"When Was This Picture Taken?" – Image Date Estimation in the Wild

Eric Müller[1(✉)], Matthias Springstein[1], and Ralph Ewerth[1,2]

[1] German National Library of Science and Technology (TIB), Hannover, Germany
{eric.mueller,matthias.springstein,ralph.ewerth}@tib.eu
[2] L3S Research Center, Hannover, Germany

Abstract. The problem of automatically estimating the creation date of photos has been addressed rarely in the past. In this paper, we introduce a novel dataset *Date Estimation in the Wild* for the task of predicting the acquisition year of images captured in the period from 1930 to 1999. In contrast to previous work, the dataset is neither restricted to color photography nor to specific visual concepts. The dataset consists of more than one million images crawled from Flickr and contains a large number of different motives. In addition, we propose two baseline approaches for regression and classification, respectively, relying on state-of-the-art deep convolutional neural networks. Experimental results demonstrate that these baselines are already superior to annotations of untrained humans.

1 Introduction

In recent years, huge datasets (e.g., *ImageNet* [8], *YFCC100M* [12]) were introduced fostering research for many computer vision tasks. In particular, such datasets are a prerequisite for the training of deep learning systems. However, estimating automatically the capturing time of (historical) photos has been rarely addressed yet and existing benchmark datasets do not contain enough images captured before 2000. But date estimation is an interesting and challenging task for historians, archivists, and even for sorting (digitized) personal photo collections chronologically. Existing approaches either rely on datasets solely containing historical color images [1,6,7] or focus on specific concepts like cities [10], cars [4], persons [2,9], or historical documents [3,5] and are therefore unable to learn the temporal differences of the broad variety of motives. For this reason, a huge dataset covering all kinds of concepts is necessary, which additionally enables the training of convolutional neural networks.

In this paper, we introduce a novel dataset *Date Estimation in the Wild* and make it publicly available to support further research. In contrast to existing datasets, it contains more than one million Flickr images captured in the period from 1930 to 1999. As shown in Fig. 1, the dataset covers a broad range of domains, e.g., city scenes, family photos, nature, and historical events. Two baseline approaches are proposed based on a deep convolutional neural network (GoogLeNet [11]) treating the task of dating images as a classification and

J.M. Jose et al. (Eds.): ECIR 2017, LNCS 10193, pp. 619–625, 2017.
DOI: 10.1007/978-3-319-56608-5_57

Fig. 1. Some example images from the *Date Estimation in the Wild* dataset.

regression problem, respectively. Experimental results show the feasibility of the suggested approaches which are superior to annotations of untrained humans.

The remainder of the paper is organized as follows. Section 2 reviews related work on dating historical images. Section 3 introduces the *Date Estimation in the Wild* dataset as well as the baseline approaches in detail. The experimental setup and results are presented in Sect. 4 along with a comparison to human annotation performance. Section 5 concludes the paper.

2 Related Work

The first work that deals with dating historical images stemming from different decades has been introduced by Schindler et al. [10]. The authors present an approach to sort a collection of city-scape images temporally by reconstructing the 3D world, requiring many overlapping images of the same location. Jae et al. [4] identify style-sensitive groups of patches for cars and street view images in order to model stylistic differences across time and space. He et al. [3] and Li et al. [5] address the task of estimating the age of historical documents. While He et al. [3] explore contour and stroke fragments, Li et al. [5] apply convolutional neural networks in combination with optical character recognition. Ginosar et al. [2] and Salem et al. [9] model the differences of human appearance and clothing style in order to predict the date of photos in yearbooks.

More closely related to our work, Palermo et al. [7] suggest an approach to automatically estimate the age of historical color photos without restrictions to specific concepts. They combine different color descriptors to model the historical color film processes. The results on the proposed dataset, which contains 1375 images from 1930 to 1980, are further improved by Fernando et al. [1] by including color derivatives and angles. Martin et al. [6] treat date estimation as a

binary task by deciding whether an image is older or newer than a reference image. However, the aforementioned approaches either rely on color photography, which was very uncommon before 1970, or focus on specific concepts.

