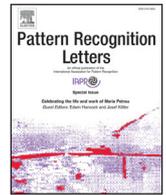




ELSEVIER

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Two sides of the same coin: Improved ancient coin classification using Graph Transduction Games

Sinem Aslan^{a,b,c,*}, Sebastiano Vascon^{a,b}, Marcello Pelillo^{a,b}^a ECLT, Ca' Foscari University of Venice, Ca' Bottacin Dorsoduro 3911, Calle Crosera 30123 Venice, Italy^b DAIS, Ca' Foscari University of Venice, via Torino 155, 30172 Mestre (VE), Italy^c International Computer Institute, Ege University, 35100, Bornova, Izmir, Turkey

ARTICLE INFO

Article history:

Received 23 July 2019

Revised 22 November 2019

Accepted 6 December 2019

Available online 6 December 2019

MSC:

41A05

41A10

65D05

65D17

ABSTRACT

In this work we tackle the problem of automatic recognition of ancient coin types using a semisupervised learning method, namely Graph Transduction Games. Such problem is complex, mainly due to the low inter-class and large intra-class variations and the task becomes even more complex due to lack of labeled large datasets from certain ancient ages. In this paper we propose a new dataset which is chiefly the extension of a previous one both in terms of quantity and diversity. Moreover, we propose a game-theoretic model that exploits both sides of a coin to achieve higher classification accuracy. We experimentally demonstrate that proposed approach brings performance improvement in this complex task even when few number of labelled images are available.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Similar to sculptures or paintings from the ancient Rome, Roman coins are important source of information for historians and archaeologists, i.e. they delineate natural disasters like eruption of volcanoes, observed environmental events like appearance of comets, and political circumstances like homicide of Caesar, etc. Conventionally, the type of a found ancient coin is detected by searching manually through index books, e.g. [8], which is a time-consuming and demanding task.

In this paper we put our effort to automatize such time consuming and exhaustive task by using a vision-based classification framework. We involve classification of Roman Republican coins (146 BC - 31 BC) based on the typology used in Michael Crawford's 1974 publication [8] where the coins are classified by the authority that was responsible for production of them. From the computer vision point of view, ancient coin classification is a highly challenging task due to high intra-class and low inter-class variations. Another difficulty arises from existence of high number of coin types (classes), ex. Roman Republic compose over 550 [8] different classes, while the number of coin specimens found for each class might be limited. This last problem creates the need for techniques that are efficient even when training data is not abundant. We

decided to face such a complex task by using a semi-supervised learning (SSL) algorithm. The motivation that drive our choice toward a SSL algorithm is that SSL is a class of machine learning algorithms that works better under a scarcity of labeled data condition. SSL algorithms estimate better classification boundaries exploiting the feature space of both labeled and unlabeled data points [5]. This property translates to require an expert to annotate few coins, saving time and money. More specifically, we model coin classification problem using a non-cooperative multiplayer game, namely Graph Transduction Games (GTG) [10], by using integrated pairwise similarities computed for each side of coins. We chose to use GTG for a series of reasons: *i*) GTG considers the global structure of the data when performing the labeling inference. Thus, the final labeling decisions are taken by considering all the choices made by all data points, looking for a consistent assignment, *ii*) prior knowledge on the label assignments (for example the softmax output of a neural network) can be plugged in straightforwardly finding the best solution close to the priors, *iii*) the superiority of GTG over a number of SSL methods has been proved recently in [25]. An important extension accomplished in this work over our previous conference paper [4] is we created a new image dataset, which we call as RRC-60 (Roman Republican Coins - 60), of ancient coins from the era of Roman Republic to be used for classification purposes. In the construction of RRC-60 we referred to the coin types of the dataset published by Zambanini and Kampel [28], which is composed of 180 images of reverse sides of 60 coin types, that we also experimented in our

* Corresponding author.

E-mail address: sinem.aslan@unive.it (S. Aslan).

Table 1
Ancient coin datasets used by computer vision community.

Reference	Period	Source	#Images	#Classes	Availability	Task	Side	Coin grade	Imaging condition & Image quality
RIC-Hq [23]	Imperial Rome (29BC-476AD)	www.acsearch.info	29,807	83 emperors	upon request	Classification	Observe	Varied	Controlled env. & High quality
RPC-Scan [23]	(Provincial) Imperial Rome (29BC-476AD)	Fitzwilliam Museum	19,164	N/A	upon request	Classification	Observe	N/A	Medium (Scans of photographs of decades ago)
RIC-Cond [23]	Imperial Rome (29BC-476AD)	Collections sold by ancient coin dealers	600	100 emperors & 3 grading classes	upon request	Classification & Grading	Observe	Varied	N/A
RIC-ACS [6]	Imperial Rome (29BC-476AD)	www.acsearch.info	200,000	N/A	-	Localization & Segmentation	N/A	Varied	Controlled env. & High quality
RIC-eBay [6]	Imperial Rome (N/A)	eBay	100	N/A	-	Localization & Segmentation	Both	N/A	Cluttered background
[16]	Imperial Rome (31BC-491AD)	Numismatic web sites	4526	314 RIC labels 96 emperors	-	Classification	Both	N/A	Controlled env. & High quality
[2]	Republican Rome (509BC-27BC)	(1) Vienna Museum of Fine Arts, (2) British Museum, (3) www.acsearch.info	2224	29	-	Classification	Reverse	N/A	Illumination and orientation variations
[21]	Sasanian Empire (531AD-628AD)	N/A	570	3	-	Classification	Both	N/A	Controlled env. & High quality
[28]	Republican Rome (146BC-31BC)	Vienna Museum of Fine Arts	180	60	Public	Classification	Reverse	Varied	Controlled env. & High quality
[13]	Time of Alexander the Great (336BC-323BC)	Fitzwilliam Museum	2400	10	-	Identification	Both	N/A	Varying cond. (acquired by different devices, illum. and orient. variations)
[15]	Asia Minor (late 7th century BC)	Fitzwilliam Museum	350	3	-	Identification	Observe	N/A	Illumination variations
[30]	Imperial Rome (30BC-300AD)	N/A	3,000	106	-	Classification	Both	Varied	Illum. and orient. variations
[24]	Merovingen dynasty (early middle ages)	N/A	4659	4	-	Classification	Both	N/A	Illumination variations
RRC-60 (proposed)	Republican Rome (146BC-31BC)	(1) www.acsearch.info , (2) CRRO	6000 image pairs	60	Public	Classification	Both	Varied	Illumination variations & High quality

