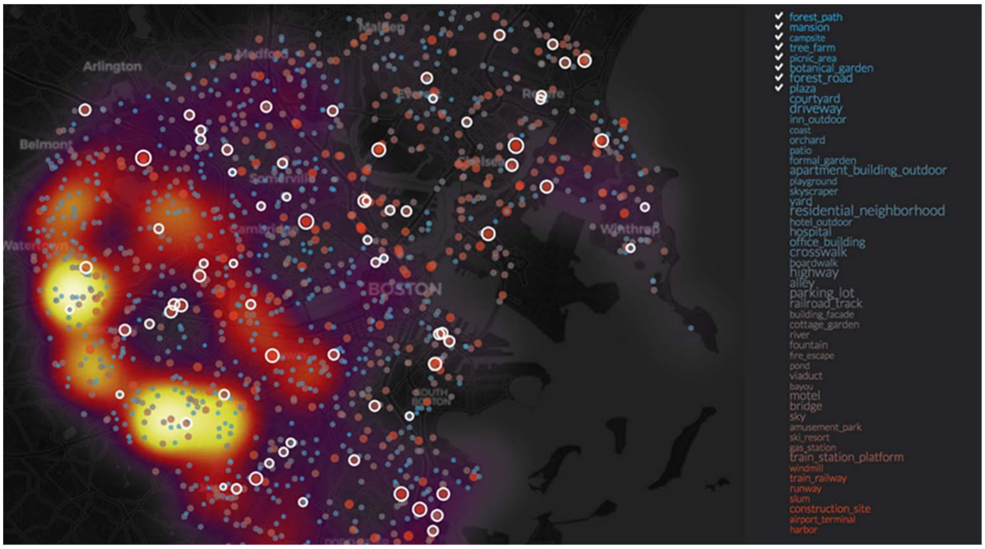

Figure 2. Overview with label list (right), activated labels (top right), and corresponding heat map (left).

This allows for the user to create cumulative heat maps for combinations of various labels, highlighting existing clusters of high and low beauty. White rings around the circles indicate locations where the underlying ML framework sees the possibility for beautification.

Goal 2: Compare Images and Track Changes

The detail-view overlay appears when a single interactive point on the map is clicked (see Figure 3). It shows a single combination of an ugly street image and its beautified alternative. It acts as a dashboard to provide a single view of all available data for this combination; besides the images before and after transformation, the calculated urban design metrics and urban elements connected to this particular combination are shown.

We use radar charts to visualize the urban design metrics (at the bottom-left corner of Figure 3). We chose this kind of visualization because it makes it possible to show and, importantly, compare all five dimensions at the same time. Radar plots use relatively little space compared to, for example, line charts. Using little space is important to support easy-to-browse-through pair interaction (see Figure 4(c)).

To display how individual urban elements change, we visualize the difference in the amount of pixels before and after transformation (at the bottom-right corner of Figure 3). The color coding is adapted to show the transformation; a decrease in the number of pixels for a given element is indicated with a red bar (as the ugly image shows a higher presence of that element), while blue bars reflect an increase.

In contrast to existing image-comparison interfaces,[7] the idea of over-imposing the before and after images is dismissed. Superposition needs additional complex interaction to compare two images. Furthermore, it is impossible to show both images at the same time. We chose a layout with juxtaposition of images to allow for the comparison without further interaction (Design Goal 2). This simplifies the interface and facilitates further low-threshold interactions; hovering on a single image triggers a state change for all data visualization in the detail view.

Hovering over one of the images highlights the corresponding radar chart (see Figure 5). At the same time, the visualization of urban elements changes—it no longer shows the relative change of elements, but rather the actual presence of elements in the hovered-on image. To conclude, we use hovering as an interaction method to achieve our goal of connecting images with the user.

