Pattern Recognition Letters 80 (2016) 84–90

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Deep Multitask Metric Learning for Offline Signature Verification[☆]

Amir Soleimani[a,*], Babak N. Araabi[a,b,*], Kazim Fouladi[a]

[a] Machine Learning and Computational Modeling Lab, Control and Intelligent Processing Center of Excellence, School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

[b] School of Cognitive Sciences, Institute for research in fundamental sciences, Tehran, Iran

ARTICLE INFO

Article history:
Received 18 November 2015
Available online 3 June 2016

Keywords:
Signature verification
Metric Learning
Multitask Learning
Transfer Learning

ABSTRACT

This paper presents a novel classification method, Deep Multitask Metric Learning (DMML), for offline signature verification. Unlike existing methods that to verify questioned signatures of an individual merely consider the training samples of that class, DMML uses the knowledge from the similarities and dissimilarities between the genuine and forged samples of other classes too. To this end, using the idea of multitask and transfer learning, DMML train a distance metric for each class together with other classes simultaneously. DMML has a structure with a shared layer acting as a writer-independent approach, that is followed by separated layers which learn writer-dependent factors. We compare the proposed method against SVM, writer-dependent and writer-independent Discriminative Deep Metric Learning method on four offline signature datasets (UTSig, MCYT-75, GPDSsynthetic, and GPDS960GraySignatures) using Histogram of Oriented Gradients (HOG) and Discrete Radon Transform (DRT) features. Results of our experiments show that DMML achieves better performance compared to other methods in verifying genuine signatures, skilled and random forgeries.

1. Introduction

Handwritten signature is one of the most socially acceptable and widespread personal attributes to authenticate individuals. Signature verification systems by comparing a questioned signature with reference samples, try to determine whether it is genuine or forgery. According to the acquisition approach, signatures are divided into offline and online. In offline, signatures are scanned and stored as grayscale or binary images, while in online, sequential information (e.g. x-y positions, velocity, acceleration, pressure, and pen inclination) describes signatures [21]. Generally, online signature verification systems have higher accuracy, whereas the performance drops considerably in offline modes.

Forgery in the signature verification literature is divided into two main categories: random forgery which is when forger signs regardless of genuine signature, and skilled (simulated, freehand, or simple) forgery which is when forger tries to simulate genuine one. However, some papers know simple forgery as a separate category defined as the time that forger has no attempt to mimic genuine signature [29] or when forger just knows genuine writer's name [30].

Hardships of offline signature verification systems lie in high intra-personal variability, limited number of training samples, inaccessibility of skilled forgeries in learning procedure, and forgers attempts to mimic genuine samples [32]. Previous works addressing such hardships can be categorized into employing better feature extraction, improving classification with limited number of samples, augmenting the datasets, and building model ensembles [16].

Writer-dependent (WD) and writer-independent (WI) are two approaches to design offline signature verification systems. In WD, a specialized classifier is trained separately for each individual based on his samples, but in WI, being trained by all authors samples, just one classifier determines the authenticity of questioned signatures [35]. WI by using samples from all authors can alleviate the problem of limited number of training samples [32], but it is probable that many writer-specific characteristics are missed in this approach.

Recently to solve the problem of face verification in the wild, some attempts have been done to use metric learning-based systems which learn good distance metric [17]. A good distance metric is a similarity measure that its output for similar samples is close to zero, and reversely for dissimilar ones, it is a large positive number. In the offline signature verification literature, various distance-based classifiers have been employed that their distance measures are mainly Euclidean and Mahalanobis [21], which are

☆ This paper has been recommended for acceptance by Jose Ruiz Jose Ruiz-Shulcloper.

* Corresponding authors. Tel.: +98-915-124-9267, +98-21-61119728.
E-mail addresses: a.soleimani.b@gmail.com (A. Soleimani), Araabi@ut.ac.ir (B.N. Araabi).

