A Century of Portraits: A Visual Historical Record of American High School Yearbooks

Shiry Ginosar Kate Rakelly Sarah Sachs Brian Yin Alexei A. Efros
University of California Berkeley Brown University University of California Berkeley

Abstract

Many details about our world are not captured in written records because they are too mundane or too abstract to describe in words. Fortunately, since the invention of the camera, an ever-increasing number of photographs capture much of this otherwise lost information. This plethora of artifacts documenting our "visual culture" is a treasure trove of knowledge as yet untapped by historians. We present a dataset of 37,921 frontal-facing American high school yearbook photos that allow us to use computation to glimpse into the historical visual record too voluminous to be evaluated manually. The collected portraits provide a constant visual frame of reference with varying content. We can therefore use them to consider issues such as a decade's defining style elements, or trends in fashion and social norms over time. We demonstrate that our historical image dataset may be used together with weakly-supervised data-driven techniques to perform scalable historical analysis of large image corpora with minimal human effort, much in the same way that large text corpora together with natural language processing revolutionized historians' workflow. Furthermore, we demonstrate the use of our dataset in dating grayscale portraits using deep learning methods.

1. Introduction

In their quest to understand the past, historians—from Herodotus to the present day—primarily rely on two sources of data that humanity has left behind through the ages: 1) textual accounts; and 2) visual and material artifacts. The invention of the daguerreotype in 1839 as a means of relatively cheap, automatic image capture heralded a new age of massive visual data creation with potentially profound implications for historians. This new format was complementary to historical texts, as it could both capture details too obvious to put down in writing, and also transmit non-verbal information that would otherwise be lost. For example, it would be hard for a future historian to understand what the term "hipster glasses" refers to, just as it is difficult for us to imagine what flapper galoshes might look like from a written description alone [5]. However, despite public adoption of photography in the past century and a half, and the abundance of online historical visual data, historians are limited by the amount of data a human curator can manually process. Typically, only comparatively small-scale image collections are employed, potentially missing numerous unseen visual connections.

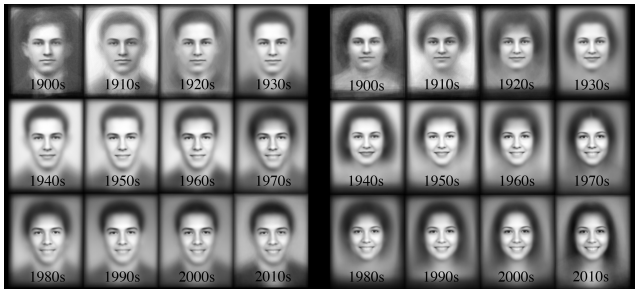

Figure 1: Average images of students by decade. The evolving fashions and facial expression throughout the 20th century are evident in this simple aggregation. For example, notice the increasing extent of smiles over the years and the tendency in recent years for women to wear their hair long. In contrast, note that the suit is the default dress code for men throughout the 20th century.

We take first steps towards a new approach to the analysis of visual historical data using data-driven methods suited to mining large image collections by creating a large visual historical dataset that can support such methods. By treating large historical photo collections as a whole, we expect to learn things that cannot be inferred from the inspection of a small number of artifacts in isolation. Similar approaches have been applied to the study of historical texts [20], but we are unaware of analyses of visual historical data.

In this paper we present a collection of one particular type of widely available yet little used historical visual data—a century's worth of high school yearbooks from around the United States (Fig 1). Yearbooks published since the wide adoption of film (the first Kodak camera was released in 1888) have contained standardized portrait photos

of the graduating class. As such, yearbook portraits provide a consistent visual format through which one can examine changes in the content ranging from personal style choices to developing social norms.

The main contributions of this paper are 1) A historical image dataset that comprises a large scale collection of yearbook portraiture from the last 120 years in the United States and which we make publicly-available. 2) An initial demonstration of the application of data-driven methods to discover historical visual patterns. In particular, we explore the gradual changes in social norms of smiling for the camera and the defining styles of different decades. 3) Finally, we demonstrate the use of the new dataset for training a deep learning algorithm for the task of image dating.