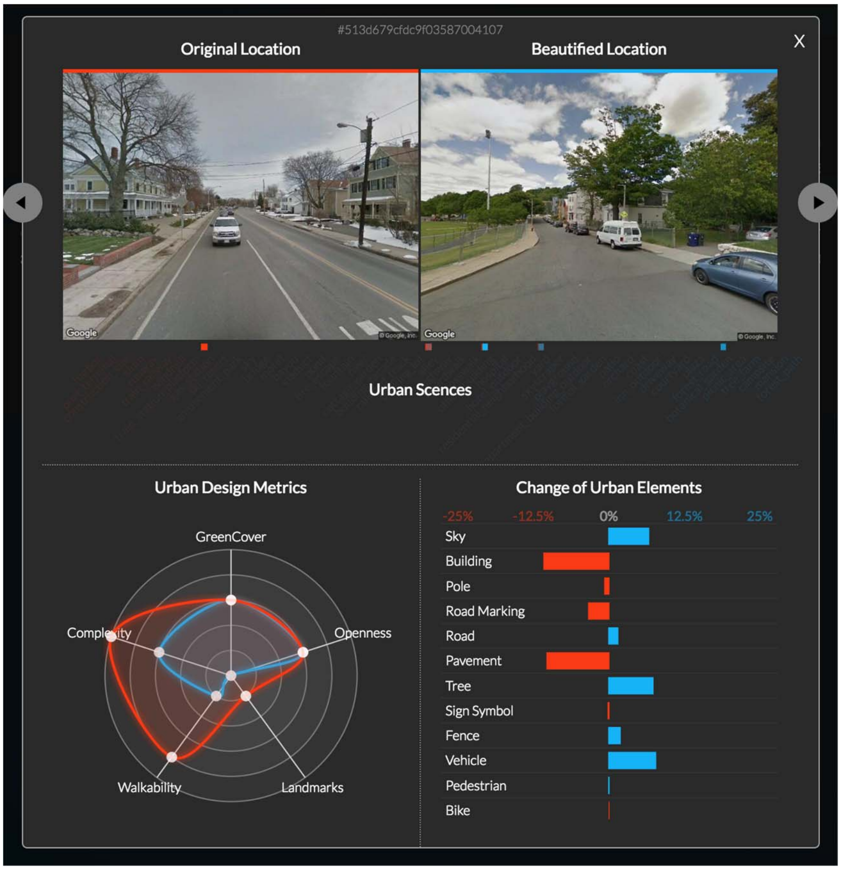

Figure 3. Detail view with image juxtaposition (top), list of detected urban scenes (middle), urban design metrics (bottom left), and change of urban elements (bottom right).

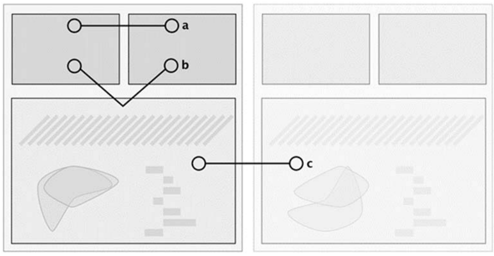

Figure 4. Supporting comparative analysis within and across detail views: (a) juxtapose images, (b) compare analysis results with actual images, and (c) compare various combinations.

The user can browse through the image collection using the arrow keys or by clicking either of the two arrows at the side of each urban scene (see Figure 3). While browsing through pairs, the position of all visualizations and the order of their dimensions remain the same. This way, the user can focus on one specific dimension on the chart and compare different examples.

The hover-and-click interaction (Goal 1) and the ability to easily browse through various image comparisons (Goal 2) are designed to meet four criteria of the "fluid interaction" paradigm:[18]

- *Balanced challenge*. The skill required by the activity and the user's skill level should be matched.

- *Transformation of time*. This enables users to “lose themselves” in the activity, essentially losing track of time.
- *Prompt feedback*. Users should be immediately informed of progress towards their goals.
- *Sense of control*. This ensures that users feel in control over the activity so that they can truly affect the outcome.

