Linking Past to Present: Discovering Style in Two Centuries of Architecture

Stefan Lee[1] Nicolas Maisonneuve[2,*] David Crandall[1] Alexei A. Efros[3] Josef Sivic[2,*]

[1] Indiana University [2] INRIA [3] UC Berkeley

Abstract

With vast quantities of imagery now available online, researchers have begun to explore whether visual patterns can be discovered automatically. Here we consider the particular domain of architecture, using huge collections of street-level imagery to find visual patterns that correspond to semantic-level architectural elements distinctive to particular time periods. We use this analysis both to date buildings, as well as to discover how functionally-similar architectural elements (e.g. windows, doors, balconies, etc.) have changed over time due to evolving styles. We validate the methods by combining a large dataset of nearly 150,000 Google Street View images from Paris with a cadastre map to infer approximate construction date for each facade. Not only could our analysis be used for dating or geo-localizing buildings based on architectural features, but it also could give architects and historians new tools for confirming known theories or even discovering new ones.

1. Introduction

With the era of "big data" at hand, many academic disciplines are beginning to use online data to perform analyses that would have been impossible with traditional methods. For example, collaborations between sociologists and computer scientists are using online social network data to measure human behavior at unprecedented scales [16], while work in health informatics is using online data to monitor outbreaks of diseases [11] and to predict their spread [22] (albeit with some controversy [21] and missteps [15]). In the humanities, analysis of online data has given insight into historical legal records [14] and the dynamics of cultural history [23]. Large-scale analysis of digitized books through several centuries has been used to quantify changes in linguistic and cultural phenomena over time [18].

The vast majority of this analysis has been on textual content, often looking at simple features like occurrences and co-occurrences of keywords [11,18]. But analyzing images could help unlock other latent sources of data for the sciences and humanities. Some work has investigated using vision to organize and navigate historical images [4,24] and to discover hidden features in artwork [13,28], but is limited in scale by the relatively small number of artifacts available. In these cases when big datasets are not involved, human experts can often perform the analysis as well and as quickly as the automated algorithms.

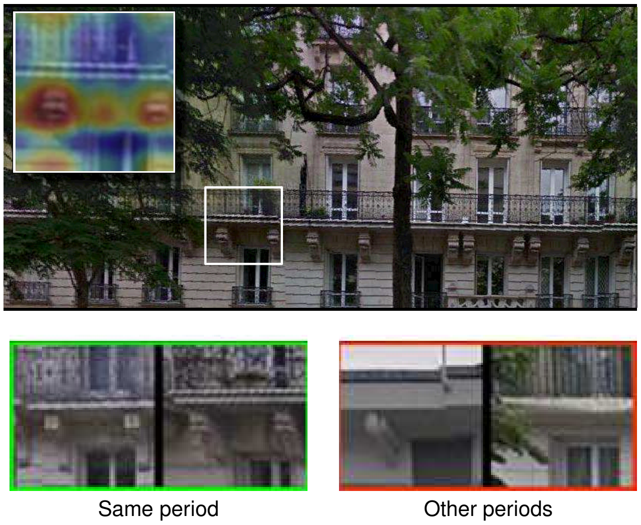


Figure 1: Using thousands of Street View images aligned to a cadastral map, we automatically find visual elements distinctive to particular architectural periods. For example, the patch in white above was found to be distinctive to the Haussmann period (late 1800's) in Paris, while the heat map (inset) reveals that the ornate balcony supports are the most distinctive features. We can also find functionally-similar elements from the same and different time periods (bottom).

Here we consider a domain in which we can study the dynamics of the past through vast image collections from the present: architecture. Architecture involves creating buildings "taking both aesthetic and practical factors into account" [20]. Thus architectural styles change over time,

[*] WILLOW project-team, Département d'Informatique de l'École Normale Supérieure, ENS/INRIA/CNRS UMR 8548, Paris, France.

reflecting the evolving artistic design, social and cultural attitudes, and technological and socioeconomic conditions of the peoples that built them. Studying features of buildings gives a window into the past, letting us observe properties of style and design at the time they were built.

Of course, manually collecting building data and images would be extremely tedious. We propose a novel analysis using Google Street View images geo-referenced against digitized maps. Street View includes street-level imagery from over 5 million miles of roads in more than 39 countries and 3,000 cities [9], capturing a huge number and variety of buildings around the world. For a given city (in our case, Paris), we associate individual building facades within Street View images and fine-grain urban planning records, letting us annotate each individual facade image with an approximate year of its construction.

With these combined data sources, we use mid-level visual features (similar to those of Doersch *et al*. [8]) to discover visual elements that are distinctive to particular architectural periods. We visualize these features over space and time, using a novel graph-based formulation to find functional elements (e.g. windows, doorways, etc.) whose style has changed across periods. In our experiments, we apply this analysis to nearly 150,000 images of 120,000 buildings of Paris. We show that these visual elements are useful for classifying the architectural style and construction date of buildings based on visual information alone (which could be useful for example for geo-localizing images by matching to city planning maps or for dating vintage photos), while potentially also generating new insights into how architectural styles have changed over time.

A key challenge is that we want to perform all of this analysis automatically. It is easy to find repeated visual elements in Street View, but most correspond to architecturally uninteresting objects like signs, bus stops, etc. We are instead interested in elements having *roughly* the same appearance because they correspond to the same function (balconies, windows, etc.), but whose appearance changes across time as styles evolved. Of course, boundaries between architectural styles are inherently blurry with a high degree of visual similarity between adjacent periods, and multiple distinct styles can exist within one time period. Important stylistic elements also wax and wane in popularity and can quickly rise to prominence only to be forgotten. Our approaches allow us to detect many diverse features, and connect functionally-similar elements through time with arbitrary start and end points. We evaluate our techniques both quantitatively and qualitatively.

To summarize our contributions, we:

1. generate large-scale architecture datasets automatically, using a novel combination of Street View image data aligned to cadastral maps;
2. discover architectural elements distinctive to particular time periods;
3. visualize these elements in detail, including which subelements make them distinctive, and how they combine to form distinctive facades;
4. find chains of visual elements corresponding to functional elements whose appearance has evolved; and
5. evaluate these techniques both quantitatively and qualitatively on a dataset of Paris.