2. Related Work

Researchers in the humanities are now able to tease out historical information from large text corpora thanks to advances in natural language processing and information retrieval. For example, these advances (together with the availability of large-scale storage and OCR technology) enabled Michel et al. to conduct a thorough study of about 4% of all books ever printed resulting in a quantitative analysis of cultural and linguistic trends [20]. Large historical image collections will enable researchers to conduct similar analyses of visual historical trends.

To date, the study of historical images has been relatively limited. Some examples include modeling the evolution over time of automobile design [19] and architecture [18] as well as dating historical color photographs [21]. Here we extend upon these works by presenting a historical dataset that can be used to answer a broader set of questions.

Several researchers recently focused on modeling fashion items. In HipsterWars, Kiapour et al. take a supervised approach and use an online game to crowd-source human annotations of five current clothing-style categories that are then used to train models for style classification [15]. Hidayati et al. take a weakly-supervised approach to discover the recent (2010-2014) trends in the New York City fashion week catwalk shows [12]. They extract color and texture features and use these to discover the representative visual style elements of each season via discriminative clustering in an approach similar to that which Doersch et al. took to discover architectural elements [3]. While we also deal with fashion and style in this paper, our focus is on changes in style through a much longer period of history. Because our dataset includes scanned images from earlier time periods, much of it consists of lower resolution and quality images than the recent datasets described above. This makes some of the above approaches unsuitable for our data.

Finally, Islam et. al. analyze the connection between facial appearance and geolocation [13].

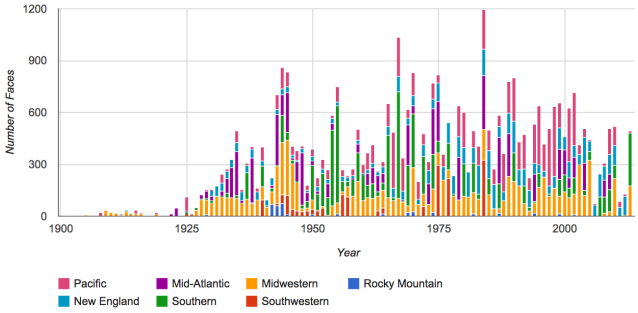

Figure 2: The distribution of portraits per year and region.

3. The Yearbook Dataset

We are at an auspicious moment for collecting historical yearbooks as it has become standard in recent years for local libraries to digitally scan their yearbook archives. This trend enabled us to download publicly available yearbooks from various online resources such as the Internet Archive and numerous local library websites. We collected 949 scanned yearbooks from American high schools ranging from 1905-2013 across 128 schools in 27 states. These contain 154,976 individual senior-class portrait photographs in total along with many more underclassmen portraits that were not used in this project. After removing all non-frontal facing we were left with a dataset of 37,921 photographs that depict individuals from 814 yearbooks across 115 high schools in 26 states.

On average, 28.8 faces are included in the dataset from each yearbook with an average of 329 faces per school across all years. The distribution of photographs over year and region is depicted in Figure 2. Overall, 46.4% of the photos come from the 100 largest cities according to US census [7].

As no dataset is bias-free, let us consider the potential biases in our data sample as compared to the high school age population of the United States. Since 1902 America's high schools have followed a standard format in terms of the population they served [9]. Yet, this does not mean that the population of high school students has always been an unbiased sample of the youth population in the US. In the early 1900s, less than 10% of all American 18-year-olds graduated from high school, but by end of the 1960s graduation rates increased to almost 50% [9]. Moreover, the standardization of high schools in the United States left out most of the African American population, especially in the South, until the middle of the 20th century [10].

In our dataset 53.4% of the photos are of women, and 46.6% are of men. As the true gender proportion in the population is only available in a census year it is difficult to determine whether this is a bias in our data. However, the gender imbalance may be due to the fact that historically

girls are disproportionately more likely than boys to attend high school through graduation [9].

3.1. Data Preprocessing

In order to turn the raw yearbooks into an image dataset we performed several pre-processing operations. First, we manually identified the scanned pages that included senior-class portraits. After converting those pages to grayscale for consistency across years, we automatically detected and cropped all faces. We then extracted facial landmarks for each face and estimated its pose with respect to the camera using the IntraFace system [26]. This allowed us to filter out all images depicting students that were not facing forward. Next, we aligned all faces to the mean shape using an affine transform based on the computed facial landmarks. Finally, we divided the photos into those depicting males and females using an SVM in the whitened HOG feature space [1, 11] and resolved difficult cases by crowdsourcing a gender classification task on Mechanical Turk.

