Determining Image Age with Rank-Consistent Ordinal Classification and Object-centered Ensemble

Shota Ashida
shota.ashida@db.soc.i.kyoto-u.ac.jp
Kyoto University
Kyoto, Japan

Adam Jatowt
jatowt@acm.org
Kyoto University
Kyoto, Japan

Antoine Doucet
antoine.doucet@univ-lr.fr
University of La Rochelle
La Rochelle, France

Masatoshi Yoshikawa
yoshikawa@i.kyoto-u.ac.jp
Kyoto University
Kyoto, Japan

ABSTRACT

A significant number of old photographs including ones that are posted online do not contain the information of the date at which they were taken, or this information needs to be verified. Many of such pictures are either scanned analog photographs or photographs taken using a digital camera with incorrect settings. Estimating the date of such pictures is useful for enhancing data quality and its consistency, improving information retrieval and for other related applications. In this study, we propose a novel approach for automatic estimation of the shooting dates of photographs based on a rank-consistent ordinal classification method for neural networks. We also introduce an ensemble approach that involves object segmentation. We conclude that assuring the rank consistency in the ordinal classification as well as combining models trained on segmented objects improve the results of the age determination task.

KEYWORDS

image dating, ordinal classification instance segmentation

ACM Reference Format:
Shota Ashida, Adam Jatowt, Antoine Doucet, and Masatoshi Yoshikawa. 2021. Determining Image Age with Rank-Consistent Ordinal Classification and Object-centered Ensemble. In *ACM Multimedia Asia (MMAsia '20), March 7–9, 2021, Virtual Event, Singapore.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3444685.3446326

1 INTRODUCTION

Social network services such as Instagram and Flickr allow users to upload and share their images online. The popularity of social networking sites is accelerating the growth in image uploads, with Instagram having already over 1 billion monthly active users and Facebook having approximately 2.6 billion monthly active users (March 2020). It is easy to imagine that these large numbers of users of social networking sites translate into huge amount of image uploads. Furthermore, nowadays, many memory institutions such as museums and libraries digitize their collections and make them available online for the public. For example, the Smithsonian Institution in the U.S. has digitized more than 40,000 items from its collection and made them available on the Web. Flickr created an archive of photos to preserve historically significant images within the frame of a project called The Commons[1] that has been jointly implemented with the Library of Congress. The Commons gathers old photos from around the world to create a collection of historical photos accessible to everyone.

If an image is shared by a museum or library, its detailed metadata information (e.g., who created the image, when it was created, what is portrays) are typically clearly stated. Similarly, most images uploaded to social networking sites are digital images accompanied by metadata (e.g., the device on which the image was taken, the date and place of taking the image). However, there are some images that lack information of their creation dates. For example, scanned born-analog photos or photos taken by a digital camera with incorrect settings[2] may not contain such information. Furthermore, sometimes one needs to verify the credibility of known creation dates (e.g., as a part of curating process of large collections of historical photos), which could be imprecise or could have been mistakenly altered.

In this work, we focus on predicting a year when an input image was created at. There are two main benefits of the creation date estimation. The first one is improving the quality and consistency of the data and enabling extended search functionalities. Consider a collection of photos portraying various real-world objects. We would not be able to chronologically arrange this dataset by the age of objects in the photos without known temporal metadata. However, if one can precisely estimate the year of shooting of each photo, then arranging the displayed objects by their age becomes possible or is at least facilitated. It should be also possible to realize temporal search mechanism for image collections that would satisfy requests such as "New York in the 1930s" or "popular car models in 1950s". In other words, the creation date estimation could improve the data value and enhance the range of possible operations.

MMAsia '20, March 7–9, 2021, Virtual Event, Singapore

ACM ISBN 978-1-4503-8308-0/21/03…$15.00
https://doi.org/10.1145/3444685.3446326

[1] https://www.flickr.com/commons
[2] Some users do not change the default date setup of 1 January 1970.

The second benefit is the curation and the protection of historically significant photographs. Old photographs can provide a variety of information about the past. Buildings, people, fashion, food and other objects in photos can tell us much about the culture, technology and customs in the past years. Date estimation of images is an important tool for supporting the curation, protection and utilization of old photographs. For example, knowing an image age should help to better assess its historical importance and uniqueness.

In this paper we improve the effectiveness of image dating by introducing two approaches for this task. In particular, we first propose using Rank-consistent Ordinal Classification method [1] for image creation date estimation. This approach allows aligning the results of the participating ordinal classifiers and removing inconsistencies in their predictions. Second, we introduce another approach which is based on the ensemble of object-centered age estimation models. It relies on training object-specific models and on combining them in ensemble scenarios for improving the estimation accuracy. The experimental evaluation using a dataset composed of 939,122 images shows that the proposed approaches can improve the results of the task of determining image age. The dataset that we use and codes are made freely available[3].