2. Related Work

Perhaps most related to this paper is work in mid-level visual mining that tries to find discriminative image patches. Doersch *et al*. [7, 8] discover patches that discriminate between different cities using geo-referenced street-level imagery from Street View. The method requires iterative refinement of the patches using SVMs. Like us, Lee *et al*. [17] consider the temporal domain, finding style-independent classifiers of style-discriminative elements present throughout multiple time periods (like automotive headlights, which have been on cars for fifty years but whose style has changed dramatically over time). Our work also tries to find patches with similar semantics through time, but we face the additional challenge that elements in architecture are much more dynamic, with certain elements such as window shutters rising to prominence for decades only to fall out of favor later.

Some work in computer vision has considered architecture applications, including classifying between different architectural styles. For example, Shalunts *et al*. [25, 26] classify specific elements like windows and domes into three architectural periods (Gothic, Baroque, and Romanesque). Their results are promising, although their dataset was small (a few hundred images manually cropped around each element), in contrast to our huge collection of unconstrained Street View images. Xu *et al*. [31] classify whole building facades in about 5,000 images from Wikimedia, where buildings are well-centered and in full view. Since Wikimedia images are selected to be good examples of specific styles, they exhibit little of the blending and other complications that one observes in practice. They use deformable part-based models to visualize which facade elements are characteristic of which styles. Other work parses facade images into predefined elements such as doors, walls, windows, roofs, and balconies [6, 27, 30]. These approaches generally require cropped and rectified facade images. Though this parsing allows for well-aligned comparison of elements, defining elements by hand can be tedious and can introduce bias, so we instead follow a data-driven approach here in which elements are discovered automatically from large-scale image data.

Other recent work has used Google Street View, but for other applications than ours. Arietta *et al*. [3] use regressors based on mid-level patches to predict geospatially dis-

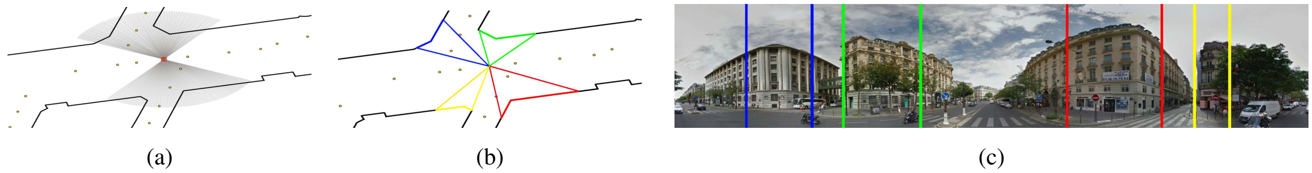


Figure 2: *Overview of data generation* from Google Street View images and cadastre maps. We first (a) cast 30m rays to the sides of each Street View capture location at 1° intervals, then (b) compute intersections with facades and select the widest view, and (c) project onto the panoramas and crop and warp the facade images according to Street View metadata.

tributed statistics such as crime rate and wealth. Ordonez and Berg [19] predict attributes of neighborhood safety, uniqueness, and wealth. Zhou *et al.* [32] demonstrate that the frequency with which Street View images contain certain attributes like green space, tall buildings, water, and social activities can be used to identify a particular city.

3. Data

We start by introducing our large-scale dataset of building facades constructed through many time periods. Rather than manually photographing buildings and researching their construction dates, we employ a noisy but automated process combining images from Google Street View, a 3D city map, and real estate records. We focus here on Paris, because it is one of the world's best-known cities and because fine-grained data on building construction dates is available (but our analysis could be easily applied to any other city covered by Street View).

Fine-grained building geometry. We use a digital cadastre (a survey of real estate boundaries) of Paris to retrieve detailed building geometry and construction dates. The cadastre was provided by the Paris Urban Planning Agency (Atelier Parisien d'Urbanisme) [2] and comes prepared in a standard GIS format including over 120,000 buildings (almost the entire city). The building geometries are recorded as 2d polygons plus maximum building height. Almost all of the buildings have a label indicating their coarse construction period, in terms of 10 periods: pre-1800, 1801-1850, 1851-1914, 1915-1939, 1940-1967, 1968-1975, 1976-1981, 1982-1989, 1990-1999, and after 2000. Some periods contain more buildings, like 1851-1914 during which the city grew from about 1 to nearly 3 million citizens. The cadastre also includes construction years for about 57% of buildings, although like much real-world data, these are noisy (for example, a suspiciously large number of years are divisible by 10).

Street View images. We collected every current Street View image and location metadata taken within the Paris city limits, yielding about 145,000 panoramas. The images were primarily captured in 2008 and 2012, with only about 11% captured in other years. The panoramas are composites of multiple images and have a resolution of about 13,300 by 6,600 pixels. They were generated from arrays of 9 to 15 cameras using Google's custom Street View vehicles [1].

Combining Street View and cadastre data. To link Street View images and building information, we need to align the images with the cadastre data. Each Street View image has a GPS coordinate, but we must still decide which buildings an image has captured and how to crop the panoramas to extract individual images. To do this, for each panorama we look up the Street View vehicle's heading from the metadata and cast rays in 160° cones from each side of the vehicle. The rays are cast at 1° intervals and are 30 meters long, which is sufficient to reach the buildings on even the larger Parisian thoroughfares (see Figure 2a). We compute the first facade encountered by each ray, and select the pair of rays from each facade with the greatest angular difference (see Figure 2b). We then crop and warp the panoramas to produce multiple facade images per Street View panorama.

A temporally labeled architectural dataset. Many of these 450,000 candidate facade images are not suitable for analysis because of extreme skews and angles of view. We remove facades which we see too narrowly, thresholding at a minimum field of view of 50°. To avoid images with high degrees of perspective warping, we remove facades that are within 5m of the camera or that are more than 10 degrees askew from the heading direction. This leaves us with images that are nearly parallel to the vehicle, such that the building facade is approximately planar. We crop the ground floor from all facades, since ground floors are often renovated and are poor representations of period architecture. From the remaining 70,000 facades, we sampled 20,000 images for analysis, evenly distributed among the construction periods such that each has 2,000 facades.