4. Mining the Visual Historical Record

We demonstrate the use of our historical dataset in answering questions of historical and social relevance.

4.1. The Quintessential Styles of Each Decade

The simplest visual-data summarization technique of facial composites dates back to the 1870s and is attributed to Sir Francis Galton [6]. Here we use this technique to organize the portraits chronologically. Figure 1 (first page) displays the pixel-mean of photographs of male and female students for each decade from the 1900s to the 2010s. These average images showcase the main modes of the popular fashions in each time period.

We can further examine each decade in more detail by asking what are the representative and visually discriminative features of that decade. These are the things that make us immediately recognize a particular style as "20s" or "60s", for example, and allow humans to effortlessly guess the decade in which a portrait was taken. They are also the things that are usually hard to put into writing and require a visual aid when describing; this makes them excellent candidates for data-driven methods.

We find the most representative women's styles in hair and facial accessories for each decade using a discriminative mode seeking algorithm [2] on yearbook portraits cropped to contain only the face and hair. Since our portraits are aligned, we can treat them as a whole rather than look for mid-level representative patches as has been done in previous work [2, 3]. The output of the discriminative mode seeking algorithm is a set of detectors and their detected portraits that make up the visual clusters for each decade. We sort these clusters according to how discriminative they are, specifically, how many portraits they contain in the top 20 detections from the target decade versus other decades. In order to ensure a good visual coverage of the target decade, we remove clusters that include in their top 60 detections more than 6 portraits (10%) that were already represented by a higher ranking cluster.

Figure 3 displays the four most representative women's hair and eyeglass styles of each decade from the 1930s until the 2000s. Each row corresponds to a visual cluster in that decade. The left-most entry in the row is the cluster average, and to its right we display the top 6 portrait detections of the discriminative detector that created the cluster. We only display a single woman from each graduating class in order to ensure that the affinity within each cluster is not due to biases in the data that result from the photographic or scanning artifacts of each physical yearbook. Looking at Figure 3, we get an immediate sense of the attributes that make each decade's style distinctive. Some of the emergent attributes are especially interesting since they would be hard to describe in words. For example, the particular style of curly bangs of the 40s or the "winged" flip hairstyle of the 60s [24]. Finding and categorizing these manually would be painstaking work. With our large dataset these attributes emerge from the data by using only the year-label supervision.

4.2. Smiling in Portraiture

These days we take for granted that we should smile when our picture is being taken; however, smiling at the camera was not always the norm. In her paper, Kotchemidova studied the appearance of smiles in photographic portraits using the traditional historical methods of analyzing sample images manually [16]. She reports that in the late 19th century people posing for photographs still followed the habits of painted portraiture subjects. These included keeping a serious expression since a smile was hard to maintain for as long as it took to paint a portrait. Also, etiquette and beauty standards dictated that the mouth be kept small – resulting in an instruction to "say prunes" (rather than cheese) when a photograph was being taken [16]. All of this changed during the 20th century when amateur photography became widespread. In fact, Kotchemidova suggests that it was the attempt to make photography ubiquitous and associate it with happy occasions like holidays and travel that led the photographic monopoly, Kodak, to educate the public through visual advertisements that the obvious expression one should have in a snapshot is a smile. This century-long advertisement campaign was a great success. By World War II, smiles were so widespread in portraiture that no one questioned whether photographs of the GIs sent to war should depict them with a smile [16].

To verify Kotchemidova's claims regarding the presence and extent of smiles in portrait photographs in a data-driven way, we devised a simple lip-curvature metric and applied