The remainder of the paper is structured as follows. In the next section we overview the related work. Section 3 describes our proposed methods. We discuss the experimental settings in Section 4 and the experimental results in Section 5. Finally, the last section concludes the paper.

2 RELATED WORK

Image age estimation is a relatively novel and understudied subfield of multimedia processing field. Initially, the amount of available data was too small to use neural network based approaches. Researches used to extract selected effective features and to use classifiers such as Support Vector Machine. Palermo *et al.* [13] extracted features such as the distribution of colors in the images and the characteristics of the cameras at different periods of time, solving the task of dating photographs as a traditional classification problem. In addition, they created an application that makes new photos look old and old photos look new by using the proposed method. Fernandor *et al.* [4] further extended the approach of Palermo *et al.* by adding color features, which take into account the light and shadow conditions of the shooting scene demonstrating the improvement in the classification accuracy. Martin *et al.* [9] employed an ordinal classification approach. For the K-ranked ordinal classification problem, $K-1$ classifiers were established to determine whether a photograph was taken before or after a certain year, and the year with the highest probability from the scores of all classifiers was used as the final estimate. However, the authors did not employ neural network approach, neither assured global consistency of individual predictions as we do.

Later some researchers tried to solve the task by gathering a large amount of data and using convolutional neural networks. Müller *et al.* [10] collected large numbers of online photographs from Flickr to create a new dataset which features various types of scenes, and used the GoogleNet [17] image recognition model to solve the image dating task as both multi-classification and regression problems. Shiry *et al.* [5] focused on human appearance. They collected 37,921 frontal-facing American high school yearbook photos and created a date estimation model using VGG. Salem *et al.* [16] also focused on a relation between human appearance and time. They found that shirt collars, glasses, and hair have a great impact on a date estimation.

A related research is also on estimating the age of persons shown in photos (typically by examining their faces). For example, Choon-Ching *et al.* [11] proposed a novel wrinkle representation and wrinkle-based age estimation approach. Haibin *et al.* [8] adopted factor analysis model to extract robust face features. In terms of age estimation function learning, the authors also proposed age-based and sequential study of rank-based age estimation learning methods and a divide-and-rule face age estimator. Niu *et al.* [12] apply ordinal regression to age estimation task with multiple output.

Other studies have also been carried out to support the field of Web image retrieval by estimating image temporal metadata from internal image features and from surrounding text [2]. In text and Web domains there were also several approaches proposed to estimate the age of text content [6, 14] or web content [7].

In the current work we use ordinal classification on the top of state-of-the-art neural network based image processing framework and assure rank consistency of ordinal classifiers in two directions for the image dating problem. We then use the rank consistent ordinal classification in combination with object-focused image dating models. To the best of our knowledge, we are the first to propose the concept of object segmentation to be used for the purpose of automatic photo time stamping.

3 METHODS

3.1 Rank-Consistent Ordinal Classification (CORAL)

Solving the date estimation task using classification is not very effective because classification does not consider the order relation between classes. Regression is also not the best choice because photography as well as the changes of photographed scenes and camera technologies over time were not subject to stationary changes. Due to these reasons, the dating task is more suitable for ordinal classification approaches. Ordinal classification is also known as ordinal regression and is used when there is some kind of order between answer labels or classes. It has been successfully applied for many objectives including estimating person age [1], judging the progression of a disease (e.g., Alzheimer's or Coulomb's disease) [3], and as a technique for advertising or recommendation [15], as well as for other purposes.

However ordinal classification can suffer from rank inconsistency when the decisions by participating classifiers are inconsistent, which may lead to sub-optimal performance. Fig. 1 shows example probabilities for an input image of exceeding particular labels (years) in the date estimation task. The left portion of the figure shows an output of a conventional rank-inconsistent ordinal classification while the right one shows an output of the rank-consistent one. Rank-consistent classification is clearly better because the probabilities should be decreasing or should remain same when the labels increase. The plot of the classification probabilities for

[3]https://github.com/ylab-public/image_dating

the rank consistency is not characterized by any peaks suggesting that rank consistency should provide more stability to the date estimation.