http://dx.doi.org/10.1016/j.patrec.2016.05.023

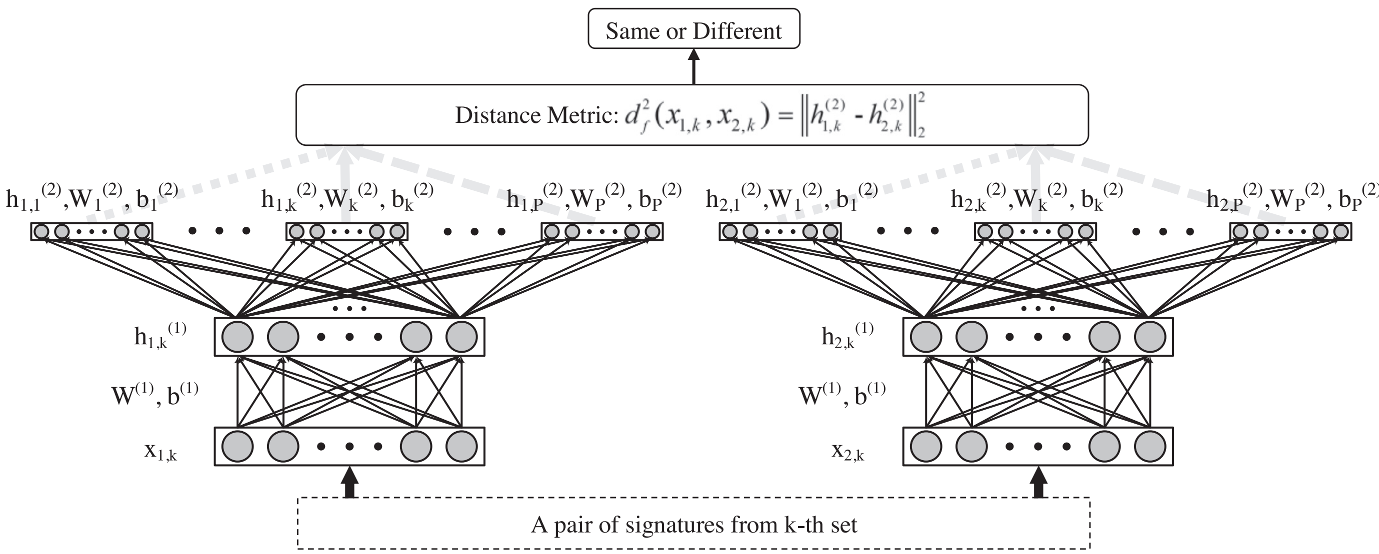

Fig. 1. Basic ideal of proposed DMML method. There is a shared layer for each pairs of signatures that is followed by separated layers which belong to distinct authentic individuals. To determine the authenticity of a signature $x_{2,\,k}$, which is claimed as a genuine sample of k-th person, questioned signature along with reference sample $x_{1,\,k}$ pass through the first layer (shared layer) and k-th unit in the second layer. Then distance metric determines whether they are the same or different. h, W, and b are respectively, output, weight and bias of a layer.

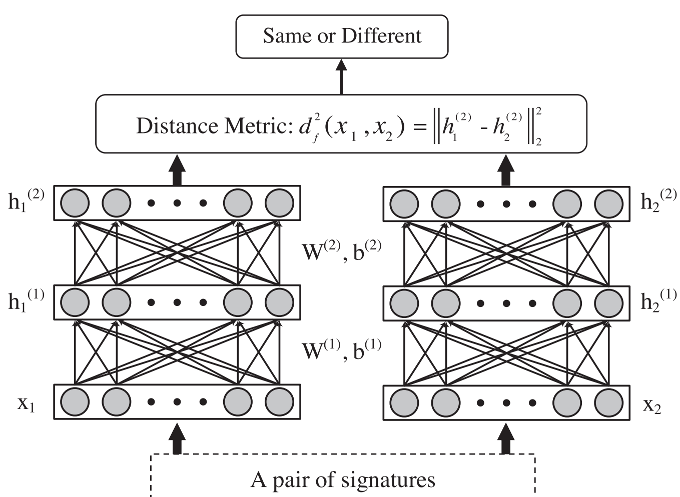

Fig. 2. The basic idea of DDML. A pair of objects (such as face, signature and ...) pass through hierarchical nonlinear transformations. Finally a distance metric determines whether they are the same or different.

not strong enough to discover the similarities and dissimilarities among genuine and forged samples. To our knowledge, this paper is the first to use metric learning based methods in the signature verification literature.

In this paper, we mix the idea of WD and WI approaches, multitask and transfer Learning with Discriminative Deep Metric learning (DDML) method [17]. Fig. 1 and Fig. 2 show the basic idea of DMML (proposed method) and DDML, respectively. DDML consists of a deep neural network that learns a set of hierarchical nonlinear transformations to make a good distance metric. We contribute to DDML by sharing a layer for all authors to involve WI approach and considering separated layers for each distinct authentic signer to handle WD approach. Our structure is benefited from the idea of multitask and transfer learning that helps to transfer the knowledge from the similarities and dissimilarities among other signers genuine and forged samples to each specific signer. Experiments on four offline signature datasets, UTSig [34], MCYT-75 [13], GPDSsynthetic [11], and GPDS960GraySignatures [12] indicate promising results for the proposed method.

The rest of this paper is organized as follows: Section 2 presents related works in metric learning specially DDML, and overviews the definition of transfer learning and multitask learning. Section 3 describes proposed DMML. Section 4 presents datasets details, experimental setups, and discusses results. Section 5 concludes the paper with some suggested works for the future.

2. Related works

2.1. Metric learning

Supervised metric learning methods aims to learn a metric or similarity measure from labeled data that outputs positive values close to zero for same-class data and large value for objects from different classes. Recently, some applications have benefited from these methods such as visual search [25], photo clustering [38], face verification [17], and person re-identification [23]. Among metric learning methods, DDML [17] by training a deep neural network to learns a set of hierarchical nonlinear transformations, shows promising results in face verification in the wild. Fig. 2 shows the basic idea of DDML. In that a pair of objects (in our application signatures), x_1 and x_2, are separately transformed from multiple layers with similar weights and biases for each object, then the similarity of the pair is calculated at the final layer ($h_1^{(2)}$ and $h_2^{(2)}$) by squared Euclidean distance.