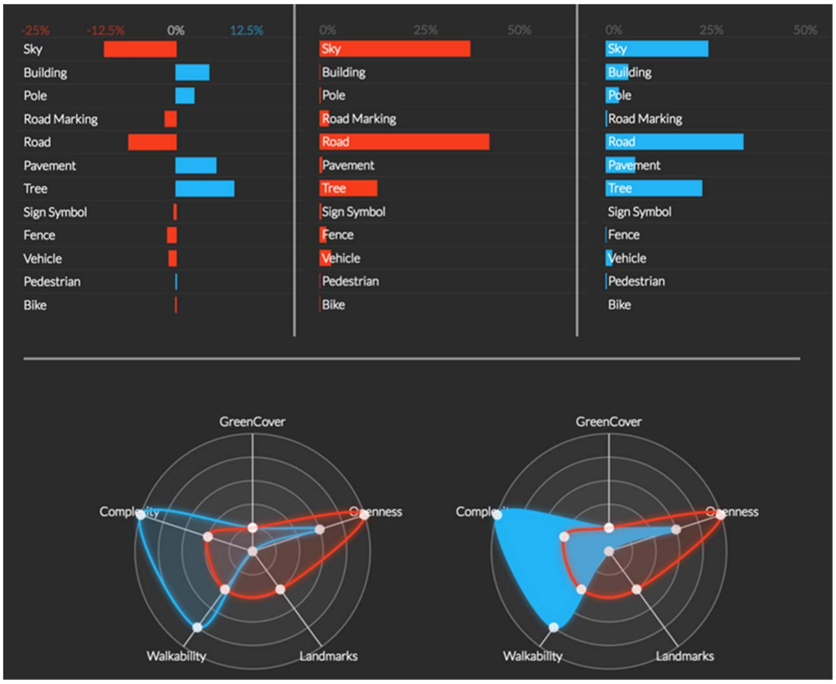

Figure 5. Top: Bar charts representing (left) the change of urban elements, (middle) their presence in the ugly original image, and (right) the beautified alternative. Bottom: Comparison view for (left) urban metrics radar chart with no image hovered on and (right) the beautiful image hovered on.

Goal 3: Comprehensive View for Contingency

We created a plot that shows how the urban elements and the design metrics changed overall during the beautification. Since this plot aims at showing which elements are generally associated with urban beauty, we call it the “DNA of urban beauty.” To enable the viewer to track overall patterns in the data, we show the computed increasing and decreasing presence of each urban element after beautification. Each urban element is displayed in a column, while its decreasing or increasing presence in a given image pair is shown as a vertical bar in each row (see Figure 6). Red bars indicate the element’s decreasing presence in beautified scenes, while blue bars indicate its increasing presence. The levels of increasing and decreasing presence are computed and displayed for the five urban design metrics.

Figure 6. DNA views for urban design metrics (left) and urban elements (right). Absolute values are on top, and the discontinued normalized scales are below.

EVALUATION

The first phase resulted in the design of the interactive FaceLift website, while the second phase was about validating whether the design goals were met (see Figure 7). In this latter phase, we tested whether our focus group was able to extract complex patterns of urban beauty from our application. Furthermore, we asked open-ended and Likert-scale questions to draw design guidelines for the future.

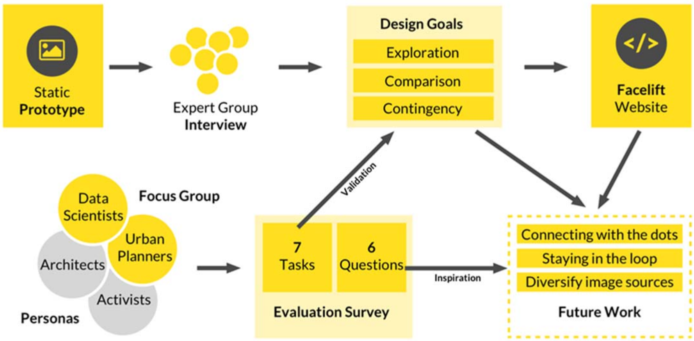

Figure 7. In the first phase (top), we generated design goals from the feedback we received from the symposium experts. We built the FaceLift website according to these goals, which were then validated by our focus group in an evaluation survey. The website, the validated design goals, and the open-ended survey questions form the guidelines for future work.