conference paper [4]. More specifically, we extended the dataset of [28] in terms of both quantity and diversity, i.e. we collected 100 image pairs (observe and reverse sides) for each coin type of [28]. Actually, we aimed to abide by the same 60 coin types with [28], however since we could not reach at least 100 number of images for five classes of [28] we replaced them with other five Roman Republican coin types having similar appearance. To the best of our knowledge RRC-60 is the largest publicly published coin image dataset from the Republican period to be used by computer vision community. Moreover, while our experimental results show that availability of images from both sides of coins is important to obtain higher classification accuracy, there is not such a dataset publicly available (See Table 1) and RRC-60 comes as a remedy this issue. We will publish the dataset, trained models, features, and codes of the experiments at <https://github.com/siinem/RRC-60>.

From the methodological point of view, while at our previous work [4] we employed GTG for classification of only the reverse sides of coins, in this work we extend our pipeline with an additional fusion step which brings to the method the information

coming from two sides of a coin. Consequently, GTG exploits the fused information to perform the classification, providing us a significant performance improvement. We present this scheme as a baseline on this new dataset to the community for further method developments.

2. Related works

2.1. Visual classification of ancient coins

Albeit a number of works addressed the problem of coin identification [13] and coin grading [20], one of the most important and well-studied vision tasks in the field of ancient coins has been coin classification. Earlier works [15,28,31,32] aimed to bypass the learning phase due to unavailability of abundant training data and followed a classifier-free procedure based on NN-search where the class of a query coin image is set to the class of its most similar coin image in the training set. These works measured similarity between coin images by adopting the notion of image

matching, e.g. [15] used number of matched (sparsely-detected) SIFT features between images to measure the similarity between them, in [28,31] matching costs (that are produced by the SIFT-Flow algorithm) between (densely-detected) SIFT features of coin images are used to define dissimilarity between them, a similarity score is derived in [32], that indicate matching quality between densely computed LIDRIC features between coin images.

Some other works adopted supervised learning for ancient coin classification, e.g. employing the Bag of Visual Words paradigm, performance of SVM and GMM are reported for classification of Roman coins in [1] and [3], respectively. More recent works working on larger datasets employed CNNs, e.g. AlexNet architecture that was pretrained on ImageNet is finetuned on around 4500 images of Roman Imperial Coins [16] and performance improvement on test set over SVM is reported. [23] proposed a CNN architecture with five consecutive convolutional blocks and max pooling pairs and trained it on the RIC-Hq dataset composed of Roman Imperial coins including around 29,000 images of 83 classes. Using the trained model [23] explored the classification performance on other two coin datasets, i.e. RPC-Scan and RIC-Cond, created in different conditions. Recently, a similar CNN architecture to the one in [23] is used in [7] to identify semantic elements (which are inferred from a highly unstructured textual description of the coins) on the reverse side of Roman Imperial coins.

While performance results reported by supervised approaches are outstanding, their performance highly depends on availability of large-scale annotated datasets. Recently in [4], we proposed to apply a semi-supervised learning scheme, namely Graph Transduction Games Framework, for classification on a very small-scale ancient coin dataset including 180 images from reverse sides of coins, which outperformed the literature work adopting NN-classifier and experimenting on the same dataset.

2.2. Ancient coin datasets

In the literature, almost each research group created their own coin dataset to evaluate performance of their method due to unavailability of a publicly released large-scale benchmark dataset for ancient coin classification. We summarize the most distinguished datasets that were employed in these works at Table 1. For each dataset, we present the covered timespan, the source where the images were acquired from, number of images and classes, the task that the dataset was used for, available coin side (*Obverse*, *Reversed* or *Both*), the grade of the coins, the environmental conditions in acquisition (*controlled* or *in the wild*) and the imaging quality. We prepared Table 1 based on the information provided in the corresponding papers that employed the datasets and when we could not reach to a particular information in that paper or in our search on the web we marked the corresponding cell by *N/A* (*Not Available*).

In Table 1, we see that the majority of the literature works experimented by using images of coins from ancient Rome, i.e. from Imperial [6,16,23,30] or Republican [2,28] periods, while few earlier works employed coins from Sassanian Empire [21], Asia Minor [15] and Merovingen dynasty [24]. The largest classification dataset with 200,000 images, i.e. RIC-ACS [6], consists of Roman Imperial coins and to the best of our knowledge it is not published publicly so far. The second and third largest datasets, i.e. RIC-Hq [23] and RIC-Scan [23] respectively, consist of the observe sides of the Roman Imperial coins, and can be obtained upon request to the authors. In our RRC-60 dataset, we publish publicly 6000 image pairs, i.e. 12,000 images, for 60 coin types from Republican Rome. Grade of involved coins is varied, i.e. in general the ones obtained from CRRO are in lower grade (see Section 3). Images are mostly in high resolution and acquired in controlled condi-

tion, i.e. with white background and in similar orientation, while illumination variations exist due to different sources obtained.