4. Discovering period-discriminative elements

Given our large dataset of building facades with time period annotations, our goal is to automatically discover discriminative architectural elements, and to capture their characteristics at multiple scales through time. We begin by finding patches that are highly representative of their respective time periods, by randomly sampling candidate patches and then performing nearest neighbor searches to evaluate discriminativeness. We then connect high quality patches to discover evolving trends across time periods.

4.1. Candidate generation

We begin with a mining approach similar to that of Doersch *et al.* [8], generating a large set $\mathcal{P}$ of candidate visual elements by sampling 25 patches at different resolutions from each of 2,000 images (which are then withheld from the remaining analysis). These images were selected uniformly across the periods, with low-gradient regions ignored during patch extraction. We represented each patch in Whitened HOG (WHO) space [10, 12], with WHO mean and covariance parameters learned over the 2,000 images. For each of the 50,000 patches, we build a set of initial "detections" by finding the closest match in each of the remaining 18,000 images, across all scales of the WHO pyramid. The result is a set of nearly 1 billion associations between images and patches, which we denote $\mathcal{C} = \{C_1, C_2, ..., C_{50000}\}$ where $C_i = (D_{i,1}, ..., D_{i,18000})$ and $D_{i,j}$ is the location and WHO distance of the best match of patch i in image j, sorted in order of increasing distance.

4.2. Identifying stylistically important elements

We define a patch to be stylistically important if it occurs often in one construction time period and rarely in others [17]. For each patch C_i, we find the closest 200 image matches (i.e. $D_{i,1}, ..., D_{i,200}$) to evaluate its discriminativeness. We observe that matches past these 200 often degrade to the point that visual correspondences are weak and would add noise to our analysis. (If our task were classification, we would not be concerned about this noise as long as it improved performance, but here we are attempting to find salient elements.) To evaluate each candidate C_i, we give each of the 200 close detections a binary label indicating whether the close detection is in the same period as the candidate. We define a simple linear classifier $f_i(x) = \text{sign}(\hat{C}_i \cdot x + b_i)$, where $\hat{C}_i$ is the whitened HOG descriptor of the candidate written as a vector, and b_i is a scalar bias. This choice of classifier is justified because the vector $\hat{C}_i$ can be interpreted as a classifier trained with the square loss from the candidate C_i as a positive example and large number of negative data points [4, 10, 12].

We apply $f_i(x)$ to each of the image matches D_{ij} and vary b_i to produce a precision-recall curve. We rank the candidates C_i by area under the curve (AUC), because candidates high AUC should be distinctive to a given period. We prune this set by finding candidates that have overlapping detections (i.e. two detections in the same image which have an intersection over union greater than 0.5) and removing the one with the weakest rank. Typically this process culls the population down to a few thousand per period.

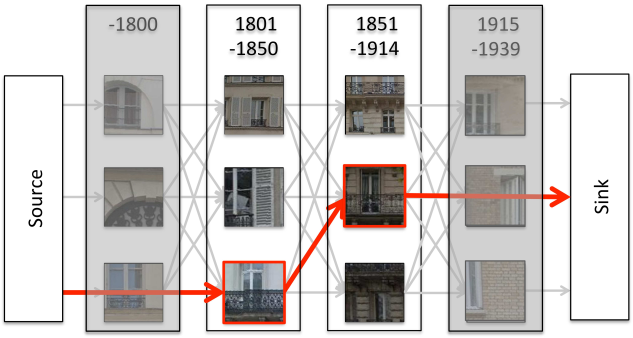


Figure 3: *Sample chain graph.* Elements in adjacent periods are fully connected with weights depending on their co-occurrence, while the source and sink connect to every node with weights that penalize the number of skipped periods. Here, the shortest path (in red) skips pre-1800 and 1915-1939 because they lack the long balconies of the other periods. (For clarity, this visualization shows only four periods (instead of ten), and only some source and sink edges.)

This procedure finds visual elements whose appearance correlates with building age, and ignores other common elements found throughout the city (like pavement, signs, bus stops, etc.) due to their very poor ranking; Figure 4 shows some examples. Some of the remaining elements are distinct to a given period, while others correspond to elements that have undergone visual evolution, as we now discuss.

4.3. Tracking the evolution of style

Functionally-identical elements of buildings can change substantially over time; for example, the styles of windows, doors, balconies, etc. vary dramatically across different architectural periods. We try to automatically identify these evolutions by looking for "chains" of elements that are discriminative to a particular time period, but that are still coarsely similar in visual appearance to elements in neighboring periods. More concretely, given the set of deduplicated and ranked candidates $\hat{\mathcal{C}}$ with period labels L generated as described in the previous section, we want to find a chain $C_1, \ldots, C_k$ of similar elements such that $L_{C_{i+1}} = L_{C_i} + 1$. We cannot fix the length k of the chain or the beginning or ending periods (L_{C_1} or L_{C_k}) in advance, since elements may appear or disappear over time. This problem is reminiscent of multiple-target tracking [5], in which detections of an object from sequential frames of video are stitched together to form trajectories, except that we are "tracking" patches over sets of images from different time periods.

We define a directed acyclic graph $G = \{\mathcal{V}, \mathcal{E}\}$ such that $\mathcal{V} = \{s, t\} \cup \hat{\mathcal{C}}$ where s and t are special source and sink nodes. The graph forms a trellis, such that each patch in any given time period has an outgoing edge to every patch in the next period, while the source and sink connect to all nodes (in all time periods) of the graph. Figure 3 presents a sample graph with four periods and three patches per period. Intuitively, the inter-period connections provide possible evolutions of corresponding elements. The source and sink nodes are added to determine the start and end of a chain, with weights such that if many matches for a patch are from the future, it is likely to be a starting point; otherwise, it tends to be an ending point.