Experimental Setup

To understand the extent to which the proposed interface helps urban planners, we conducted an online survey. Participants were first asked to perform seven tasks on the interface (see Table 1), as well as to fill out the survey. Depending on the task, the questions were in the form of either multiple choice or Likert scales. The survey included several open questions answerable with free text fields (see Table 2). The survey was sent to a group of 120 experts. These experts work in one of the following fields: urban planning, architecture, transport engineering, or urban informatics. They were chosen among the authors' past collaborators and acquaintances. As such, the set is biased towards academics and practitioners in England and the United States. Among the academics, there is a disproportionate representation from the University College London, Massachusetts Institute of Technology (MIT), and University of Cambridge. Finally, the respondents were unpaid, and they answered the survey mainly because they wanted to be helpful and they have a professional interest in the topic of the survey. The survey was completed by 20 respondents, 13 (65 percent) of which have "very good" or "good" knowledge of urban planning, four (20 percent) of which are practitioners in the fields of ML or AI, and three (15 percent) of which are from unrelated fields.

We designed the six tasks (Table 1) over two iterations, each separated by a test-run with one person not involved in the study. The criteria for the final set of tasks (1) have to target one specific design goal, (2) have to be translatable into quantifiable metrics (see Evaluation Metrics 1 through 3), and eventually (3) need to be understandable to members of our focus group. The six accompanying questions (Table 2) are designed in the same fashion and were split into (1) Likert questions, which aim at ascertain the effectiveness of our visualization in the focus group's fields of work, and (2) open-ended questions, which are designed to result in points for future work.

Evaluation Metrics

1. We measure the participants' accuracy in identifying the most frequent beautiful labels by using the label list. We invite participants to explore the labels next to the map, and the labels are listed in decreasing order of beauty. We ask for which label they see the biggest activation on the corresponding heat map. The answer is then entered into a

free-text field. Based on the data, we expect the highest number of mentions for “residential_neighborhood,” as it appears 590 times with a mean TrueSkill score of 27 (“driveway” has 326 appearances with a 23.8 mean TrueSkill, “plaza” has 75 appearances with a 28.5 mean TrueSkill, and “forest_path” has 26 mentions with a 30.7 mean TrueSkill).

2. We measure the participants’ accuracy in correctly identifying spatial clusters of beauty. We invite participants to combine multiple labels to create cumulative heat maps. We ask which cardinal direction they see as the most beautiful in the city. The answer can be given by checking a radio box. We expect the participants to name “west” as the correct answer, as previous analysis showed that this area has the highest average TrueSkill score of 27.1 (“south” has a 26.2 mean TrueSkill, “north” has a 25.9 mean TrueSkill, and “east” has a 24.2 mean TrueSkill).
3. We measure the participants’ accuracy in correctly identifying patterns of change in our urban design metrics and urban elements. We do this for both visualization styles (detail view and DNA view) and evaluate which one performs better in conveying the overall patterns.
4. We ask participants to evaluate the usefulness of this technology for real urban-planning applications.

Results

We aggregated the results of the above tasks across all participants and found the following:

1. Seventy-five percent of the participants correctly identified “residential neighborhood” as the most frequent beautiful label. Since we asked for the “most frequent” but also “beautiful” label, participants were free to prioritize on their own. We observe that, except for one response, participants use the heat map to identify the most frequent label (“residential neighborhood”).
2. Similarly, the responses show an accuracy of 75 percent for correctly identifying “west” as the most beautiful area. Based on comments added by some respondents, we see that different techniques are used to perform this task. Some click several labels that they consider to be associated with beauty on the label list, and they then interpret the resulting heat maps; meanwhile, others simply look at the static map and the colored dots to infer the most beautiful area. Hence, the agreement for this metric is lower compared to the previous one. This is in line with the observation of one participant who says that “the maps reveal patterns that might not otherwise be apparent” and that “the tool helps for focusing on parameters to identify beauty in the city.”
3. We compare the detail view (Figure 5) and the DNA view (Figure 6) in terms of how well they help in understanding the underlying data. We compute the Spearman Rank Correlation “R” between the number of respondents who saw an increase in an urban design metric (such as openness) and the actual average increase for that metric in the data. We find that the viewer perception is positively correlated with the DNA visualization ($R = 0.7$) and negatively correlated with the detailed view. Similarly, we compute the rank correlation between our respondents’ perceived increase of each urban element and the actual average increase in the data. We find that the DNA view yields a relatively high correlation ($R = 0.43$) as opposed to the detail view ($R = 0.02$). This means that while the detail view might be appropriate for examining the changes and absolute values of a single pair of pictures, it is deceiving when trying to infer patterns of change across the whole dataset. For this task, the DNA view is able to expose knowledge in a much more accurate way.
4. We ask to what extent the tool could be used in three predetermined urban-planning scenarios. We found that a vast majority of participants see a general potential for the technology, and 85 percent of participants state that it is probably better than existing tools used for “participatory approaches to urban planning.” Seventy percent say the same about its utilization for decision-making, and another 70 percent also see potential in its ability to “promote green cities.” Other responses state that the technology could be used “in historic and neighborhood preservation” or to help the administration manage their “maintenance priorities.”

Table 1. List of tasks in our survey administered to practitioners.

Question	Task	Design Goal
What area is the most beautiful one?	Your task is to understand which of the four areas (west, east, south, or north) on the map (goodcity-life.org/facelift/#start) are the most beautiful. Select multiple labels to show cumulative heat maps.	1
How helpful are the urban scenes?	Explore at least five beautification examples (goodcity-life.org/facelift/#combinations) and hover over the before image and the after image. You will see lists of items changing. How helpful is this in giving you an extra understanding of the items that change in the beautification process?	2
Comparison View: Which urban design metrics usually increase or decrease?	Explore at least five beautification examples (goodcity-life.org/facelift/#combinations) and focus on the radar plots. Identify which ones usually increase or decrease in the beautified image.	2
Comparison View: What are three urban elements that increase most often?	Explore at least five beautification examples (goodcity-life.org/facelift/#combinations) and focus on the bar charts. Identify three urban elements that increase most often (comma-separated).	2
DNA View: Which urban design metrics usually increase or decrease?	Explore the DNA of beautification (goodcity-life.org/facelift/#dna) and focus on the urban design metric chart on the left. Identify which urban design metrics usually increase or decrease in the beautified image.	3
DNA View: What are three urban elements that increase most often?	Explore the DNA of beautification (goodcity-life.org/facelift/#dna) and focus on the bar charts. Identify three urban elements that increase most often (comma-separated).	3

DISCUSSION

Although the general idea of merging ML and information visualization to better understand complex data has been explored in past work, we are among the first to evaluate this approach from a data-literacy point of view. The single most striking result is that our visualization enables planners to grasp the complexity of urban beauty. One participant wrote:

> *The metrics are nice. It made me think more about beautiful places needing a combination of criteria, rather than a high score on one or two dimensions. It made me realize that these criteria are probably spatially correlated.*

However, as we state in our third design goal, the interface is intended to spark discussion about how the underlying technology can be further developed. Even though the interface "helped me think about what kind of a city we want to live in and what needs to be done to achieve that," there are various principle implications we learned from conducting our evaluation.

Table 2. List of questions (open-ended and Likert) in our survey administered to practitioners.