3. Roman Republican Coins (RRC-60)

In this work, we created a new dataset, that we call by Roman Republican Coins (RRC-60), for ancient coin classification which will be published publicly at the web site¹ with the features and fine-tuned DL models that we employed in this paper. In the creation of this dataset we referred to a previous public² dataset of [28] with 180 images of reverse (motif) sides of 60 coin types from Roman Republican period which we also experimented on at out conference paper [4]. Thus, abiding by the same coin types with [28] we created a new dataset with larger quantity and diversity, i.e. we collected 100 images for each side of the 60 coin types, which in the end resulted with 6000 image pairs for each coin specimen. We collected our dataset from following two sources:

1. *acsearch.info* is an online auction site for coins, banknotes and antiques. It is one of the largest sources for ancient coin images and a number previous literature works, i.e. [2,6,23], have already exploited to acsearch to create their own datasets in large numbers. Around 2 million ancient coins (Greek, Roman Republican, Roman Imperial, Roman Provincial, Byzantine, Celtic) are presented to the interest of collectors. Coin images are mostly in high resolution, with white-coloured background, sometimes with some stamps at the background, and at a standard orientation. We collected around 90% percent (5483 of 6000) of images of RRC-60 from acsearch.info.
2. *Coinage of the Roman Republic Online (CRRO)* is another largest source for coin images from Roman Republican period which is created to be an online version of Crawford's reference book [8]. The collection includes 49,162 coins which are collected from 29 contributors, e.g. Bibliotheque Nationale de France (20,237 coins), British Museum (12,537 coins), American Numismatic Society (6188 coins), etc. (full list is available³). Although information on grades of coins are not presented in CRRO, we observe that coins in CRRO seem as in lower grades, i.e. has more deterioration on their surfaces, compared to coins in acsearch.info. Moreover, since the images are from a variety sources they have illumination variations, not in high resolution as always as it was in acsearch.info, not always but mostly with white-coloured background, and at a standard orientation (the majority of the images that are presented in Fig. 2 are acquired from CRRO). When we could not retrieve sufficient quantity of images for a class at acsearch.info, we complemented it by referring to CRRO. Eventually, 517 images for 33 classes are obtained from CRRO which constitutes to around 10% of RRC-60.

In spite of we aimed to abide by the coin types of [28], we could not reach sufficient quantity of images for a few classes of [28], e.g. we could only reach to 40 images for 258/1 (class 4), 30 images for 370/1a-b (class 5), 11 images for 169/1 (class 15), 46 images for 451/1 (class 44), and 41 images for 462/1a-c (class 55). Thus, by discarding those types we selected another five coin types from Republican period which look quite similar to the previous ones. We present an example image from previous and recent classes in Fig. 1 with their Crawford numbers and issuer name.

Ancient coin classification is a highly challenging task from the computer vision point of view. Missing parts and degradations on the aged specimens hinder to extract discriminative visual information for classification and such spatial variations arising

¹ <https://github.com/siinem/RRC-60>.

² <https://cvl.tuwien.ac.at/research/cvl-databases/coin-image-dataset/>.

³ <http://numismatics.org/crro/contributors>.



Fig. 1. Obverse-side and reverse-side images of a coin selected from five replaced classes of the dataset in [28]. First row: Classes of the dataset in [28], Second row: New classes in RRC-60 dataset. First column: 258/1 (Caesar) and 257/1 (Vargunteius), Second column: 370/1a-b (Servilius) and 264/1 (Servilius), Third column: 169/1 (Anonymous) and 219/1a-e (Antestius), Fourth column: 451/1 (Pansa) and 449/1a (Pansa), Fifth column: 462/1a-c (Cato) and 343/1c (Cato).



Fig. 2. Intra-class variations in RRC-60 due to illumination differences, degradations and dirt on the surfaces of the coins. Each row depicts images of observe-side and reverse-side of three coins selected from the same class. First row: Class 60 (Cra 489/5-6); Second row: Class 59 (Cra 489/2-3); Third row: Class 53 (Cra 543/1).



Fig. 3. Intra-class variations in RRC-60 due to manual manufacturing of coins by different engravers. Left: Observe sides of three coins from class 25 (Cra 379/1), Right: Differences in spatial relations on the reverse side of each in respective order.

from degradations on the coins yield to large intra-class variation. Additionally, dirt on the coins and imaging specimens under different lighting conditions cause high variations in images of the same class. We present Fig. 2 to demonstrate such variations at observe and reverse side images of three coins selected from three different classes of RRC-60. In ancient times dies to strike the coins were manufactured manually by different engravers. Therefore, differences in spatial relations on the images from the same class exist which is another factor that causes high intra-class variations. In Fig. 3, we demonstrate such variations on observe and reverse side images of three coins selected from the same class of RRC-60 images.

In addition to large intra-class variation, another challenge met in the classification on RRC-60 dataset is high visual similarity between a number of classes. As an example, we present visual similarities between images of five classes of RRC-60 in Fig. 4.

4. Graph Transduction Games (GTG)

The Graph Transduction Game (GTG) [10], is a graph-based semi-supervised learning algorithm which has recently found a renewed interest in different contexts, e.g. bioinformatics [26], the domain adaptation [25] and the label augmentation [9] problems.



Fig. 4. Examples to low inter-class variation in RRC-60. Left to Right: Observe and Reverse side images of a coin selected from Class 1 (Cra 387/1), Class 2 (Cra 300/1), Class 13 (Cra 352/1a-c), Class 16 (Cra 275/1), and Class 17 (273/1).

Even though the process above resemble a nearest neighbour classification, the main difference (and peculiarity) is that in GTG the label consistency is a global property which is not related to a single player but achieved for all of the players. The GTG method views the classification task as a non-cooperative multiplayer game, in which the objects (or images of a dataset) are the players while the possible strategies are the class labels. Each players pick a certain strategy and receive a reward proportional to the choices made by the opponents. Indeed, is on each player's interests to pick the strategy (label) that maximize his/her payoff, hence the compatibility between object and label. This non-cooperative game is played until all the players (objects) have chosen a strategy (label) and none of them would like to change it because otherwise they will receive a lower gain. This particular condition is known as *Nash Equilibria* [19]. Once the game reaches an equilibrium, every player plays its best strategy which correspond to a consistent labeling [14]. We recap here some basic concepts on game-theory. Given a set of players $I = \{1, \dots, n\}$ (i.e. coin images in our dataset) and a set of possible pure strategies $S = \{1, \dots, m\}$ (the set of labels):

1. *mixed strategy*: a mixed strategy x_i is a probability distribution over the possible strategies for player i . Then, $x_i \in \Delta^m$, where $\Delta^m = \{\sum_{h=1}^m x_i(h) = 1, x_i(h) \geq 0, h = \{1, \dots, m\}\}$ is the standard m -dimensional simplex and $x_i(h)$ is the probability of player i to choose the pure strategy (label) h .
2. *strategy space*: it corresponds to the union of all mixed strategies of the players $x = \{x_1, \dots, x_n\}$. The strategy space is a stochastic matrix of size $n \times m$ and represents the starting point of the game.
3. *utility function*: it is responsible to provide the reward to the i -th player when it chooses a mixed strategy x_i . In particular $u : x_i \rightarrow \mathbb{R}_{\geq 0}$.