For the edge weights, we need a measure of visual similarity that will connect patches likely to correspond to the same functional elements (e.g. windows, balconies, etc.). We could use visual similarity in WHO space, except that the appearance of some functional elements varies rather dramatically across time. We thus instead use relationships between images and candidate patches, looking at the spatial consistency of detections between candidates as a measure of similarity. For example, even if a window from 1939 looks quite different from one from the 1800s, it will still likely find its closest detection on a window region instead of anywhere else on a facade.

In particular, for pairs of patches C_i and C_j with $(C_i, C_j) \in \mathcal{E}$, we compute the intersection-over-union between their top 200 detections, $(D_{i,1}...D_{i,200})$ and $(D_{j,1}...D_{j,200})$. We store this overlap value and the spatial offsets of the centers of overlapping pairs as θ and Δ, where $\theta \in \mathbb{R}^N$ and $\Delta \in \mathbb{R}^{N \times 2}$ and N is the number of overlapping pairs. We also compute the component-wise trimmed mean $\bar{\Delta}$, discarding outliers outside the 0.1 and 0.9 quantiles. To capture the consistency of the overlaps, we take a mean deviation $\mu_\Delta = \frac{1}{n} \sum ||\Delta_i - \bar{\Delta}||$, where higher values indicate less consistency, and compute an edge weight,

$$w_{ij} = \mu_\Delta * e^{\lambda / \sum \theta},$$

where λ is a constant (5 in our experiments) controlling how steeply the cost rises for low confidence overlaps. The exponential term incorporates our level of confidence in μ_Δ based on the number and quality of the overlaps. Intuitively, two elements i and j that co-occur in the same images with a consistent spatial offset will have a low weight.

The source and sink nodes are attached to every other node with weights dependent on each candidate's closest detections. In particular, the weight from source to C_i is defined as $n_< * \beta * f_<$ where $f_<$ is the fraction of the top 200 detections for C_i that are from periods before L_{C_i} and $n_<$ is the number of periods before L_{C_i}. The weights to the sink are defined similarly, considering periods and fractions of those periods past L_{C_i}. Generally these weights will be low when a patch lacks many detections in earlier or later periods respectively. We set β empirically as a typical edge cost in a high quality chain, so that these weights balance the total cost of continuing the chain, amortized by how likely the chain should continue. To generate chains we greedily find the shortest path from source to sink, remove it, and repeat.

5. Results

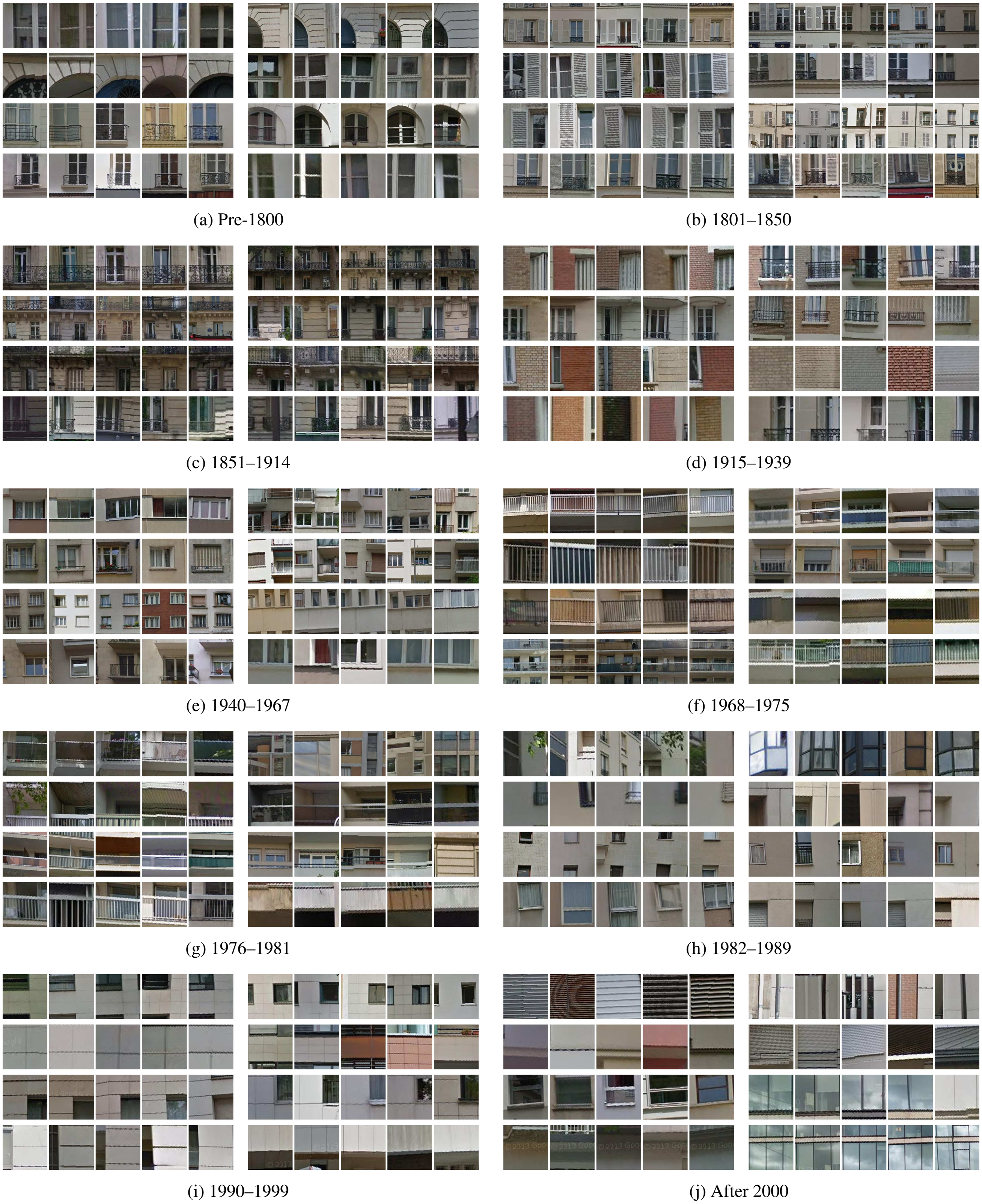
We now evaluate our dataset construction and visual element discovery methodology on our large-scale dataset from Paris. Our goal of automatically producing informative architectural visualizations is inherently qualitative, so we evaluate it in several different ways. First we view the problem in terms of the discriminativeness of our patches. Then we look in detail at sample output and identify historical trends that our automatic methods have captured. Finally, we report on qualitative results based on the feedback of an expert on Parisian architecture, who used an interactive version of our analysis.