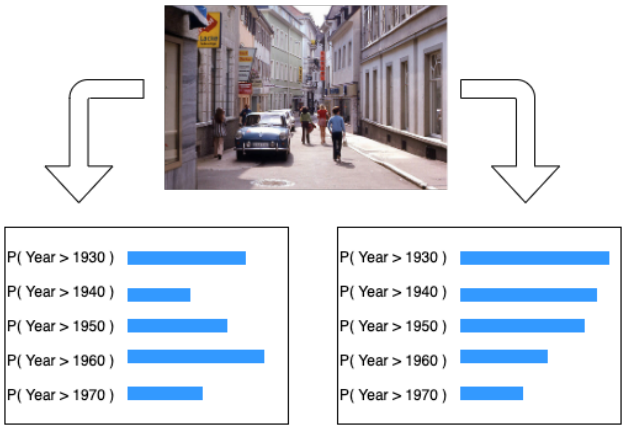

Figure 1: Examples of the output by rank inconsistent (left) and rank consistent (right) ordinal classification

To achieve the rank consistency in the ordinal classification, each classifier shares the weights of the fully connected layer in a CNN to satisfy the rank consistency unlike the conventional method (see Fig. 2). The mathematical proof is given in the paper [1].

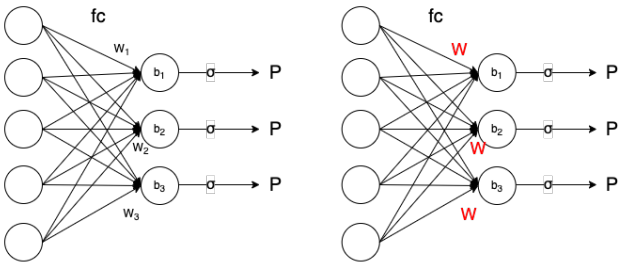

Figure 2: The left figure shows a usual fully connected layer and the right one shows the one with shared weights

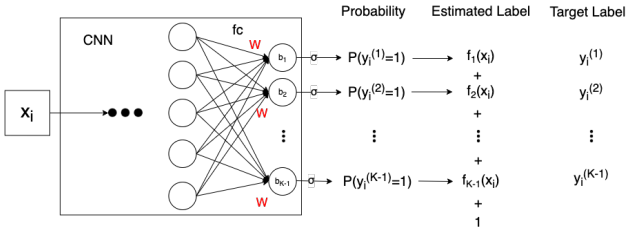

Figure 3: Outline of Rank-consistent Ordinal Classification (forward direction)

In the following, we explain how to achieve rank consistency in ordinal classification. Let us assume $D = \{x_i, y_i\}_i^N$ is a data set, where x_i is an input image and y_i is the correct label of x_i. y_i satisfies $\{y_i \in 1, 2, ..., K\}$ and the labels have an order $1 < 2 < ... < K$. Then we prepare $K-1$ binary classifiers for the K-class rank-consistent ordinal classification (CORAL). A k-th classifier ($k \in \{1, ..., K-1\}$) outputs the probability of exceeding label (year) k. Fig. 3 shows the complete model of CORAL.

During the training, we map y_i to $K-1$ binary labels for assigning the correct label of each classifier. $y_i^{(k)}$ is then the correct label of k-th classifier and represents whether or not the creation date of x exceeds label k. If it does, then $y_i^{(k)}$ is 1, otherwise it is 0. The loss function, $L(W, b)$, is given in Eq. 1.

$$L(W,b) = -\sum_{i=1}^{N}\sum_{k=1}^{K-1} \log(P(y_i^{(k)} = 1))y_i^{(k)} + \log(1 - P(y_i^{(k)} = 1))(1 - y_i^{(k)}) \quad (1)$$

$P(y_i^{(k)} = 1)$ is the probability of y_i exceeding the label k. The final creation date is calculated using Eq. (2).

$$E_{x_i} = 1 + \sum_{k=1}^{K-1} f_k(x_i) \quad (2)$$

$$f_k(x_i) = \begin{cases} 1 & (P(y_i^{(k)} = 1) > 0.5) \\ 0 & (otherwise) \end{cases}$$

We propose two variants of the rank-consistent ordinal classification: *forward direction Rank-consistent Ordinal Classification (CORAL)* and *backward direction Rank-consistent Ordinal Classification (R-CORAL)*. The forward direction is the same as the one described above, where the k-th classifier outputs the probability of whether the estimated date of the input is greater than or equal to rank r_k.

The backward direction is the opposite of the forward direction, such that k-th classifier determines whether the input is less than the rank r_k. The loss function is the same as in Eq. 1, however the label for the correct answer is the opposite of the one described in the forward direction. $y_i^{(k)}$ represents whether or not x_i was created before rank r_k. If it does, then $y_i^{(k)}$ is 1, otherwise 0.