In the GTG method, the players are separated into *labeled* ($\mathcal{L} \subset I$) and *unlabeled* ($\mathcal{U} \subset I$) sets. The strategy space x is initialized in two different ways based on the fact that an object is *labeled* or *unlabeled*. A one-hot vector is assigned to each labeled objects, since their labels are known:

$$x_i(h) = \begin{cases} 1, & \text{if } i \text{ has label } h \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

while for the unlabeled players, we can assign the same probability for all the labels:

$$x_i(h) = \frac{1}{m} \quad \forall h \in S \quad (2)$$

indeed, if prior knowledge is available, e.g. a bias toward certain classes for an object, they can be injected here.

Payoff function. The payoff function is responsible to quantify the reward for a player (object) when he/she choose a particular strategy (label). The payoff should also considers the similarities between labeled and unlabeled players. The rationale is that the more similar the players are, the more they will influence each other in picking one of the possible strategies (labels). Formally, given a player i and a strategy h the utility function is defined as follows:

$$u_i(h) = \sum_{j \in U} (A_{ij}x_j)_h + \sum_{\gamma=1}^m \sum_{j \in L_\gamma} A_{ij}(h, k) \quad (3)$$

$$u_i(x) = \sum_{j \in U} x_j^T A_{ij} x_j + \sum_{\gamma=1}^m \sum_{j \in L_\gamma} x_j^T (A_{ij})_\gamma \quad (4)$$

where L_γ is the set of labeled objects with class γ and $A_{ij} \in \mathbb{R}^{m \times m}$ is the *partial payoff matrix* between the pair of players (i, j) . In particular, $A_{ij} = I_m \times \omega_{ij}$ with I_m being the identity matrix and ω_{ij} the similarity of player i and j .

Players similarity. To measure the similarity between pair of players (coin images in our context) we use their feature representation. Let's consider two players i and j , the similarity (or affinity) between them can be computed by Eq. 5, where $d(f_i, f_j)$ denotes the distance between features f_i and f_j and σ_i is the distance between i and its 7-nearest-neighbors [33].

$$\omega(i, j) = \exp \left\{ -\frac{d(f_i, f_j)}{\sigma_i \sigma_j} \right\} \quad (5)$$

This formulation allows us to avoid setting the kernel parameter, which is a time consuming and data specific task.

Similarity Sparsification The sparsification of the affinity between the players is an important step in the performances of the algorithm, which has been also testified in [26]. The goal is to filter out the small similarities which may bias the utilities in Eq. 4 toward an incorrect labeling. Here, we follow the method proposed in [26] which is a statistical connectivity principle in random graph stating that a graph is connected if each node has at least $k = \lfloor \log_2(n) \rfloor + 1$ nearest neighbours. The rationale of this choice is the following: since GTG propagates the labels through the graph from the labeled elements to the unlabeled ones, having a disconnected graph might stuck the propagation at a certain point. The aforementioned sparsification scheme ensures that the graph is connected after the sparsification, hence all the nodes are reached at the equilibrium condition of the dynamical system (Eq. 6).

Finding Nash Equilibria In order to find a Nash Equilibria of the game, which corresponds to a consistent labeling of the objects (coin images), we used a result from Evolutionary Game Theory [29], named Replicator Dynamics (RD). The RD are a family of dynamical systems that mimic a Darwinian selection process over the set of strategies for each player. The underlying idea is that it leads the fittest strategies to survive, while the others to get extinct. The RD are more formally defined as follows:

$$\dot{x}_i(h)^t = x_i(h)^t \frac{u_i(h)^t - u_i(x^t)}{u_i(x^t)} \quad (6)$$

where $x_i(h)^t$ is the probability of strategy h at time t for player i (see Eq. 3) and $u_i(x^t)$ is the expected payoff of the entire mixed strategy (see Eq. 4). The Eq. 6 is iterated until convergence.⁴ Once Eq. 6 reaches the convergence, we simply get the index of the maximum value on each i -th row of x in order to label the i -th object.

⁴ Convergence criteria: i) the distance between two successive steps is $\|x^{t+1} - x^t\|_2 \leq \epsilon$ or ii) a certain amount of iterations is reached. See [22] for a detailed analysis

5. Ancient coin classification using GTG

In this section, we first review implementation of GTG using one side of coins, then we present how we used it for both sides of coins.