5.1. Period-wise Analysis

One way of evaluating our discriminative patches is to test their classification effectiveness, i.e. how well they discern visually similar elements from different periods. As discussed in Section 4, we score our candidate patches by taking the nearest 200 detections for each candidate as our collection of visually similar elements, and compute the area under a precision-recall curve for each patch. We show the top eight elements ranked by AUC for each period in Figure 4.

In Figure 5a we plot the mean AUC for the highest ranked k patches from each period, varying k between 1 and 200. Note that after about the 10th to 30th best patch, the mean AUC drops substantially for all periods. We take this to imply that periods are surprisingly well characterized by only about a dozen key elements. Also interesting is the relatively higher mean AUC for the two periods spanning 1801-1914. These periods include Haussmann's renovation of Paris (during which Emperor Napoléon III commissioned Georges-Eugène Haussmann to redesign much of the city center) and perhaps are more cohesive because of the heavy influence of a single person. Figure 5b shows the distribution of period labels for the 200 nearest detections for the top 200 patches in each period. As we would expect, much of the mass is around the diagonals, indicating that most confusion is with adjacent periods. Both Figures 5a and 5b show that later periods are more challenging than earlier ones, perhaps because they are shorter and thus presumably more similar to neighboring periods.

Facade-level analysis. While these results suggest that we are discovering discriminative patches, one of our larger goals is to find patches that are relevant and useful to studies of architecture. This is difficult to quantify, so we showed our discovered patches to an expert on Parisian architecture and asked for feedback [29]. They informed us that many of the patches did capture key elements known to be prevalent



(a) Pre-1800

(b) 1801–1850

(c) 1851–1914

(d) 1915–1939

(e) 1940–1967

(f) 1968–1975

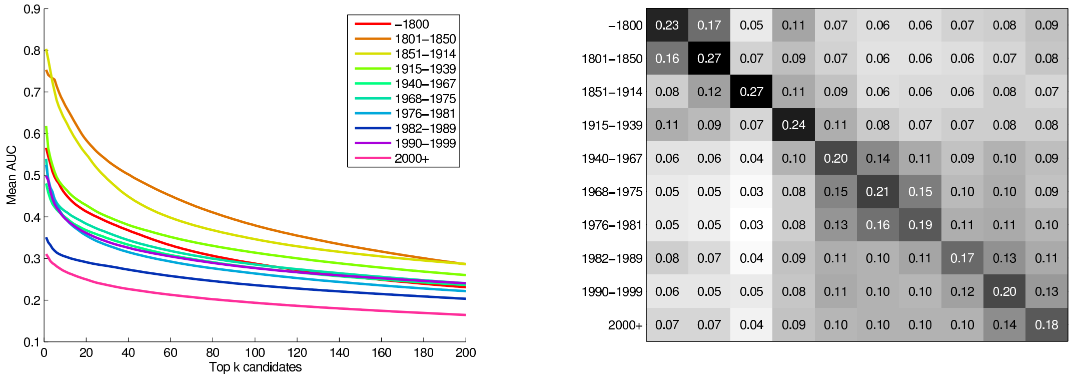
(g) 1976–1981

(h) 1982–1989

(i) 1990–1999

(j) After 2000

Figure 4: *Top 8 discriminative elements for each period.* For each element, five example patches are shown in a row. Notice how as we move towards modern times the patches largely become repeated geometries.



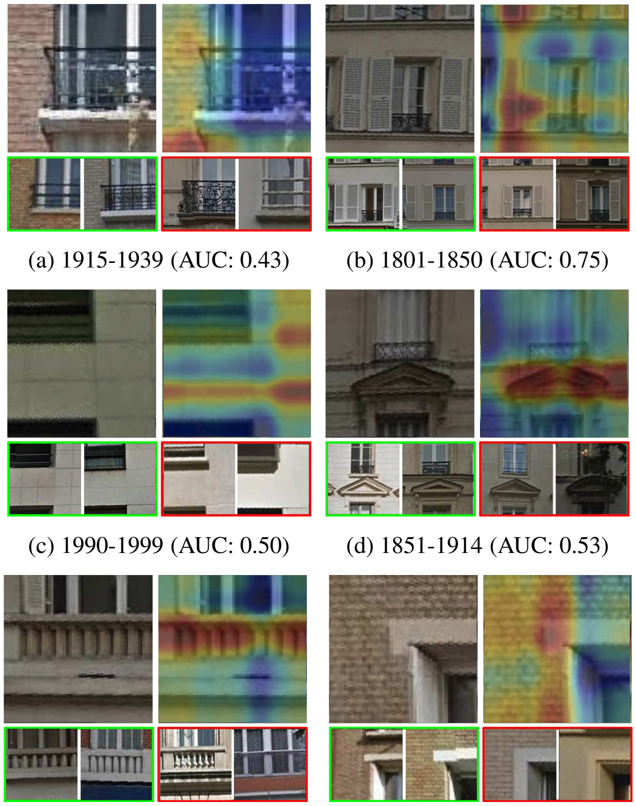
–1800	0.23	0.17	0.05	0.11	0.07	0.06	0.06	0.07	0.08	0.09
1801–1850	0.16	0.27	0.07	0.09	0.07	0.06	0.06	0.06	0.07	0.08
1851–1914	0.08	0.12	0.27	0.11	0.09	0.06	0.06	0.06	0.08	0.07
1915–1939	0.11	0.09	0.07	0.24	0.11	0.08	0.07	0.07	0.08	0.08
1940–1967	0.06	0.06	0.04	0.10	0.20	0.14	0.11	0.09	0.10	0.09
1968–1975	0.05	0.05	0.03	0.08	0.15	0.21	0.15	0.10	0.10	0.09
1976–1981	0.05	0.05	0.03	0.08	0.13	0.16	0.19	0.11	0.11	0.10
1982–1989	0.08	0.07	0.04	0.09	0.11	0.10	0.11	0.17	0.13	0.11
1990–1999	0.06	0.05	0.05	0.08	0.11	0.10	0.10	0.12	0.20	0.13
2000+	0.07	0.07	0.04	0.09	0.10	0.10	0.10	0.10	0.14	0.18

(a) (b)

Figure 5: *Evaluation of top 200 patches per period:* (a) Mean AUC as a function of number of top candidates. (b) Distribution of labels in top detections for each period.

in their respective periods, but that architects usually look at entire facades where many details combine to indicate the facade's period, as opposed to studying individual features.