3.2 Object-centered Ensemble Approach

3.2.1 Objects in Age Detection Process. We now discuss the second approach that we propose in this work. When humans make inferences about the age of a target photo, they usually pay attention to a variety of signals including whether the image is black and white or in color, what is the overall quality of the photo and the level of its deterioration. They also look carefully at contents of the image such as persons' outlook, fashion and hairstyle, the shapes of portrayed objects such as cars, bicycles, furniture and so on. As an example, let's look at the photograph shown in Fig. 4 to try to reconstruct the likely inference process carried by humans. If we could know the make and the model of the car in this image, we would deduce that the photo was taken at least after a certain date (e.g., the date of the first production of this particular car type). We assume here that one can roughly guess an approximate time when the car was being produced even without having the exact

information about the car model and its history. However, there are also several newer car models shown in the background, which leads us to reason that the photo is actually more recent than it would seem based on the initial guess. This inference process could be also aided by considering the appearance of a person that stands behind the car. Both the cars and the person's outlook would be then used to make overall guess as for when the picture was taken.

Figure 4: An example image from the dataset (picture taken in 1989)

In order to replicate the part of the process of date inference by a human that relates to image objects, we propose an estimation approach summarized by the procedure shown below and sketched in Fig. 5.

(1) Detect and crop out the objects in target photo.
(2) Compute first the estimate based on the entire image using the full-size image model.
(3) Input each detected object into its corresponding object-centered image dating model and compute date estimate based on the object.
(4) Combine all the generated estimates to provide the final guess of the image creation date.

We detect and clip the objects in step 1 using an object segmentation module. The details of the segmentation method that we use are described in Sec. 4.2. In step 2, we use an age estimation model which was trained on original images, that is, images that are not cropped. In step 3, we use object-specific age estimation models. Their training data is different compared to one used for the model in step 2. These models are learned based on images of specific types of objects. Finally, step 4 combines the estimates from the model trained on entire images and the ones delivered by the models dedicated to each type of object that was found in the image. The ensemble methods are described in Sec. 3.2.2.

3.2.2 Ensemble methods. Several possible approaches can be devised for implementing effective ensemble. In our case, an obvious one would be to use the most recent estimate as the final judgment. Since older objects may exist until the present (e.g., old buildings or classic cars can survive unchanged till now), the object that was deemed as the most recent could serve as the basis for the final estimate of the creation year of the entire image. For example, suppose there is a photo that shows a person dressed in clothes typical for a fashion style of the 1980s, a car that was popular in the 1970s, as

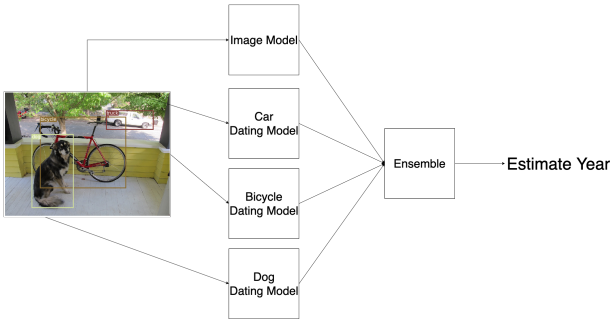

Figure 5: Toy example of the procedure for computing the image creation date based on the object-centered ensemble approach

well as a desktop computer that was produced in the 2000s. The final answer for the creation date of this photograph would then be the 2000s (or later) due to the presence of the desktop computer. E in Eq. 3 is the final date estimate for an input image where E_i is the estimated date by a particular model and N is the number of estimates.

$$E = \max_{1 \le i \le N} E_i \tag{3}$$

The next combination method is to integrate the estimates as a weighted average using the inverse of the Mean Absolute Errors (MAE) of each used model as weights. Since the amount of data used by a model and the accuracy of the model differ for different object types, we assign greater weights to the models that have higher reliability. Ensemble calculation is given by Eq. 4, where MAE_i represents the MAE of each model.

$$E = \frac{\sum_i \frac{1}{MAE_i} * E_i}{\sum_i \frac{1}{MAE_i}} \tag{4}$$

4 EXPERIMENTAL SETTINGS

We next describe the settings chosen for the experimental evaluation of the proposed methods.

4.1 Dataset

We use the "Date Estimation in the Wild" dataset" [10] which includes urls to images publicly available on Flickr[4] that were taken from 1930 to 1999. The total number of the images in the original dataset was 1,029,710. However, some of the photos are no longer available in Flickr and cannot be retrieved. We managed to collect 939,122 of the original images.

The authors of the original dataset [10] used 1,120 images as the test data, 8,494 images as the validation data and the rest of them as the training data. In our case, we use 1,040 images as the test data, 7,905 images as the validation data and the rest as the training data. Fig. 6, 7, 8 give the distributions of the training, validation and test portions of our dataset.

[4] http://www.flickr.com

The following pre-processing procedure was executed. All images were scaled by the ratio 256/min(w, h) (w and h are image dimensions) and then randomly cropped to 224×224×3 pixels for the model training. During model evaluation, the 256/min(w, h) scaled RGB images were center-cropped to a model input size of 224×224×3.