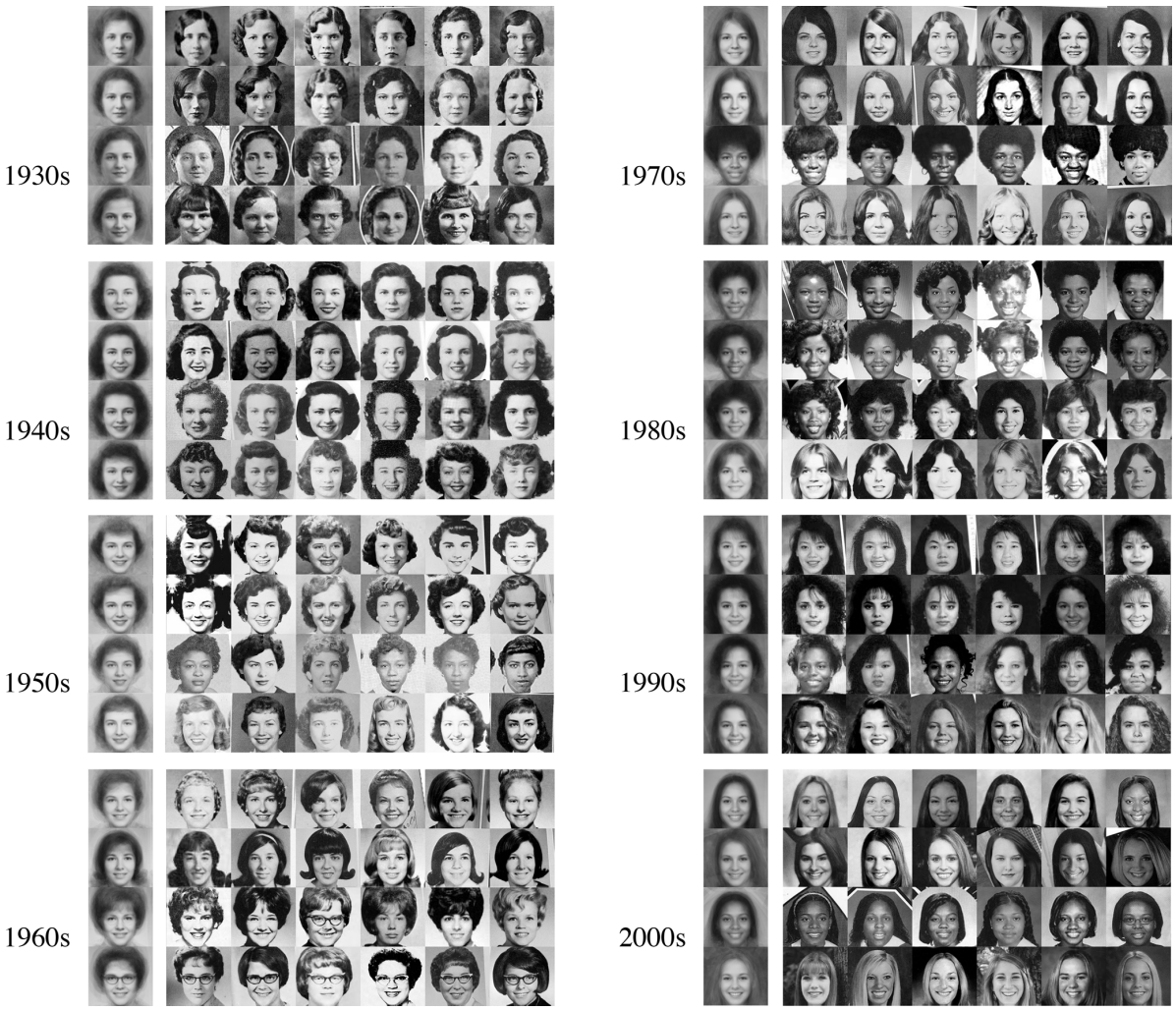

Figure 3: Discriminative clusters of high school girls' styles from each decade of the 20th century. Each row corresponds to a single detector and the cluster of its top 6 detections over the entire dataset. Only one girl per graduating class is shown in the top detections. The left-most entry in each row displays the cluster average. Note that the clusters correspond to the quintessential hair and accessory styles of each decade. Notable examples according to the Encyclopedia of Hair [24] are: The finger waves of the 30s. The pin curls of the 40s and 50s. The bob, "winged" flip, bubble cut and signature glasses of the 60s. The long hair, Afros and bouffants of the 70s. The perms and bangs of the 80s and 90s and the straight long hair fashionable in the 2000s. These decade-specific fashions emerge from the data in a weakly-supervised, data-driven process.

it to our dataset. We compute the lip curvature by taking the average of the two angles indicated in Figure 4 (Left) where the point that forms the hypotenuse of the triangle is the midpoint between the bottom of the top lip and the top of the bottom lip of the student. The same facial keypoints were used here as in the image alignment process (see section 3.1). Figure 4 (Right) is a montage of students ordered in ascending order of lip curvature value from left to right. It demonstrates that the lip-curvature metric quantifies the smile intensities in our data in a meaningful way.