Inspired by this observation, we used our ranked candidates to evaluate the 'periodness' of whole facades. For each facade, we found the top 100 detected patches. We sum the AUC of the detected patches for each period in a facade to produce an unnormalized distribution over how well each period's patches fit the given facade. In Figure 6 we show the highest likelihood facade for each period. Each image is accompanied by an over-painting of patch detections with colors corresponding to source period (using the same color coding as in Figure 5a) and a "reconstruction" of the image made by averaging these detections. For instance, notice how the 1851-1914 facade demonstrates the similarity in the periods spanning pre-1800 to 1914; its overpainting has colors corresponding to patch detections from pre-1800 (red) and 1801-1850 (orange) in addition to its own (yellow). The figure presents a sense of the progression in style and types of buildings constructed in Paris over the last two centuries, as modern materials gradually overtake old. The confusion about later periods is again seen here, with later periods exhibiting higher degrees of confusion (indicated by more mixture of color in the over-painting).



(a) 1915-1939 (AUC: 0.43) (b) 1801-1850 (AUC: 0.75)

(c) 1990-1999 (AUC: 0.50) (d) 1851-1914 (AUC: 0.53)

(e) 1915-1939 (AUC: 0.30) (f) 1915-1939 (AUC: 0.34)

Figure 7: *Sample discriminative elements.* Each figure shows a patch (top left), a fine-grained importance map (top right), close examples from the same (bottom left) and other (bottom right) time periods. Best viewed in color.

Fine-grained substructures. We also took a finer-grained perspective, looking for the most discriminative substructures within each patch. In particular, we drop each spatial cell of the WHO feature in sequence, by recomputing the patch's distances and AUC while ignoring that cell. This shows us which spatial cells are important, by observing drops in the AUC relative to the entire patch. We visualize this with a histogram of these differences where the i, j-th entry is the difference in AUC when the i, j-th spatial bin is masked. A sample of these visualizations is shown in Figure 7 along with some candidates and their nearest detections. The 1915-1939 period is characterized by new raw brick facades, highlighted in Figure 7a. Figures 7b and 7c give importance to details not identified by our expert. In Figure 7b, the spacing between adjacent window shutters appears to be influential in discerning between periods. In

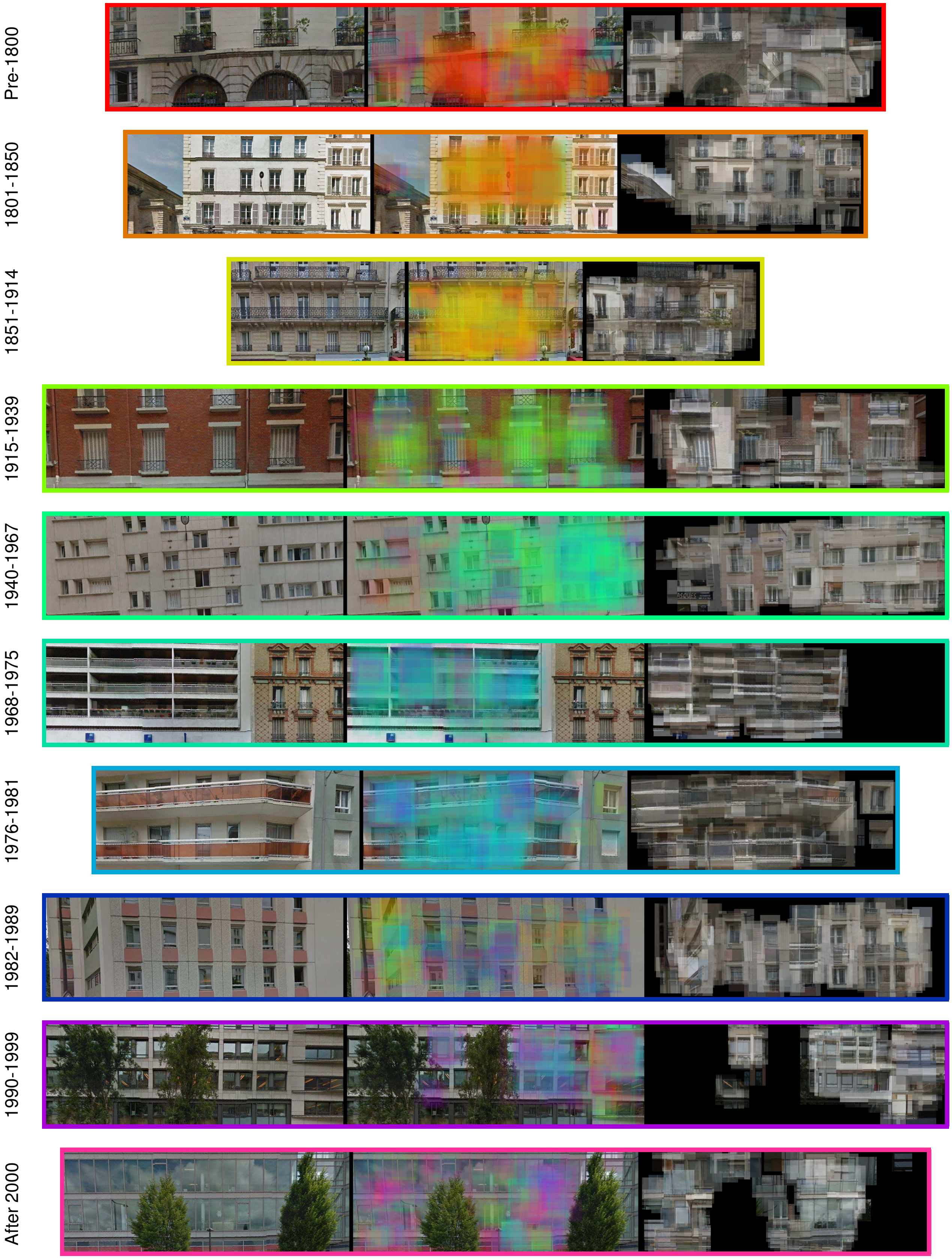


Figure 6: *Exemplars of each period according to facade level analysis.* In each row we show the original facade (left), the original overpainted with the periods of the top 100 detections (middle), and a reconstructed version of the facade where the period patches are replaced by their average images (right). Note the progression of style and the types of building constructed over the 200 year span. (Best viewed in color.)