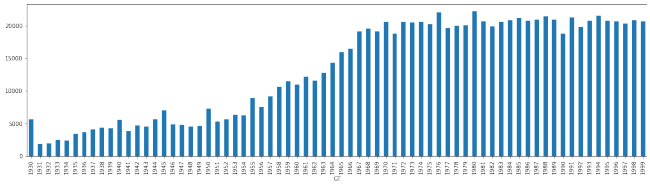

Figure 6: Image distribution over time in the training dataset

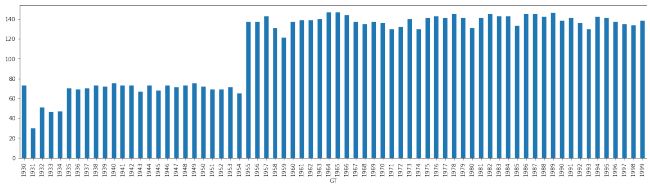

Figure 7: Image distribution over time in the validation dataset

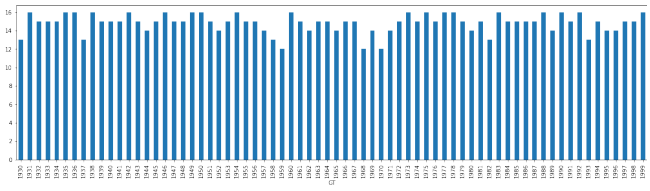

Figure 8: Image distribution over time in the test dataset

4.2 Object Detection

We have performed object segmentation on all the images using Detectron2 [18], which has been released by the Facebook research team. The trained model is an instance segmentation model of ResNet50+FPN provided by Detectron2 to obtain a mask image and bounding box for each image. The type of each segmented object, the bounding box of the object, and the assessed probability of that object were obtained to generate the data, as exemplified in Tab. 1.

The total number of object types which can be recognized by the trained model is 80. Tab. 2 shows the top 10 most common objects in our dataset. As one might expect, the majority of the found objects are human beings. Since ties and handbags are typically worn by persons, their recognition rate is also high. Besides humans, vehicles such as cars, trains and boats are other most commonly recognized object classes.

In the current experiment, we use a person model and a car model for ensemble approaches since persons and cars are the most common objects detected by the segmentation algorithm.

The input data to the person model is selected by the following process: (1) We made sure that the confidence score of object identification is over 0.99. (2) The height of the person should be also at least twice as large as its width. 344,146 images were collected for the person model following this process.

The input data to the car model is selected by the following process: (1) The confidence score of object identification must be higher than 0.95. (2) The width of the car should be larger than its height. In total, 42,534 images were collected for the car model following the above procedure. Fig. 9 ~12 show example images from the person and car datasets.

The same procedure of re-scaling, cropping and centering as discussed before was applied to the newly created images of cars and persons.

Figure 9: An example person from an image taken in 1988

Figure 10: An example person from an image taken in 1933

Figure 11: An example car from an image taken in 1964

Figure 12: An example car from an image taken in 1987

4.3 Tested Models

We trained CORAL, R-CORAL, as well as classification and regression models for the comparison. In each CNN model we prepared 69 output nodes for the CORAL and R-CORAL model as well as 70 output nodes for the classification model because the dataset have 70 classes (years from 1930 to 1999). 1 output node was prepared for the regression model.

For all the above-mentioned tested models, we used ResNet50[5] as the CNN model. Each model was trained using PyTorch[6] on a pre-trained ImageNet[7] model. The weights of the fully connected (FC) layers were randomly re-initialized. We randomly selected 128 images per batch and iterations were set to 1M. As optimization algorithms, Adam[10] with 0.0005 learning rate was applied to

[5]https://github.com/keras-team/keras-applications/blob/master/keras_applications/resnet50.py
[6]https://pytorch.org/
[7]http://www.image-net.org/

filepath	label	GT	class	score	left	right	top	bottom
8/29/8296357577.jpg	48	1978	person	0.99217	92	349	165	293

Table 1: An example of a data record obtained from a photograph from our dataset (picture taken in 1978). The label is the offset number of a year; GT denotes the date at which the photograph was taken; the class means the detected type of object; the score is the confidence value of segmentation and the last four columns denote the bounding box of the detected object.

Object	Object count
Person	2,994,072
Car	389,363
Train	207,438
Tie	149,634
Boat	125,111
Bus	89,146
Truck	88,383
Book	59,204
Handbag	59,017

Table 2: Top 10 most common objects in the dataset and their frequency

CORAL and R-CORAL, and SGD with 0.001 learning rate was applied to the classification and regression. Eq. 1 was used in CORAL and R-CORAL as loss functions, while cross entropy was used for classification, and euclidean loss function was applied for regression.