5.1. Using one side of coins

At the conference paper [4] of this work, we applied GTG using images from reverse side of coins. In order to accomplish ancient coin classification using GTG framework, we consider the training set images as labeled players and the test set images as the unlabelled players. Then, labels from the training images are propagated through the unlabelled test images by playing a non-cooperative multiplayer game where the players are the images and the labels are the possible strategies. Main steps of implementation in [4] and our recent extensions are listed as follows:

1. **Feature extraction.** In [4] we computed two type of features, i.e. *dense SIFT features* [18] and *off-the-shelf* features (which were obtained from the output of the last fully-connected layer of DenseNet201[12] that was pretrained on ImageNet). In this paper we used dense SIFT features in replication of [4] and [28] at the experiments on the *RRC-60-small* dataset. At the experiments on *RRC-60* dataset we used features acquired from last fully-connected layer of Resnet152 architecture [11] after the network is finetuned by a number of training sets of *RRC-60* with varying sizes. Moreover, we applied standardization on each feature dimension accounting for mean and standard deviation of the labeled set.
2. **Initialization of the strategy space.** Initialization of the strategy space provides assigning a starting point for the game. In [4] by assuming that no prior knowledge was available, we initialized strategy space using Eq. 1 and 2. In this work, we additionally used softmax output from the fine-tuned ResNet152 model as prior to set a starting point for the game (see “*w prior*” in the experiments).
3. **Computation of similarity between images.** In order to measure dissimilarity between image pairs, we used matching scores of SIFT flow [17] similar to [4,28] when we use dense SIFT features and Euclidean distance when we use CNN-based features. Next, we switch from dissimilarity space to affinity (similarity) space by using Eq. 5. Finally, we sparsify the obtained affinity matrix as mentioned in Section 4, i.e. in this work we perform sparsification in a more principled way (see Section 4) while in [4] we used $k = 2$ setting.
4. **Execution of the transduction game.** Initializing the strategy space and computing the affinity matrix, GTG starts to be played between the players until reaching an equilibrium state where final probabilities of the strategies are obtained for each unlabelled player. Then, each unlabelled player is labelled by its strategy with the highest probability.

5.2. Using both sides of coins

In order to exploit information from both sides of a coin in the GTG framework, one can refer to *similarity fusion* scheme where the similarity of two coins i, j is computed as follows:

$$\omega_F(i, j) = \alpha \omega_O(i, j) + (1 - \alpha) \omega_R(i, j) \quad (7)$$

Here, $\omega_O(i, j)$ and $\omega_R(i, j)$ are the similarity functions which consider only the *observe* and the *reverse* side, respectively (see Eq. 5). α balances the convex combination of the two functions, hence adjusts the affect of each coin's side into the final similarity $\omega_F(i, j)$. A similar type of fusion has been already successfully applied in [27].

From both a qualitative and quantitative point of view, we noted that the *observe*-side of the coins look more similar to

each other and so the *reverse*-side becomes more discriminative in assessing the label of a coin. To achieve some flexibility on contribution of each side we mainly adopted aforementioned similarity fusion scheme. However, it is also reasonable to apply feature concatenation scheme where L_2 -normalized features of images of each coin side are concatenated before the procedures of standardization and similarity computation using Eq. 5 (See Section 5.1). Thereby, we also experimented using feature concatenation scheme on RRC-60 dataset (See Section 6.2).

6. Experiments

In order to show improvements after conference paper of this work we first experimented on a smaller subset of our dataset, i.e. we call by *RRC-60-small*, which is created to be in the same size with the coin dataset published by [28] that we experimented on in [4]. Next, we experimented on our proposed dataset RRC-60. In the implementation of GTG, we always used $k = \lfloor \log_2(n) \rfloor + 1$ as claimed in Section 4.

6.1. RRC-60-small

In [4] we performed experiments on a small dataset made of 180 images of *reverse* side of 60 coin types. In a similar manner, we randomly selected three images of reverse and corresponding observe sides for each of the 60 coin types to create RRC-60-small. Main reason of creating RRC-60-small is the lack of observe side images of the coins in [28] which we need to experiment on to make comparative analysis. Furthermore, we have done this random selection five times, i.e. five randomly selected RRC-60-small is created, in order to account for the randomness in the selection of the training and test sets, and we report averaged performances over 5 independent runs including the standard deviation.

In order to offer a baseline for performance comparison we first replicated the methods of [28] (we used the code shared by the authors of [28]) and [4] on the observe and the reverse side images of RRC-60-small coins. More specifically, 1) we computed Dense SIFT features [18] on images of each side of the coins; 2) using Dense SIFT features, we computed matching costs between images by applying SIFT-Flow [17], which are used as dissimilarities between image pairs; 3) We applied NN-based classification as in [28], and by following the implementational stages explained in Section 5.1 and similarity fusion scheme in 5.2 we applied GTG; 4) Finally, the classification accuracy is computed by following the experimental setting that was adopting one and two training image per class in [4,28]. The results are reported in Table 2.

In Table 2, best performances obtained at observe and reverse side experiments are shown by italic font, while best performances obtained for each training set is shown by black-colored bold font. In a similar manner with [4], we got better performances with GTG compared to NN-based classifier employed in [28] in both observe and reverse side experiments. From the high standard deviations, we see that selection of images affect the performance significantly. Following the same implementation stages for both 1 and 2 labeled images per class experiments at the reverse side of coins (row 3 in Table 2), we get around 16% lower (when

largest deviations in performances are considered) performances for each method, i.e. [28] and [4], compared to results reported in [4]. We justify this performance gap by the existent differences in images of the RRC-60-small and the dataset used in [4,28], e.g. there are five non-overlapping classes between two datasets in addition to image differences in two datasets. Eventually, we base our performance comparisons between GTG-w/o prior [4], Zambanini et al. [28] and fusion scheme referring to the results in Table 2 instead of the ones in [4].

Next, we explored performance of GTG using both sides of coins. We experimented with a number of α values, i.e. $\{0, 0.1, \dots, 0.9, 1\}$ where $\alpha = 0$ and $\alpha = 1$ constitutes to classification using only reverse and observe sides, respectively. In particular at each experiment using different number of labeled images, we chose an α value for each of the five independent runs regarding to the best performance achieved. It can be seen in Table 2 that adopting information from both sides yields significant performance improvement over using single side information which demonstrates the importance of availability of images from both sides of coins for detecting their type automatically with higher accuracy.

6.2. RRC-60

Experimental setup. We present performance evaluation on RRC-60 using training sets in gradually increasing sizes, i.e. we first created the training, validation and test set splits (80%, 10% and 10% of the set of 6000 images of observe side of coins and 6000 images of corresponding reverse-sides of coins). Then, from the training set of each coin side, we randomly sampled 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70 and 80 images per class, that yields to 18 training sets with the sizes of 60, 120, ..., 4800. Moreover, similar to what we have done at the RRC-60-small experiments we created aforementioned sets five times by random selection of images.