Figure 7c the additional horizontal line is missing in many similar pre-1990 facades. The cap in Figure 7d is highlighted as well. Interestingly the highlight extends off the right-hand side indicating that the continued horizontal may also be important. The railing in Figure 7e sets itself apart from other similar elements by the plainness of its columns as compared to close negative patches. Figure 7f is unique among the examples because the map highlights an area because of what is *not* present: in the close negative examples, the white trim extends down the side of the window.

5.2. Style chains

Next we evaluated our technique for finding "chains" of similar functional elements whose appearance has evolved over time. Figure 8 shows sample chains of varying length and differing elements. Figures 8a and 8i show increasingly ornate window dressings starting from very plain structures before 1800, to multiple decorative structures in the 1851–1914 period. Figure 8c shows the long window balconies of the 1850s to 1940s, while Figure 8d shows an evolution of short balconies. Many similar chains are produced as there is a great deal of variety in balcony shapes over time. Some chains show consistent directions of change, for instance Figure 8e demonstrates the increasing depth of windows. The last chain in Figure 8f highlights railings for large buildings after 1940, with the railings transitioning to glass in 1982–1989 and into metal in the 1990s.

6. Conclusion

We presented simple but effective methods to automatically discover and track visually important architectural elements using an automatically annotated collection of thousands of street-level images of Paris. The images are mapped to buildings in a fine-grain urban planning model that annotates each with a rough construction date. Using these combined data sources, we mine for period specific stylistics elements, analyze facade-level architectural influences, and find evolutions of elements across times. This work is one step towards a longer-term goal of developing automatic techniques to mine large-scale image collections in order to help experts in other disciplines discover important and meaningful visual patterns.

Acknowledgments. We are grateful to Yves Ubelmann for discussions and thoughtful comments. This research was supported in part by the National Science Foundation through CAREER grant IIS-1253549, by the Intelligence Advanced Research Projects Activity (IARPA) via Air Force Research Laboratory, contract FA8650-12-C-7212, and by the European Research Council (ERC grant LEAP no. 336845), Agence Nationale de la Recherche (Semapolis project, ANR-13-CORD-0003) and the Inria CityLab IPL. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, AFRL, NSF, or the U.S. Government.

References

[1] D. Anguelov, C. Dulong, D. Filip, C. Frueh, S. Lafon, et al. Google Street View: Capturing the world at street level. *Computer*, 43(6):32–38, 2010. 3

[2] Atelier Parisien d'Urbanisme. http://www.apur.org/. 3

[3] S. Arietta, A. A. Efros, M. Agrawala, and R. Ramamoorthi. City forensics: Using visual elements to predict non-visual city attributes. *IEEE Transactions on Visualization and Computer Graphics*, 2014. 2

[4] M. Aubry, B. Russell, and J. Sivic. Painting-to-3d model alignment via discriminative visual elements. *ACM TOG*, 33(2), 2014. 1, 4

[5] J. Berclaz, F. Fleuret, E. Turetken, and P. Fua. Multiple object tracking using k-shortest paths optimization. *PAMI*, 2011. 4

[6] A. C. Berg, F. Grabler, and J. Malik. Parsing images of architectural scenes. In *ICCV*, 2007. 2

[7] C. Doersch, A. Gupta, and A. Efros. Mid-level visual element discovery as discriminative mode seeking. In *NIPS*, 2013. 2

[8] C. Doersch, S. Singh, A. Gupta, J. Sivic, and A. Efros. What makes Paris look like Paris? *ACM TOG*, 31(3), 2012. 2, 4

[9] D. Farber. Google takes Street View off-road with backpack rig. *CNET*, 2012. 2

[10] M. Gharbi, T. Malisiewicz, S. Paris, and F. Durand. A gaussian approximation of feature space for fast image similarity. *Technical report, MIT*, 2012. 4

[11] J. Ginsberg, M. Mohebbi, R. Patel, L. Brammer, M. Smolinski, and L. Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 457:1012–1014, 2009. 1

[12] B. Hariharan, J. Malik, and D. Ramanan. Discriminative decorrelation for clustering and classification. In *ECCV*, 2012. 4

