The experiment was presented as shown in Figure 1. Participants were instructed: “In the following page, you will be presented three pictures of buildings. Imagine that you are meeting with a friend in a year’s time in front of the building highlighted in red. Your task is to help him find that building. To do so, please select the words that best describe what is special about this building compared to the two others.”.
Figure 1: Interface for the Amazon Mechanical Turk facade unusualness collection. As instruction, participants were asked to “select the words that best describe what is special about this building compared to two others” so that a friend could find it.

Figure 2: Our automatic description of the facade in the middle: “The facade with tall blue shop and wide brown shop.” On right, the heatmap of the most unusual features.
Figure 3: Zurich20 Analysis. Columns: Asset, feature type, colour and size usage. The vertical axis in the histograms refers to the number of times an asset or attribute was cited. Rows: Top: study results, Middle: LOF, Bottom: After regression. The regression helps to give descriptions closer the human-given annotations. Note that the material and shape are not used in the automatic unusualness inference.

Figure 4: TopK scores obtained for facade mining with our method (blue) against a randomness-based baseline (red) for Graz50 (left), ECP2011 (middle) and Zurich20 (right). The TopK score quantifies the agreement between the human and automatic rankings for the top K items. A perfect agreement gives a score of 1.
Figure 5: TopK scores obtained for facade characterization with our method (blue) against a randomness-based baseline (red) for Graz50 (left), ECP2011 (middle) and Zurich20 (right).