We compared performance of GTG with ResNet152 [11] which is a state of the art classification architecture adopting deep learning scheme. Specifically, by initializing the architecture with ImageNet pretrained model, we finetuned it by RRC-60 training set images. Network training is accomplished by Stochastic Gradient Descent optimizer with momentum 0.9 and a learning rate of 0.001. The batch-size and the number of epochs are set to 20 and 30, respectively.

By using the features obtained by the fine-tuned models, we applied GTG by initializing strategy space in two ways: (i) *GTG-without prior*: with the assumption of no prior knowledge was available we assign same probability to all labels using Eqs. 1 and 2 (as in [4]); (ii) *GTG-with prior*: we initialized strategy space with the softmax output of the finetuned model.

Fusion schemes for GTG. As mentioned in Section 5.2, one way to apply GTG in a fusion scheme can be convex combination of affinity matrices obtained using features of each coin side. Another option could be computing an affinity matrix using concatenated (L_2 -normalized) features of each coin side. Using fine-tuned ResNet152 features, we present performances (in terms of classification accuracy on test set) obtained by both fusion schemes, i.e. Similarity fusion and Feature concatenation, and performances

Table 2
Classification accuracy on RRC-60-small, ([†]Re-implementation of [28] on RRC-60-small).

Method	Side	1 labeled img. per class	2 labeled img. per class
Zambanini et al. [28] [†]	Observe	45.9 ± 2.6	62.7 ± 3.1
GTG-w/o prior [4]	Observe	53.3 ± 2.9	65.3 ± 2.6
Zambanini et al. [28] [†]	Reverse	52.2 ± 2.7	65.8 ± 0.8
GTG-w/o prior [4]	Reverse	59.1 ± 3.5	70.7 ± 3.4
GTG w/o prior - Similarity fusion	Both	68.5 ± 2.5	81.3 ± 1.9

Table 3

Classification accuracy on RRC-60 test set using features of ResNet152 that is finetuned on increased number of training images. (O: Observe side, R: Reverse Side, B: Both sides).

Method	Number of labeled images per class									
	1	2	3	4	5	6	7	8	9	10
O-ResNet	10.7 ± 1.9	17.7 ± 1.6	27.5 ± 2.1	44.0 ± 2.8	59.1 ± 3.1	66.0 ± 2.9	73.9 ± 1.4	80.3 ± 1.9	81.8 ± 0.8	86.6 ± 1.9
O-GTG-w/o prior	14.0 ± 0.9	18.2 ± 1.8	29.3 ± 2.0	48.0 ± 3.0	64.3 ± 2.4	70.4 ± 1.5	76.4 ± 2.1	82.7 ± 1.8	84.5 ± 0.8	88.5 ± 2.0
O-GTG-w prior	12.8 ± 1.5	17.6 ± 1.9	27.6 ± 2.1	46.2 ± 2.0	62.9 ± 2.4	69.3 ± 1.9	76.4 ± 2.6	82.2 ± 2.0	84.1 ± 0.4	88.0 ± 1.8
R-ResNet	15.7 ± 1.1	31.4 ± 2.4	46.5 ± 2.1	63.1 ± 3.8	75.1 ± 2.8	80.9 ± 0.9	88.0 ± 1.0	90.0 ± 1.5	91.1 ± 1.3	93.3 ± 1.0
R-GTG-w/o prior	20.8 ± 1.2	34.1 ± 1.3	51.3 ± 3.6	68.8 ± 3.7	80.4 ± 2.3	84.8 ± 1.3	90.2 ± 0.7	92.1 ± 0.9	92.4 ± 1.3	94.1 ± 1.5
R-GTG-w prior	17.9 ± 1.3	32.8 ± 1.9	50.6 ± 4.3	67.3 ± 5.3	79.4 ± 2.0	84.7 ± 1.4	89.6 ± 0.7	92.3 ± 0.7	92.2 ± 1.5	94.1 ± 1.6
B-ResNet prior	16.9 ± 1.7	33.1 ± 2.7	49.3 ± 1.3	67.7 ± 2.6	79.9 ± 2.6	85.8 ± 0.9	91.6 ± 1.1	94.0 ± 1.3	94.7 ± 0.7	96.6 ± 0.8
B-GTG-w/o prior	22.2 ± 1.4	36.2 ± 2.3	58.3 ± 3.1	76.6 ± 3.2	86.4 ± 1.5	92.1 ± 1.1	94.9 ± 0.6	95.7 ± 1.4	96.6 ± 1.1	97.2 ± 0.5
B-GTG-w prior	20.4 ± 0.7	34.1 ± 2.4	56.6 ± 3.3	75.1 ± 2.0	86.6 ± 1.8	91.9 ± 1.2	95.0 ± 0.6	95.9 ± 1.5	96.7 ± 1.2	97.4 ± 0.4