In addition to the classification and regression approaches we use the method of Müller *et al.* [10] as another baseline. This method uses a fine-tuned version of GoogleNet.

4.3.1 Object-centered Approach. Both the car model and the person model use R-CORAL. Adam with 0.0005 learning rate was applied and the number of iterations were 1M. The loss function that we used is given in Eq. 1. All ensemble combinations of classification, CORAL, R-CORAL, person and car models were compared in the experiments. Same as before, we used ResNet50 as the CNN model for all the approaches.

In the evaluation process, if an image had no object then only the full-size image models were applied.

4.4 Evaluation Metrics

For model evaluation and comparison, we computed the mean absolute error (MAE) and root mean square error (RMSE) on the test set by the model which had the best validation score:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - E_{x_i}| \quad (5)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - E_{x_i})^2} \quad (6)$$

5 EXPERIMENTAL RESULTS

5.1 Results of CORAL and R-CORAL

Fig. 13 shows the results of each method in terms of MAE and RMSE. The results indicate that the classification approach has the lowest MAE. On the other hand, in terms of RMSE, the proposed model R-CORAL outperforms other models. This suggests that rank consistency allows obtaining more stable results. The approach based on GoogleNet (Müller *et al.*) [10] could not produce satisfactory results. Interestingly, the regression approach was the worst performing method.

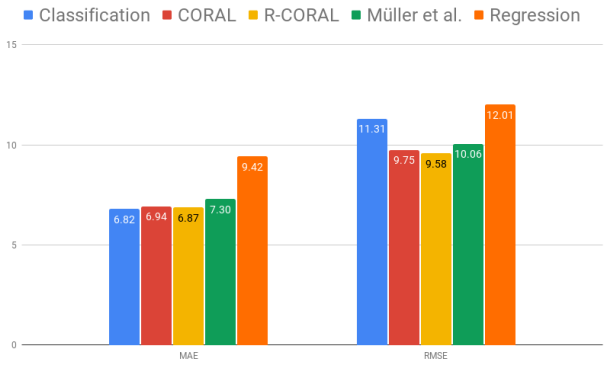

Figure 13: MAE & RMSE of CORAL and R-CORAL

We can observe that R-CORAL produces consistently better results than CORAL in both error measures. This indicates that reversing the order of ranks helps to improve the results to a certain degree. It is likely due to the distribution of the images over time in the dataset such that more images are from the recent years than from older years. Tab. 3 shows that CORAL is worse than R-CORAL for old images and R-CORAL is worse than CORAL for new images. Fig. 14 shows also the differences in MAE values between CORAL and R-CORAL in each year.

	CORAL	R-CORAL
1930~1964	7.98	**7.32**
1965~1999	**5.89**	6.41

Table 3: MAE for the first half and the second half of the timeline of our dataset

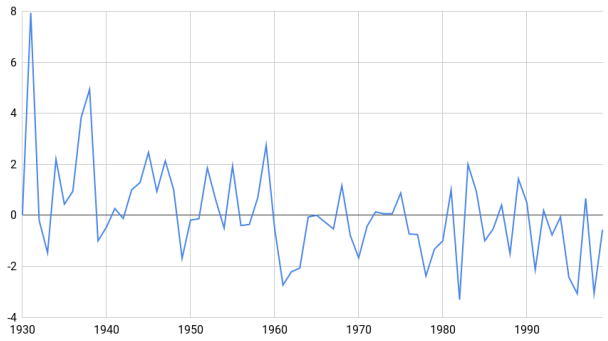

Figure 14: The difference between MAE of CORAL and MAE of R-CORAL for different years

5.2 Results of Object-centered Ensemble Approach

The newest (Eq. 3) and the weighted ensemble (Eq. 4) approaches as discussed in Sec. 3.2.2 are compared now for all combinations of the models. The results are shown in Tab. 4 ~7 (cls denotes classification). We show ensemble combinations of classification, CORAL and R-CORAL in rows, while the columns indicate different used objects. As shown in Tab. 4 and 5, the case of using all the models and all the objects in the weighted ensemble produces the best MAE and RMSE scores. In the weighted ensemble, person objects are more effective than car objects, although we also notice that car objects are more effective than person objects in the newest ensemble. This might be somehow related to the observation that cars are more common in the recent years when compared to the old years, while the opposite is true for persons. Fig. 15 and 16 show the average numbers of objects (persons, cars, respectively) per image at different years based on the entire dataset, i.e., including training, validation and test parts.