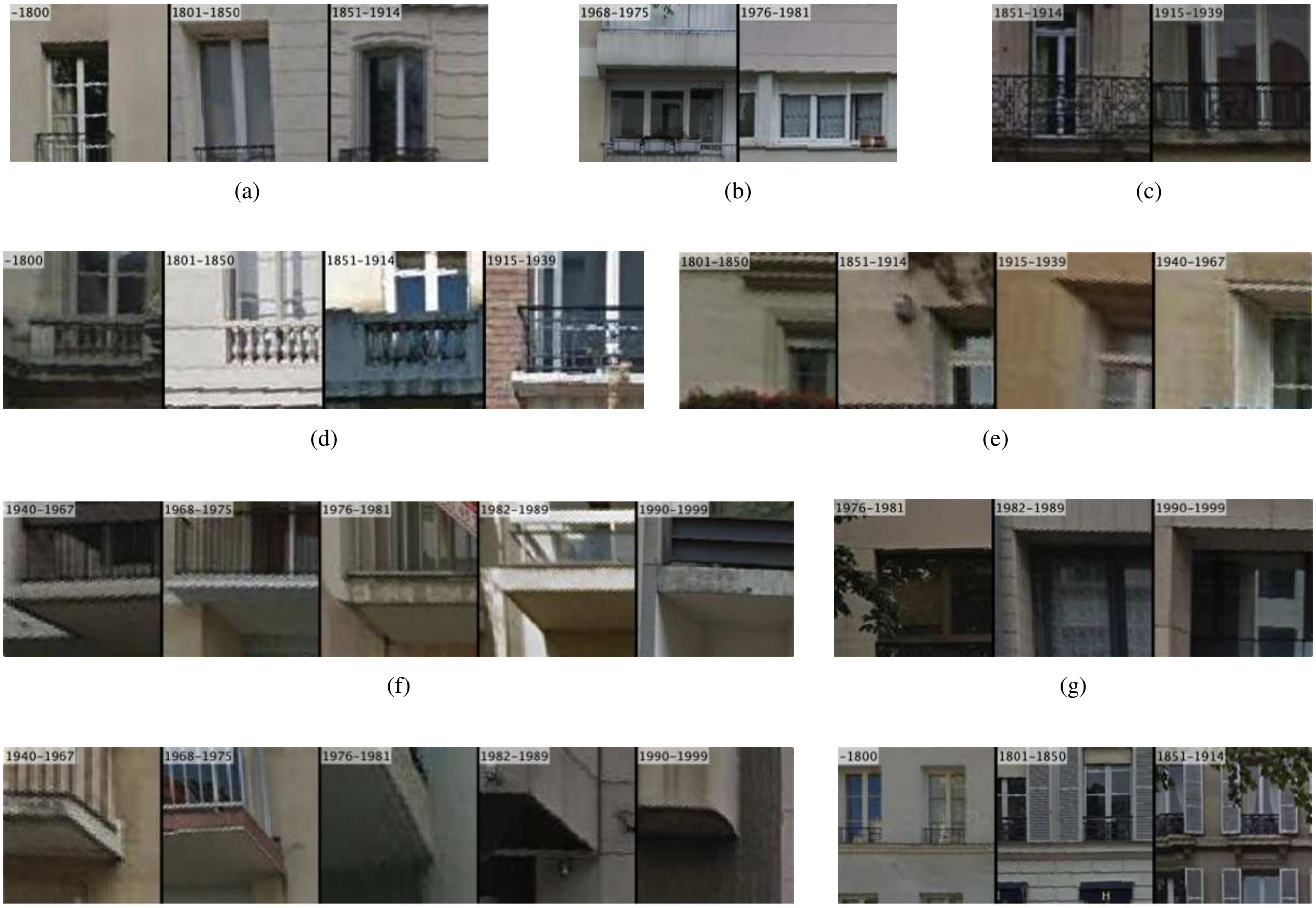
[13] D. Kim, S. Son, and H. Jeong. Large-scale quantitative analysis of painting arts. *Scientific Reports*, 4(7370), 2014. 1

[14] S. Klingenstein, T. Hitchcock, and S. DeDeo. The civilizing process in London's Old Bailey. *PNAS*, 111(26), 2014. 1

[15] D. Lazer, R. Kennedy, G. King, and A. Vespignani. The Parable of Google Flu: Traps in big data analysis. *Science*, 343(6176), 2014. 1

[16] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabasi, et al. Computational social science. *Science*, 323(5915):721–723, Feb. 2009. 1

[17] Y. J. Lee, A. A. Efros, and M. Hebert. Style-aware mid-level representation for discovering visual connections in space and time. In *ICCV*, 2013. 2, 4



(a) (b) (c)

(d) (e)

(f) (g)

(h) (i)

Figure 8: *Sample chains of architectural elements across time periods,* showing how our technique can discover functionally-similar elements whose visual appearance has evolved over time (see Section 5.2).

[18] J.-B. Michel, Y. Shen, A. Aiden, A. Veres, M. Gray, et al. Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014), 2011. 1

[19] V. Ordonez and T. L. Berg. Learning high-level judgments of urban perception. In *ECCV*, 2014. 3

[20] *The Oxford Dictionary of English*. 2010. 1

[21] D. Ruths and J. Pfeffer. Social media for large scale studies of behavior. *Science*, 346, 2014. 1

[22] M. Salathé, L. Bengtsson, T. Bodnar, D. Brewer, J. Brownstein, et al. Digital epidemiology. *PLOS Computational Biology*, 8(7), 2012. 1

[23] M. Schich, C. Song, Y.-Y. Ahn, A. Mirksy, M. Martino, et al. A network framework of cultural history. *Science*, 345(6196), 2014. 1

[24] G. Schindler and F. Dellaert. 4d cities: analyzing, visualizing, and interacting with historical urban photo collections. *Journal of Multimedia*, 7(2), 2012. 1

[25] G. Shalunts, Y. Haxhimusa, and R. Sablatnig. Architectural style classification of building facade windows. In *Advances in Visual Computing*, pages 280–289, 2011. 2

[26] G. Shalunts, Y. Haxhimusa, and R. Sablatnig. Architectural style classification of domes. In *Advances in Visual Computing*, pages 420–429. Springer, 2012. 2

[27] L. Simon, O. Teboul, P. Koutsourakis, L. Van Gool, and N. Paragios. Parameter-free/pareto-driven procedural 3d reconstruction of buildings from ground-level sequences. In *CVPR*, 2012. 2

[28] D. Stork. Computer vision and computer graphics analysis of paintings and drawings: An introduction to the literature. In *Computer Analysis of Images and Patterns*, 2009. 1

[29] Y. Ubelmann. *Personal communication*, Dec 2014. 5

[30] J. Weissenberg, H. Riemenschneider, M. Prasad, and L. Van Gool. Is there a procedural logic to architecture? In *CVPR*, 2013. 2

[31] Z. Xu, D. Tao, Y. Zhang, J. Wu, and A. C. Tsoi. Architectural style classification using multinomial latent logistic regression. In *ECCV*, 2014. 2

[32] B. Zhou, L. Liu, A. Oliva, and A. Torralba. Recognizing city identity via attribute analysis of geo-tagged images. In *ECCV*, 2014. 3