3 Image Date Estimation in the Wild

In this section, the *Date Estimation in the Wild* dataset (Sect. 3.1) and the two proposed baseline approaches to predict the acquisition year of images (Sect. 3.2) are described in detail.

3.1 Image Date Estimation in the Wild Dataset

The Flickr API was utilized to download images for each year of the period from 1930 to 1999. We have observed that many historical images are supplemented with time information, either in the title or in the related tags and descriptions. Therefore, we used the current year as an additional query term to reduce the number of "spam" images. The only kind of filtering that we applied was restricting the search to photos. As a consequence, the dataset is noisy since it contains, for example, close-ups of plants or animals as well as historical documents. In order to avoid a bias towards more recent images, the maximum number of images per year was limited to 25000. Finally, the dataset consists of 1029710 images with a high diversity of concepts. Information about the granularity $g \in \{0, 4, 6, 8\}$ according to the Flickr annotation of the date entry is stored as well. The distribution of images per year and the related granularity of dates are depicted in Fig. 2.

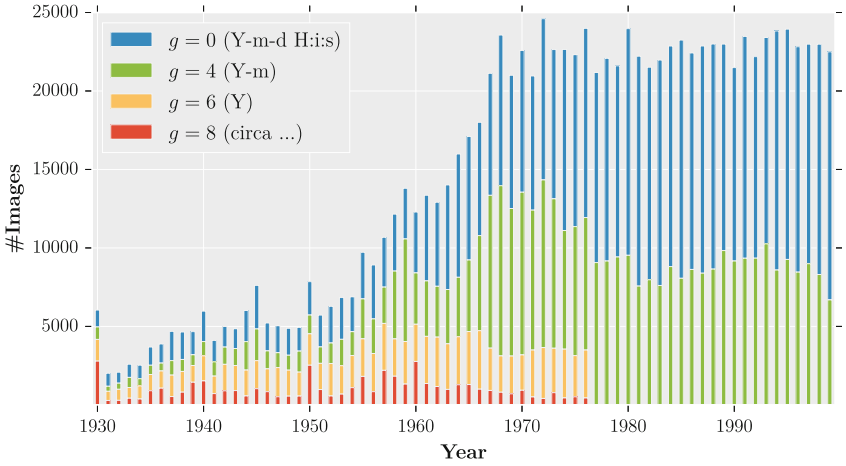

Fig. 2. Number of crawled images and the accuracy of the provided timestamps for each year in the *Date Estimation in the Wild* dataset.

In order to obtain reliable validation and test sets that match the dataset distribution, a maximum number of 75 *unique images* for 1930 to 1954 and 150 *unique images* for the remaining years were extracted. A *unique image* is defined as an image with a date granularity of $g = 0$ (Y-m-d H:i:s) or $g = 4$ (Y-m), for which no visual near-duplicates (detected by comparing the features from the last pooling layer of a GoogLeNet pre-trained on *ImageNet*) exist in the entire dataset. Subsequently, 8495 *unique images* were extracted for the validation set and another 16 per year were selected manually to obtain the test dataset containing 1120 images. The remaining 1020095 images constitute the training set. The dataset[1] is available at https://doi.org/10.22000/0001abcde.

3.2 Baseline Approaches

Two baseline approaches are realized by training a GoogLeNet [11] and treating image date estimation as a classification or regression problem, respectively.

Convolutional neural networks require many images per class c to learn appropriate models for the classification task. However, the dataset lacks images for the first three decades (Fig. 2). For this reason, we decided to use $|c| = 14$ classes by quantizing the image acquisition year into 5-year periods to reduce the classification complexity, while still maintaining a good temporal resolution. For the classification task, GoogLeNet was trained using Caffe on a pre-trained *ImageNet* model [8]. We randomly selected 128 images per batch for training, which were scaled by the ratio $256/\min(w, h)$ (w and h are image dimensions). To augment training data, the images were horizontally flipped and cropped randomly to fit in the reception field of $224 \times 224 \times 3$ pixel. The stochastic gradient descent algorithm was employed using 1M iterations with a momentum of 0.9 and a base learning rate of 0.001 to reduce the classification loss. The weights of the fully connected (fc) layers are re-initialized and their corresponding learning rates are multiplied by 10. The output size of the fc layers is set to the number of classes and the learning rates were reduced by a factor of 2 every 100k iterations.