Full-size Image Model	None	Person	Car	Person and Car
cls	6.82	6.71	6.77	6.68
CORAL	6.94	6.88	6.90	6.85
R-CORAL	6.87	6.71	6.80	6.67
cls & CORAL	6.37	6.36	6.34	6.34
cls & R-CORAL	6.30	6.23	6.25	6.20
CORAL & R-CORAL	6.52	6.43	6.48	6.41
cls & CORAL & R-CORAL	6.19	6.16	6.16	**6.14**

Table 4: MAE of the weighted ensemble

5.3 Case Study

As a case study we discuss an image taken in 1961 (Fig. 17). R-CORAL trained with full-size images estimates that it was taken in 1945. Person model outputs 1990, 1959, 1965, 1966 and 1965 as estimated dates for person objects found in this image (see Fig. 18). MAE of the full-size image model computed over the validation dataset is 8.37, while the MAE of the person model over the validation dataset is 8.31. Eq. 7 gives the calculation of the date by the weighted ensemble for this image, the result of which is quite

Full-size Image Model	None	Person	Car	Person and Car
cls	11.31	10.73	11.19	10.64
CORAL	9.75	9.66	9.72	9.64
R-CORAL	9.58	9.38	9.48	9.33
cls & CORAL	9.53	9.39	9.49	9.35
cls & R-CORAL	9.31	9.13	9.26	9.10
CORAL & R-CORAL	9.08	8.99	9.04	8.96
cls & CORAL & R-CORAL	8.98	8.89	8.95	**8.88**

Table 5: RMSE of the weighted ensemble

Full-size Image Model	None	Person	Car	Person and Car
cls	6.82	7.85	6.80	7.28
CORAL	6.94	8.00	6.94	7.46
R-CORAL	6.87	7.78	6.87	7.25
cls & CORAL	7.36	8.44	7.41	7.95
cls & R-CORAL	7.08	8.16	7.11	7.66
CORAL & R-CORAL	7.05	8.13	7.08	7.62
cls & CORAL & R-CORAL	7.67	8.72	7.71	8.23

Table 6: MAE of the newest ensemble

Full-size Image Model	None	Person	Car	Person and Car
cls	11.31	12.40	11.23	11.59
CORAL	9.75	11.33	9.78	10.47
R-CORAL	9.58	10.98	9.59	10.11
cls & CORAL	11.37	12.67	11.41	11.97
cls & R-CORAL	11.03	12.36	11.07	11.64
CORAL & R-CORAL	9.98	11.56	10.03	10.74
cls & CORAL & R-CORAL	11.60	12.88	11.64	12.20

Table 7: RMSE of the newest ensemble

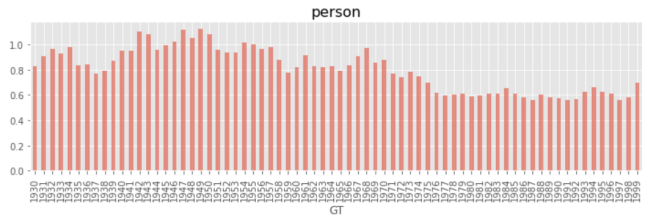

Figure 15: The average number of person objects per image

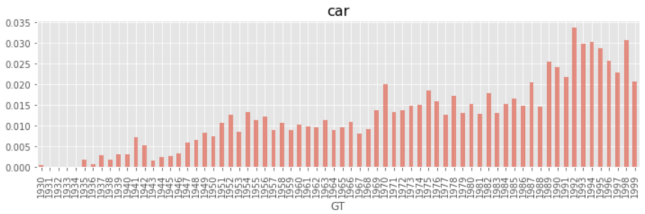

Figure 16: The average number of car objects per image

close to the ground truth date (1961). On the other hand, the newest ensemble outputs 1990 as the estimated date for this image, which is very far from 1961. 1990 is selected as the answer for this ensemble due to the person image (the first one on the left in Fig. 18) which has a high resolution because the man is standing near the camera. Naturally, our dataset contains more images that have objects shown in high resolution in the recent years when compared to older years (due to the progress of camera technology). This likely causes the problem when different objects in the same image are positioned at different distances from the camera point.

$$\frac{\frac{1}{8.37} * 1945 + \frac{1}{8.31} * 1969}{\frac{1}{8.37} + \frac{1}{8.31}} = 1957 \quad (7)$$

Figure 17: An example image taken in 1961

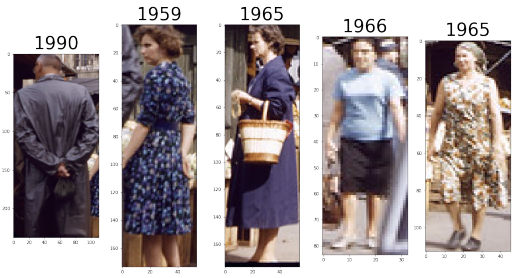

Figure 18: Persons with their estimated years extracted from the image shown in Fig. 17