We verify that our metric generalizes beyond yearbook portraits by testing it on the BP4D-Spontaneous dataset that contains images of participants showing various degrees of facial expressions with ground truth labels of expression intensity [27]. BP4D uses the Facial Action Coding System, commonly used in facial expression analysis, for ground truth annotations [4]. This coding system consists of Action Units (AU) which correspond to the intensity of contraction of various facial muscles. Following previous work done on smile intensity estimation [8], we compared our smile intensity metric with the activation of AU12 (Lip corner puller) as it corresponds to the contraction of muscles that raise the

Figure 4: Smile intensity metric. Left: the lip curvature metric is the average of the two marked angles. Right: women and men portraits sorted by increasing lip curvature.

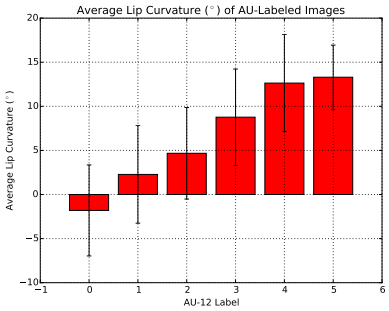

Figure 5: Average lip curvature correlates with AU-12 labels on BP4D data (error bars denote standard deviation).

corners of the mouth into a smile. A higher AU12 value represents a higher contraction of muscles around the corner of the mouth, resulting in a larger smile. Figure 5 displays the average lip curvature for each value of AU12 for 3 male and 3 female subjects in the dataset, corresponding to 2,500-3,000 samples for each AU12 value (0-5). As the simple lip-curvature metric we used correlates with increasing AU12 values on BP4D images, it is a decent indicator for smile intensities beyond our yearbook dataset.

Using our verified lip-curvature metric we plot the trend of average smile intensities in our data over the past century in Figure 6. Corresponding montages of smile intensities over the years are included in Figure 7, where we picked the student with the smile intensity closest to the average for each 10-year bucket from 1905 to 2005. These figures corroborate Kotchemidova's theory and demonstrate the rapid increase in the popularity and intensity of smiles in portraiture from the 1900s to the 1950s, a trend that still continues today; however, they also reveal another trend—women significantly and consistently smile more than men. This phenomenon has been discussed extensively in the literature (see the meta-review in [17]), but until now required intensive manual annotation in order to discover and analyze. For example, in her 1982 article Ragan manually analyzed 1,296 high school and university yearbooks and media files in order to reveal a similar result [22]. By use of a large historical data collection and a simple smile-detector we arrived at the same conclusion with a minimal amount of annotation and virtually no manual effort.

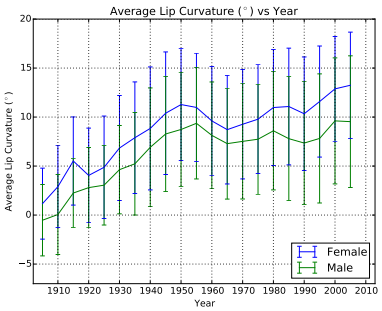

Figure 6: Smiles increasing over time, but women always smile more than men: Male and female Average lip curvature by year with one standard deviation error bars. Note the fall in smile extent from the 50s to the 60s, for which we did not find prior mention.

Figure 7: Images selected as having the closest smile to the mean of that period (10-year bins from 1905 (left) to 2005 (right)). Note the increasing extent of smiles over time.

5. Dating Historical Images

One practical use for such a large historical dataset is to develop models for dating historical images. We extend the work of Palermo et al. [21] in dating color photographs to the realm of black and white portraiture photography where we cannot rely on the changes in image color profiles over time. We chose to train a deep neural network classification model for dating photographs based on the recent success of such models in other areas of computer vision. Here we pose the task of dating the portraits of female students as an 83-way year-classification task between the years 1928 and 2010, for which we have more than 50 female images per year. Again, we use portraits that were cropped to the face and hair alone. We set aside 20% of the portraits taken between 1982 and 2010 as the *yearbook test set* and use the remaining 80% for training. We deliberately exclude from the test set images from the schools we train on within a period of 10 years, in order to minimize photographic and scanning training biases. To minimize training biases due to photographic and scanning artifacts, we separate test and training images drawn from the same school by at least a decade. To further minimize these biases, we use the built-in Photoshop noise reduction filter on all the yearbook images and resize them to 96 by 96 pixels. To evaluate the generalizability of our fine-tuned models to non-yearbook portraits,