Test images are scaled by the ratio $224/\min(w, h)$ and three 224×224 pixel regions depending on the images' orientations are passed to the trained model. To estimate a specific acquisition year y_E, the averaged class probabilities $p(c)$ of the three crops for each class $c \in [0, 13]$ are interpolated by:

$$y_E = 1930 + \left\lfloor 0.5 + \frac{1999 - 1930}{|c| - 1} \cdot \sum_{i=0}^{|c|-1} i \cdot p(i) \right\rfloor, \quad \text{with} \sum_{i=0}^{|c|-1} p(i) = 1. \tag{1}$$

For the regression task, the Euclidean loss between the predicted and ground truth image date was minimized. We used the same parameters for learning as in classification except for: The base learning rate was reduced to 0.0001 and a bias of 1975 (middle year) for the fc layers was used to stabilize training. Finally, the output size was set to 1 for regression to directly predict the year.

[1] Images or links (depending on the copyright status) and metadata are provided.

4 Experimental Results

In the experiments, the trained GoogLeNet models were applied to the test set. In contrast to Palermo et al. [7], we do not report the classification accuracy for predicting the correct 5-year period. For example, imagine that the ground truth date of an image is 1989 and the model predicts the class 1990–1994. Although the difference is possibly only one year the prediction would be false in this case. For this reason, we argue that the absolute mean error (ME) as well as the number of images with an absolute estimation error of at most n years (EE_n) are more meaningful for evaluation.

Table 1. Absolute mean error (ME) [y] and number of images estimated with an absolute estimation error of at most n years (EE_n) [%] for human annotators and for the baselines GoogLeNet classification (cls) and regression (reg) approaches on the *Date Estimation in the Wild* test set, with respect to each quantized 5-year period.

	Human performance				GoogLeNet cls				GoogLeNet reg			
Year	ME	EE_0	EE_5	EE_{10}	ME	EE_0	EE_5	EE_{10}	ME	EE_0	EE_5	EE_{10}
30–34	15.7	3.0	24.8	40.7	15.0	0.0	5.0	37.5	14.4	0.0	7.5	41.3
35–39	12.2	2.7	34.1	53.2	11.1	2.5	23.8	52.5	10.7	3.8	26.3	58.8
40–44	9.6	4.1	43.2	66.6	8.8	2.5	40.0	67.5	9.1	7.5	42.5	66.3
45–49	11.7	3.9	31.1	54.3	8.2	6.3	51.3	71.3	8.5	3.8	43.8	70.0
50–54	12.2	2.5	29.6	49.8	7.5	3.8	47.5	77.5	7.3	2.5	52.5	73.8
55–59	13.3	1.4	27.1	49.5	6.1	6.3	60.0	86.3	7.0	7.5	50.0	77.5
60–64	13.6	1.4	24.1	43.0	7.3	5.0	51.3	73.8	7.2	1.3	47.5	75.0
65–69	12.5	2.7	24.6	46.4	5.4	12.5	63.8	82.5	6.0	1.3	52.5	83.8
70–74	10.5	4.8	33.2	55.9	5.6	3.8	58.8	85.0	5.4	8.8	61.3	85.0
75–79	9.4	4.1	37.9	62.1	4.7	8.8	71.3	90.0	5.0	7.5	63.8	90.0
80–84	7.5	5.2	45.5	76.1	4.4	8.8	62.5	95.0	4.5	6.3	61.3	93.8
85–89	7.6	5.0	49.6	77.3	4.8	10.0	71.3	83.8	4.9	8.8	68.8	90.0
90–94	7.5	5.9	51.3	76.1	5.6	5.0	66.3	85.0	5.7	6.3	61.3	83.8
95–99	9.4	6.1	39.5	62.9	7.5	11.3	52.5	75.0	8.7	1.3	36.3	73.8
Overall	10.9	3.8	35.4	58.1	7.3	6.2	51.8	75.9	7.5	4.7	48.2	75.9

Human performance was investigated as well. Seven untrained annotators of different age (ranging from 26 to 58) were asked to label all 1120 images of the test set and to make a break after each batch of 100 images. The average human performance and the results of our baseline approaches are displayed in Table 1.