Topic	**Question**
Learning	What new ideas about the concept of urban beauty did you get by looking at the visualization?
Fields of application	To what extent could this technology be used in the following fields: participatory approaches to urban planning, decision-making, and promoting green cities?
Other future applications	Could the beautified scenes be useful to other applications? Please elaborate.
Explorability	Does the tool help you explore the city in a way you didn't think about before? Please elaborate.
Unintended consequences	Do you see any negative consequences in the beautification approach?
None	What would you like to see visualized that isn't shown at the moment?

Limitations

The limitations of the underlying technology already discussed by Joglekar *et al.*[3] also apply to our visualization; we can only show existing data, which is limited and biased. The training data of the current model are acquired through an interface open to the general public, and we cannot correct for biases that come from culture, location, or sentiments. That is less than ideal, especially considering that we are studying a highly subjective quality (beauty).

Furthermore, the beautified scene is not purely computer-generated; it is an existing scene that best matches the automatically generated template image, which is too coarse to look realistic and, as such, to be shown to the user. This results in two limitations. First, the interface suggests that the beautified image is computer-generated, while only its retrieval is so (a differentiation that might not be beneficial to individuals who are not supposed to understand the complexity of the underlying technology). Second, whenever the ML model is inaccurate, viewers have a hard time imagining that a particular location is beautified. As one participant put it, "The two images are always so different that it is REALLY hard to understand that they refer to 'the same place before and after.'" Future work should focus on how to best visualize the output of ML models, which is bound to be less than perfect. Indeed, there is an interesting tradeoff for visualization designers: the less accurate an ML model is, the more likely it is that the corresponding visualization breaks users' expectations.

Given these limitations, the application at hand can only be seen as a proof of concept that needs further development to be truly applicable to urban planning. We promote a set of design guidelines and future technology implications to make this possible.

Future Work

Based on our evaluation, we identify three main points for future work:

- *Connecting with the dots*. In his widely discussed article, "Connecting with the dots,"[19] Harris talks about the necessity to show not only abstract data points but also the people they represent to enable viewers to actually connect with the data. Our underlying data-driven technology comes with a high level of abstraction, so one vital function of the visualization has to be to connect this data to the actual geographical locations and how they look. We do so by showing not only the visualization of the data but also the before and after images. In general, as the visualization is intended to serve both the scientific community and practitioners, we recommend a user-centered and iterative approach to getting to know the audience and facilitating an empathetic visualization.
- *Staying in the loop*. As of now, the visualization and the underlying technology are kept separate. The output is generated from existing data and is then displayed. By contrast, an integration of feedback loops could rapidly increase data quality and thus further improve our results. Possible scenarios are (a) viewers enhancing the existing training data by comparing and rating more images and (b) practitioners flagging wrong output data (such as images that are not actually more beautiful or detected scenes that do not match the image). As we already have a user frontend, this would be an optimal starting point for creating an iterative process in which viewers could become contributors. This crowdsourced approach could be made even more attractive by implementing elements of gamification, as previous projects successfully demonstrated.[2,20]
- *Diversifying image sources*. The used body of Google Street View images provides us with a comprehensive set of images that are similar in quality, angle, and depth of focus. This fact, though, is a double-edged sword, since it also means that we can only show the city from a car-centric perspective, usually taken early in the morning. Other views of the city, which include areas only accessible by foot or areas that only develop their beauty throughout the day, are omitted. Augmenting and thus diversifying the data with user-generated content might well open new and more nuanced perspectives.

REFERENCES

1. A. De Botton, "The architecture of happiness," *Vintage*, 2008.
2. D. Quercia, N.K. O'Hare, and H. Cramer, "Aesthetic capital: what makes London look beautiful, quiet, and happy?," *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, 2014.
3. S. Joglekar and D. Quercia, "FaceLift: A transparent deep learning framework recreating the urban spaces people intuitively love," *Under Review*, 2018.
4. R. Ewing and O. Clemente, *Measuring urban design: Metrics for livable places*, Island Press, 2013.
5. L. Liu et al., *A machine learning-based method for the large-scale evaluation of the qualities of the urban environment*, 65, Computers, Environment and Urban Systems, 2017.
6. R. König et al., "Cognitive Computing for Urban Planning," *The Virtual and the Real in Planning and Urban Design*, Claudia Yamu, Alenka Poplin, Oswald Devisch, Gert De Roo, Routledge, 2017.