Model	Yearbook Accuracy	Yearbook L1 Med	Celebrity Accuracy	Celebrity L1 Med
Chance	1.20%	-	1.20%	-
Baseline	6.18%	6 [yr]	1.79%	14.50 [yr]
FT on YB	11.31%	4 [yr]	1.79%	9 [yr]

Table 1: Classification accuracy and L1 average and median distance from the ground truth year for the yearbook and celebrity test sets. Note that fine-tuning on yearbooks improves the classification results on the yearbook test data, but only improves the L1 median distance between the predictions and the ground truth for the celebrity dataset.

we further evaluate them on the *celebrity test set* – a small test set of 56 gray-scale head shots of female celebrities, annotated with year labels, that we cropped and aligned to the yearbook images.

We use the Caffe [14] implementation of the VGG network architecture (modified to allow for $96px$ inputs) [25] that was pre-trained on the ILSVRC dataset [23] in all our experiments. As a baseline, we *fine-tune* only the last classification layer (fc_8) of the ILSVRC-trained network on our yearbook training data to predict the year at which a yearbook photograph was taken in an 83-way classification task. In Table 1 we refer to this result as the *baseline*. We compare this baseline to the classification performance of the same network that we trained by *fine-tuning* **all** the layers for the same classification task. We train the network in both conditions for $100K$ iterations using SGD with image-mirroring during training, learning rate of 0.001 ($\gamma = 0.1$, stepsize $= 20K$) and momentum of 0.9. As expected, fine-tuning on the yearbook data improves the classification accuracy on the yearbook test set by a large margin. We further compare our classification performance to *chance*, which we define as the inverse of the number of classes. The confusion matrix for the fine-tuned classification model on the yearbook test set is shown in Figure 8. The diagonal structure of the matrix indicates that most of the confusion occurs between neighboring years which matches our expectations that visual trends transcend the single-year boundary.

Given the success in dating yearbook portraits, we try using our model to date the images in the celebrity test set. Unfortunately, the classification model which was fine-tuned to the yearbook data does not generalize well to the images in the celebrity test set. This may be because celebrity glamour shots may not be the best validation set for portraiture dating as celebrity hairstyles can be quite different than those of the general public, or because our celebrity test set is simply too small. However, we do find that fine-tuning on the yearbook data reduces the median L1 distance between the predicted and ground truth year for the celebrity portraits (Table 1).

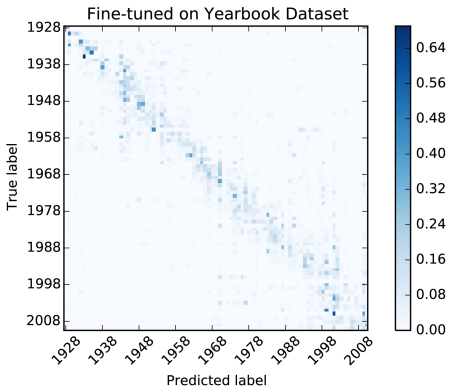

Figure 8: Normalized soft confusion matrix, fine-tuned and tested on yearbooks. The diagonal structure demonstrates that confusion mostly occurs between neighboring years.

6. Conclusion

In this paper, we presented a large-scale historical image dataset of yearbook portraits, which we have made publicly available. These provide us with a unique opportunity to observe how styles and portrait-posing habits change over time in a restricted, fixed visual framework. We demonstrated the use of various techniques for mining visual patterns and trends in the data that significantly decrease the time and effort needed to arrive at the type of conclusions often researched in the humanities. Moreover, we showed how this dataset can be used along with deep learning techniques to date black and white portraits.

Much remains to be done with visual historical datasets, and in particular the one at hand. For example, historical yearbook portraits can be used to discover the cycle-length of fashion fads and can be used as a basis of data-driven style transfer algorithms. In addition, while our dating results are promising for similarly posed yearbook portraits, generalizing our models to other types of portraits remains for future work. Ultimately, we believe that the use of large-scale historical image datasets such as ours in conjunction with data-driven methods, can radically change the methodologies in which visual cultural artifacts are employed for humanities research.