The results clearly show the feasibility of our baselines outperforming human annotations in nearly all periods and reducing the mean error by more than three years on the entire dataset. Another observation is that there is a correlation between the number of images and the results for each 5-year period. For this

reason, an increased mean error for images between 1930 to 1964 is noticeable. Besides, the potential error can be higher for classes at the interval boundaries (1930 and 1999), which explains the slightly worse results for 1990 to 1999. A similar observation can be made for human annotations, since they are more familiar with images, TV material, and their own experiences starting from 1960. Interestingly, the human error is noticeably lower for images covering the period from 1940 and 1944, which frequently show scenes from World War II.

Despite the problem caused by the interval bounds of the entire time period which affects the interpolation step, the classification results are slightly better than for regression. This is attributed to the easier task of minimizing the classification loss of 14 classes compared to minimizing the Euclidean loss.

5 Conclusions

In this paper, we have introduced a novel dataset entitled *Date Estimation in the Wild* to foster research regarding the challenging task of image date estimation. In contrast to previous work, the dataset is neither restricted to color imagery nor to specific concepts, but includes images covering a broad range of motives for the period from 1930 to 1999. In a first attempt to tackle this challenging problem, we have proposed two approaches relying on deep convolutional neural networks to predict an image's acquisition year, considering the task as a classification as well as a regression problem. Both approaches achieved a mean error of less than 8 years and were superior to annotations of untrained humans. In the future, it is planned to exploit different specific classifiers for frequent concepts such as persons or cars to further enhance the performance of our systems.

References

1. Fernando, B., Muselet, D., Khan, R., Tuytelaars, T.: Color features for dating historical color images. In: IEEE International Conference on Image Processing, pp. 2589–2593 (2014)
2. Ginosar, S., Rakelly, K., Sachs, S., Yin, B., Efros, A.A.: A century of portraits: a visual historical record of American high school yearbooks. In: IEEE International Conference on Computer Vision Workshops, pp. 1–7 (2015)
3. He, S., Samara, P., Burgers, J., Schomaker, L.: Image-based historical manuscript dating using contour and stroke fragments. Pattern Recogn. **58**, 159–171 (2016)
4. Jae Lee, Y., Efros, A.A., Hebert, M.: Style-aware mid-level representation for discovering visual connections in space and time. In: IEEE International Conference on Computer Vision, pp. 1857–1864 (2013)
5. Li, Y., Genzel, D., Fujii, Y., Popat, A.C.: Publication date estimation for printed historical documents using convolutional neural networks. In: Third International Workshop on Historical Document Imaging and Processing, pp. 99–106 (2015)
6. Martin, P., Doucet, A., Jurie, F.: Dating color images with ordinal classification. In: International Conference on Multimedia Retrieval, pp. 447–450 (2014)
7. Palermo, F., Hays, J., Efros, A.A.: Dating historical color images. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7577, pp. 499–512. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33783-3_36

8. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., et al.: Imagenet large scale visual recognition challenge. Int. J. Comput. Vis. **115**(3), 211–252 (2015)
9. Salem, T., Workman, S., Zhai, M., Jacobs, N.: Analyzing human appearance as a cue for dating images. In: 2016 IEEE Winter Conference on Applications of Computer Vision, pp. 1–8 (2016)
10. Schindler, G., Dellaert, F., Kang, S.B.: Inferring temporal order of images from 3D structure. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–7 (2007)
11. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., et al.: Going deeper with convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
12. Thomee, B., Shamma, D.A., Friedland, G., Elizalde, B., Ni, K., Poland, D., Borth, D., Li, L.J.: YFCC100M: the new data in multimedia research. Commun. ACM **59**(2), 64–73 (2016)