7. N. Naik et al., “Computer vision uncovers predictors of physical urban change,” *Proceedings of the National Academy of Sciences*, no. 29, 2017, pp. 7571–7576.
8. I. Seiferling et al., “Green streets − Quantifying and mapping urban trees with street-level imagery and computer vision,” *Landscape and Urban Planning*, vol. 165, 165, 2017, pp. 93–101.
9. A. Kachkaev and J. Wood, “Automated planning of leisure walks based on crowd-sourced photographic content,” *46th Annual Universities' Transport Study Group Conference*, 2014; http://openaccess.city.ac.uk/4943/.
10. A. Radford, L, Metz, and S. Chintala, *Unsupervised representation learning with deep convolutional generative adversarial networks*, arXiv, 2015.
11. N. Naik et al., “Streetscore-predicting the perceived safety of one million streetscapes,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014.
12. R. Herbrich, Z. Minka, and T. Graepel, “TrueSkill™: a Bayesian skill rating system,” *Advances in neural information processing systems*, 2007.
13. A. Nguyen et al., “Synthesizing the preferred inputs for neurons in neural networks via deep generator networks,” *Advances in Neural Information Processing Systems*, 2016.
14. B. Zhou et al., “Learning Deep Features for Scene Recognition using Places Database,” *Advances in Neural Information Processing Systems*, 2014.
15. V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A deep convolutional encoder-decoder architecture for image segmentation,” *arXiv*, 2015.
16. A.E. Stamps III, “Advances in visual diversity and entropy,” *Environment and Planning B: Planning and Design*, vol. 30.3, 2003, pp. 449–463.
17. M. Dörk et al., “Critical InfoVis: exploring the politics of visualization,” *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, 2013.
18. N. Elmqvist et al., “Fluid interaction for information visualization,” *Information Visualization*, vol. 10, no. 4, 2011, pp. 327–340.
19. J. Harris, “Connecting with the Dots”; https://source.opennews.org/articles/connecting-dots.
20. D. Quercia et al., “Psychological Maps 2.0: A Web Engagement Enterprise Starting in London,” *Proceedings of the 22nd international conference on World Wide Web*, 2013.

ABOUT THE AUTHORS

Tobias Kauer is an urban futures student at the University of Applied Sciences Potsdam. He is an interface designer and conducts interdisciplinary research in urban informatics, urban sensing, data visualization, and HCI. Contact him at tobias.kauer@fh-potsdam.de.

Sagar Joglekar is a King's India Scholar and PhD student in the Department of Informatics at King's College London. He is interested in representation learning and complex networks as they relate to quantifying intangible human properties. He has a master's degree from the University of California, Santa Barbara. Contact him at sagar.joglekar@kcl.ac.uk.

Miriam Redi is a research scientist at Nokia Bell Labs Cambridge. Formerly, she worked at Yahoo Labs and the Wikimedia Foundation. She has a PhD from EURECOM, Sophia Antipolis. She conducts research in visual data science, computational aesthetics, social multimedia analysis, and forensic image analysis. Contact her at miriam.redi@gmail.com.

Luca Maria Aiello is a senior research scientist at Nokia Bell Labs Cambridge and a research fellow of the Institute for Scientific Interchange (ISI) Foundation. He has a PhD in computer science from the University of Torino. He conducts interdisciplinary research in computational social science. Contact him at luca.aiello@nokia-bell-labs.com.

Daniele Quercia is the department head of social dynamics at Nokia Bell Labs Cambridge. Previously, he worked at Yahoo Labs, the University of Cambridge, and MIT. He has a PhD from University College London. He has been named one of *Fortune* magazine's Data All-Stars. His main research area is urban computing. Contact him at quercia@cantab.net.