7. Acknowledgments

The authors would like to thank Bharath Hariharan, Carl Doersch and Evan Shelhamer for their insightful comments. This material is based upon work supported by the NSF Graduate Research Fellowship DGE 1106400, ONR MURI N000141010934 and an NVidia hardware grant.

References

[1] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, pages 886–893, 2005.

[2] C. Doersch, A. Gupta, and A. A. Efros. Mid-level visual element discovery as discriminative mode seeking. In *Neural Information Processing Systems (NIPS)*, 2013.

[3] C. Doersch, S. Singh, A. Gupta, J. Sivic, and A. A. Efros. What makes paris look like paris? *SIGGRAPH*, 31(4), 2012.

[4] P. Ekman and W. V. Friesen. Facial action coding system: A technique for the measurement of facial movement, 1978.

[5] Flappers flaunt fads in footwear. *The New York Times*, page 34, Sunday, January 29, 1922.

[6] F. Galton. Composite portraits made by combining those of many different persons into a single figure. *Nature*, 18(447):97–100, 1878.

[7] C. Gibson. Population of the 100 largest cities and other urban places in the united states: 1790 to 1990. `https://www.census.gov/population/www/documentation/twps0027/twps0027.html`. Accessed: 2014-12-11.

[8] J. M. Girard. *Automatic Detection and Intensity Estimation of Spontaneous Smiles*. PhD thesis, University of Pittsburgh, 2014.

[9] C. Goldin. America's graduation from high school: The evolution and spread of secondary schooling in the twentieth century. *Journal of Economic History*, 1998.

[10] C. Goldin and L. F. Katz. The race between education and technology: The evolution of u.s. educational wage differentials, 1890 to 2005. Working Paper 12984, National Bureau of Economic Research, March 2007.

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

[12] S. C. Hidayati, K.-L. Hua, W.-H. Cheng, and S.-W. Sun. What are the fashion trends in new york? In *Proceedings of the ACM International Conference on Multimedia*, MM '14, pages 197–200, New York, NY, USA, 2014. ACM.

[13] M. T. Islam, C. Greenwell, R. Souvenir, and N. Jacobs. Large-Scale Geo-Facial Image Analysis. *EURASIP Journal on Image and Video Processing (JIVP)*, 2015(1):14, 2015.

[14] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.

[15] M. H. Kiapour, K. Yamaguchi, A. C. Berg, and T. L. Berg. Hipster wars: Discovering elements of fashion styles. In *ECCV (1)*, pages 472–488, 2014.

[16] C. Kotchemidova. Why we say "cheese": Producing the smile in snapshot photography. *Critical Studies in Media Communication*, 22(1):2–25, 2005.

[17] M. Lafrance, M. A. Hecht, and E. L. Paluck. The contingent smile: A meta-analysis of sex differences in smiling. *Psychological Bulletin*, pages 305–334, 2003.

[18] S. Lee, N. Maisonneuve, D. Crandall, A. Efros, and J. Sivic. Linking past to present: Discovering style in two centuries of architecture. In *IEEE International Conference on Computational Photography (ICCP)*, 2015.

[19] Y. J. Lee, A. A. Efros, and M. Hebert. Style-aware mid-level representation for discovering visual connections in space and time. In *ICCV*, pages 1857–1864, 2013.

[20] J.-B. Michel, Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, T. G. B. Team, J. P. Pickett, D. Holberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M. A. Nowak, and E. L. Aiden. Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014):176–182, 2010.

[21] F. Palermo, J. Hays, and A. A. Efros. Dating historical color images. In *ECCV (6)*, pages 499–512, 2012.

[22] J. M. Ragan. Gender displays in portrait photographs. *Sex Roles*, 8(1):33–43, 1982.

[23] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge, 2014.

[24] V. Sherrow. *Encyclopedia of Hair: A Cultural History*. Greenwood Press, 2006.

[25] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

[26] X. Xiong and F. De la Torre. Supervised descent method and its applications to face alignment. In *CVPR*, 2013.

[27] X. Zhang, L. Yin, J. F. Cohn, S. Canavan, M. Reale, A. Horowitz, P. Liu, and J. M. Girard. BP4D-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database. *Image and Vision Computing*, 32(10):692–706, 2014.