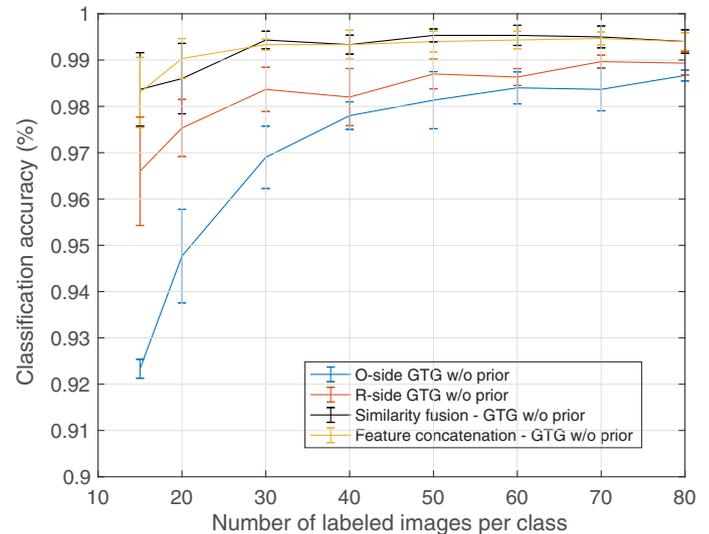
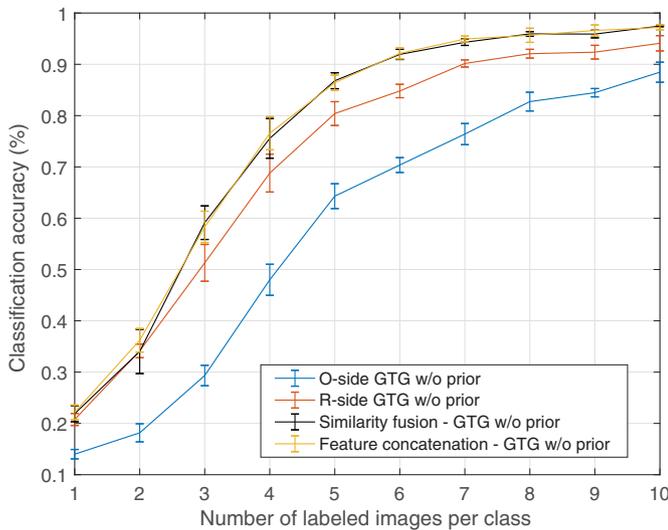


Fig. 5. Performance of fusion schemes for GTG using: (left) 1 to 10 labeled images per class; (right) 15 to 80 labeled images per class.

obtained by using one side, i.e. observe (O) and reverse (R), images at Fig. 5. In these experiments α parameter in Eq. 7 is tuned on validation set. In Fig. 5, we observe that while performances obtained using reverse side images are always better than observe side images, fusing information from both sides brings further performance improvement. Moreover, performances obtained by each fusion scheme is quite similar. We continued experiments using fusion of affinity matrices computed for each side.

Comparison of GTG with ResNet152. We present performances of ResNet152 and GTG (with and without prior) obtained by using each coin sides (O for observe and R for reverse) and fusion in Table 3 (using 1 to 10 labeled images per class) and Fig. 6 (using 15 to 80 labeled images per class). In Table 3 best performances obtained at observe and reverse side experiments are shown by italic font, while best performances obtained for each training set is shown by black-colored bold font.

In order to explore performance of ResNet152 by considering both coin side information, we applied convex combination (See Eq. 7) of softmax outputs obtained for each coin side by tuning α parameter on the validation set. Then, the class label is decided based on maximum value in the fused-softmax outputs. For GTG-with prior experiments we used this fused softmax-outputs as prior knowledge to start GTG. In Table 3 and Fig. 6, we see that performances obtained using reverse side images are always better than observe side images which is an expected outcome since inter-class variation at observe side images is lower. In Table 3, GTG is always better than ResNet152, especially performance gap is significant when only one labeled image per class is used. While

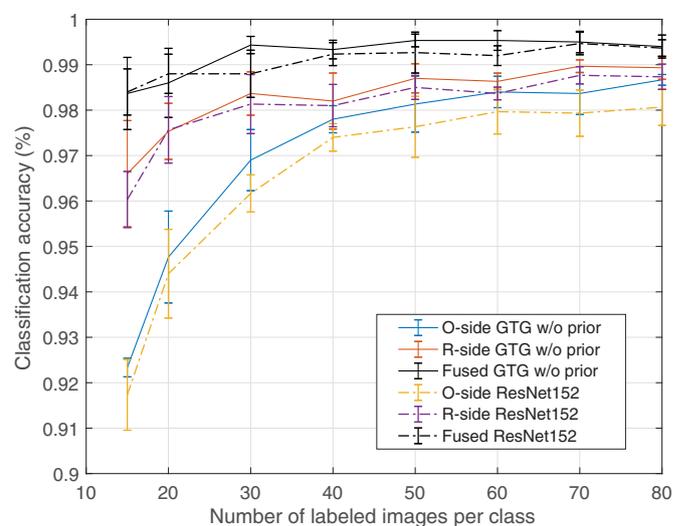


Fig. 6. Classification accuracy on RRC-60 test set using features of ResNet152 that is finetuned on increased number of labeled images. (Fusion performances are shown by Black-colored lines, ResNet152 performances are shown by dashed-lines.)

fusion always brings improvement for both GTG and ResNet152, performance improvement achieved by GTG-fusion scheme is quite high at experiments using 3 to 6 number of labeled images. GTG-without Prior usually performs better than GTG-with Prior

when the network is trained with few labeled images (1 to 6) per class. Training the network with few labeled images results with a peaked softmax towards wrong classes. GTG in case of very peaked prior is not able to correct the probabilities leading to still better than ResNet152 but poorer than GTG-without prior results. When the network is trained with 7 or more labeled images per class the performance of GTG-with prior scheme starts working as expected. For the larger sets, both GTG schemes perform very similar, thus we presented only without prior scheme for GTG at Fig. 6.

7. Conclusions

In this paper we publish a new image dataset for ancient coin classification exploiting both sides of the same coin. Based on the coin classes of a previously published dataset, we prepare an extension of it in terms of both quantity (from 180 to 6000 images) and diversity (includes both observe and reverse sides). Moreover, we proposed a principled extension of our previous method [4] by employing GTG using integrated pairwise similarities computed for each coin side. Our experimental results showed that the proposed fusion scheme provides significant performance improvement over the baseline which exploit only one side of each coin, for both shallow and deep features. Eventually, this result also demonstrates the importance of availability of a dataset composed of images from both sides of the same coin for better ancient coin classification accuracy.

Declaration of Competing Interest

This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.

Acknowledgment

The authors would like to thank Sebastian Zambanini for his invaluable advices on various technical issues and code sharing.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.patrec.2019.12.007.