6 CONCLUSIONS

In this paper, we introduce two methods for the task of estimating dates at which photos were taken: the Rank-consistent Ordinal Classification with two kinds of settings (CORAL and R-CORAL), and ensembles of object-centered date estimates. In the experiments we demonstrate that in terms of RMSE the Rank-consistent Ordinal Classification achieves the best results compared to baselines with a quite good improvement, while its MAE results are still comparable to the ones of the best performing method. The small RMSE means that the Rank-consistent Ordinal Classification is relatively more stable than the other approaches, and has fewer outliers. Another advantage of the Rank-consistent Ordinal Classification method is that it consumes less memory than the general classification task, since it has fewer parameters in the final layer. The comparison between the forward and reverse directions shows that the reverse direction is more effective. This is likely due to the distribution of the images in dataset such that more images are from the recent years than from older years. Finally, the object-centered approach that relies on identifying objects in images and using their specific models was found to help in further improving the results.

In the future we plan to train additional models which focus on other object types besides persons and cars. We plan also to introduce a more effective ensemble method.

REFERENCES

[1] Wenzhi Cao, Vahid Mirjalili, and Sebastian Raschka. 2019. Rank-consistent ordinal regression for neural networks. *arXiv preprint arXiv:1901.07884* (2019).
[2] Gaël Dias, José G Moreno, Adam Jatowt, and Ricardo Campos. 2012. Temporal web image retrieval. In *International Symposium on String Processing and Information Retrieval*. Springer, 199–204.
[3] Orla M Doyle, Eric Westman, Andre F Marquand, Patrizia Mecocci, Bruno Vellas, Magda Tsolaki, Iwona Kłoszewska, Hilkka Soininen, Simon Lovestone, Steve CR Williams, et al. 2014. Predicting progression of Alzheimer's disease using ordinal regression. *PloS one* 9, 8 (2014), e105542.
[4] B. Fernando, D. Muselet, R. Khan, and T. Tuytelaars. 2014. Color features for dating historical color images. In *2014 IEEE International Conference on Image Processing (ICIP)*. 2589–2593.
[5] Shiry Ginosar, Kate Rakelly, Sarah Sachs, Brian Yin, and Alexei A Efros. 2015. A century of portraits: A visual historical record of american high school yearbooks. In *Proceedings of the IEEE Int. Conference on Computer Vision Workshops*. 1–7.
[6] Adam Jatowt and Ricardo Campos. 2017. Interactive System for Reasoning about Document Age. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM '17)*. 2471–2474.
[7] Adam Jatowt, Yukiko Kawai, and Katsumi Tanaka. 2007. Detecting age of page content. In *Proceedings of the 9th annual ACM international workshop on Web information and data management*. 137–144.
[8] Haibin Liao, Yuchen Yan, Wenhua Dai, and Ping Fan. 2018. Age estimation of face images based on CNN and divide-and-rule strategy. *Mathematical Problems in Engineering* 2018 (2018).
[9] Paul Martin, Antoine Doucet, and Frédéric Jurie. 2014. Dating color images with ordinal classification. In *Proceedings of International Conference on Multimedia Retrieval*. 447–450.
[10] Eric Müller, Matthias Springstein, and Ralph Ewerth. 2017. "When Was This Picture Taken?" – Image Date Estimation in the Wild. In *Advances in Information Retrieval*, Joemon M Jose, Claudia Hauff, Ismail Sengor Altıngovde, Dawei Song, Dyaa Albakour, Stuart Watt, and John Tait (Eds.). Springer International Publishing, Cham, 619–625.
[11] Choon-Ching Ng, Moi Hoon Yap, Yi-Tseng Cheng, and Gee-Sern Hsu. 2018. Hybrid ageing patterns for face age estimation. *Image and Vision Computing* 69 (2018), 92–102.
[12] Zhenxing Niu, Mo Zhou, Le Wang, Xinbo Gao, and Gang Hua. 2016. Ordinal regression with multiple output cnn for age estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4920–4928.
[13] Frank Palermo, James Hays, and Alexei A Efros. 2012. Dating historical color images. In *European Conference on Computer Vision*. Springer, 499–512.
[14] Octavian Popescu and Carlo Strapparava. 2015. Semeval 2015, task 7: Diachronic text evaluation. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*. 870–878.
[15] Ruth Rettie, Ursula Grandcolas, and Bethan Deakins. 2005. Text message advertising: Response rates and branding effects. *Journal of targeting, measurement and analysis for marketing* 13, 4 (2005), 304–312.
[16] Tawfiq Salem, Scott Workman, Menghua Zhai, and Nathan Jacobs. 2016. Analyzing human appearance as a cue for dating images. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 1–8.
[17] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1–9.
[18] Yuxin Wu, Alexander Kirillov, Francisco Massa andWan Yen Lo, and Ross Girshick. 2019. Detectron2.