References

- [1] H. Anwar, S. Zambanini, M. Kampel, Coarse-grained ancient coin classification using image-based reverse side motif recognition, *Mach. Vision Appl.* 26 (2–3) (2015) 295–304.
- [2] H. Anwar, S. Zambanini, M. Kampel, K. Vondrovec, Ancient coin classification using reverse motif recognition: image-based classification of roman republican coins, *IEEE Signal Process. Mag.* 32 (4) (2015) 64–74.
- [3] O. Arandjelovic, Automatic attribution of ancient roman imperial coins, in: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 1728–1734.
- [4] S. Aslan, S. Vascon, M. Pelillo, Ancient coin classification using graph transduction games, in: *2018 IEEE Int. Conf. on Metrology for Archaeology and Cultural Heritage*, 2018 (In Press).
- [5] O. Chapelle, B. Schölkopf, A. Zien, *Semi-Supervised Learning*, 1st, The MIT Press, 2010.
- [6] B. Conn, O. Arandjelović, Towards computer vision based ancient coin recognition in the wild-automatic reliable image preprocessing and normalization, in: *2017 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2017, pp. 1457–1464.
- [7] J. Cooper, O. Arandjelovic, Understanding ancient coin images, in: *INNS Big Data and Deep Learning conference*, 2019, pp. 330–340.
- [8] M.H. Crawford, *Roman republican coinage*, 2, Cambridge University Press, 1974.
- [9] I. Elezi, A. Torcinovich, S. Vascon, M. Pelillo, Transductive label augmentation for improved deep network learning, in: *24th Int. Conf. on Pattern Recognition (ICPR)*, 2018, pp. 1432–1437, doi:10.1109/ICPR.2018.8545524.
- [10] A. Erdem, M. Pelillo, Graph transduction as a noncooperative game, *Neural Comput.* 24 (3) (2012) 700–723.
- [11] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [12] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: *IEEE Conf. on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.
- [13] R. Huber-Mörk, S. Zambanini, M. Zaharieva, M. Kampel, Identification of ancient coins based on fusion of shape and local features, *Mach. Vision Appl.* 22 (6) (2011) 983–994.
- [14] R.A. Hummel, S.W. Zucker, On the foundations of relaxation labeling processes, *IEEE Trans. Pattern Anal. Mach. Intell.* (3) (1983) 267–287.
- [15] M. Kampel, M. Zaharieva, Recognizing ancient coins based on local features, in: *Int. Symp. on Visual Computing*, 2008, pp. 11–22.
- [16] J. Kim, V. Pavlovic, Discovering characteristic landmarks on ancient coins using convolutional networks, *J. Electron. Imag.* 26 (1) (2016) 011018.
- [17] C. Liu, J. Yuen, A. Torralba, Sift flow: dense correspondence across scenes and its applications, *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (5) (2011) 978–994.
- [18] D.G. Lowe, et al., Object recognition from local scale-invariant features., in: *IEEE Int. Conf. on Comp. Vision*, 99, 1999, pp. 1150–1157.
- [19] J. Nash, Non-cooperative games, *Ann. Math.* (1951) 286–295.
- [20] X. Pan, L. Tougne, Image analysis and deep learning for aiding professional coin grading, in: *2018 International Conference on Image and Video Processing, and Artificial Intelligence*, 10836, International Society for Optics and Photonics, 2018, p. 1083605.
- [21] S.-S. Parsa, M. Rastgarpour, M.M. Dehshibi, Statistical feature fusion for sassanian coin classification, in: *Recent Advances in Information and Communication Technology 2015*, Springer, 2015, pp. 75–84.
- [22] M. Pelillo, The dynamics of nonlinear relaxation labeling processes, *J. Math. Imag. Vision* 7 (4) (1997) 309–323.
- [23] I. Schlag, O. Arandjelovic, Ancient roman coin recognition in the wild using deep learning based recognition of artistically depicted face profiles, in: *IEEE Int. Conf. on Computer Vision Workshop (ICCVW)*, 2017, pp. 2898–2906.
- [24] L. Van Der Maaten, E. Postma, Towards automatic coin classification, *CiteSeer*, 2006.
- [25] S. Vascon, S. Aslan, A. Torcinovich, T. v. Laarhoven, E. Marchiori, M. Pelillo, Unsupervised domain adaptation using graph transduction games, in: *2019 International Joint Conference on Neural Networks (IJCNN)*, 2019, pp. 1–8, doi:10.1109/IJCNN.2019.8852075.
- [26] S. Vascon, M. Frasca, R. Tripodi, G. Valentini, M. Pelillo, Protein function prediction as a graph-transduction game, *Pattern Recognit. Lett.* (2018 (in press)).
- [27] S. Vascon, E.Z. Mequanint, M. Cristani, H. Hung, M. Pelillo, V. Murino, A game-theoretic probabilistic approach for detecting conversational groups, in: *ACCV 2014*, 2015, pp. 658–675.
- [28] S. Zambanini, M. Kampel, Coarse-to-fine correspondence search for classifying ancient coins, in: *ACCV*, 2012, pp. 25–36.
- [29] J. Weibull, *Evolutionary Game Theory*, MIT Press, 1997.
- [30] M. Zaharieva, M. Kampel, S. Zambanini, Image based recognition of ancient coins, in: *International Conference on Computer Analysis of Images and Patterns*, Springer, 2007, pp. 547–554.
- [31] S. Zambanini, M. Kampel, Automatic coin classification by image matching, in: *12th International conference on Virtual Reality, Archaeology and Cultural Heritage*, 2011, pp. 65–72.
- [32] S. Zambanini, A. Kavelar, M. Kampel, Classifying ancient coins by local feature matching and pairwise geometric consistency evaluation, in: *22nd Int. IEEE Conf. on Pattern Recognition (ICPR)*, 2014, pp. 3032–3037.
- [33] L. Zelnik-Manor, P. Perona, Self-tuning spectral clustering, in: *Advances in Neural Information Processing Systems (NIPS)*, 2005, pp. 1601–1608.