Unlike DDML, which uses only common layers for all pairs, we use the idea of multitask learning to consider separated layers after a shared layer for different pairs from different sets (each set belongs to one authentic signer). The basic idea of our method is shown in Fig. 1. In that, squared Euclidean distance is calculated for pairs $x_{1,\,k}$ and $x_{2,\,k}$ from k-th set (k = 1,2,…,P) which pass through the k-th separate layer.

2.2. Multitask and Transfer Learning

When source domain and target domain, or source task and target task are not equal, transfer learning approaches can be used to transfer the knowledge from the source domain or the source task to improve the target task. Moreover multitask learning, which can be considered as a kind of transfer learning, without insisting on just one target, aims to learn all source and target task simultaneously using a shared representation [28].

In signature verification problem, transfer learning, in general, can help to transfer the knowledge from the similarities and dissimilarities between genuine and forged signatures of one or more authentic individuals to better verify samples from a new person. In other words, it is similar to a forensic expert who learns from his experiences in previous forensic cases to determine a new sample's authenticity. Multitask learning, as a method of transfer learning, focuses to learn similarities and dissimilarities between genuine and forged signatures from all authentic individual, including the new person, simultaneously. In other words, multitask learning does not just emphasis on a new person by learning and transferring knowledge from previous individuals, it seeks to use shared knowledge among different individuals simultaneously.

So far some transfer learning approaches have been applied to many applications such as image classification [40], natural language processing (NLP) [5], WiFi localization [27], cross-language classification [22], opinion mining [20], expression recognition [7], letter-recognition and news groups [9], sentiment classification [41] and action similarity [1]. Specially a deep transfer metric learning (DTML) for cross-dataset visual recognition is introduced in [18]. In that, such as DDML a deep network by a set of hierarchical nonlinear transformation is used, but in contrast, by a new cost function, it transfers knowledge from labeled source domain to unlabeled target domain. To this end, it tries to minimize the distribution divergence between labeled source domain and unlabeled target domain, minimize intra-class variations, and maximize inter-class variations. DTML does not seem to be useful in signature verification problem since to achieve good accuracy, many different authentic individuals' samples must be considered as source domain, and this produces source domain distribution useless, specially when DTML framework minimizes the distribution divergence between the source domain and target domain by minimizing their mean distance.

Multitask Learning methods have been recently employed in many areas such as WiFi localization [42], computer aided design (CAD) [31], Spam filtering [3], text classification [19], web search [6], multilabel image classification [24], and phoneme recognition [33].

In this paper we propose multitask learning version of DDML to use both the power of DDML in learning complex distance metric, and the ability of multitask learning approach to improve the accuracy.

3. Proposed methods

Consider P sets, $X_k = \{g_{1,k}, g_{2,k}, \dots, g_{G_k,k}, f_{1,k}, f_{2,k}, \dots, f_{F_k,k}\}$ (k = 1,2,…,P) as the k-th training set corresponding to k-th authentic individual, where $g_{i,\,k} \in \Re^{d}$ and $f_{i,\,k} \in \Re^{d}$ are d-dimensional feature vectors extracted from i-th genuine and forged images, respectively. For k-th set there are G_k genuine samples ($g_{i,k}, i = 1, 2, \dots, G_k$) and F_k forged samples ($f_{i,k}, i = 1, 2, \dots, F_k$). Forgeries can be whether skilled forgery or genuine and forged samples of other individuals as random forgery.

For each set, we consider $\binom{G_k}{2} + G_k$ similar pairs $(x_{i,k}, x_{j,k}) = (g_{n,k}, g_{m,k})$, $(n = 1, 2, \dots, G_k\ ,\ m = 1, 2, \dots, G_k)$ with label 1 $(l_{ij} = 1)$ and G_kF_k dissimilar pairs $(x_{i,k}, x_{j,k}) = (g_{n,k}, f_{m,k})$, $(n = 1, 2, \dots, G_k$, $m = 1, 2, \dots, F_k)$ with label −1 $(l_{ij} = -1)$. As shown in Fig. 1, our approach is a multitask learning-based version of DDML, in which passing through common layers, pairs from k-th set construct $h^{(2)}_{1,k}$ and $h^{(2)}_{2,k}$ at the second layer by k-th separated units.

Considering a shared layer for pairs from k sets, leads the network to learn a representation that discover underlying shared subsets of factors among different tasks [4], in addition separated layers help the network to learn writer-specific factors for different sets correspond to authentic individuals, better. In other words, shared layer acts as a WI approach and separated ones insist on WD approach.

Assume there are $p^{(1)}$ and $p^{(2)}$ nodes in first layer and second layers, respectively. The outputs of first layer are $h^{(1)}_{1,k} = s(W^{(1)}x_{1,k} + b^{(1)})$ and $h^{(1)}_{2,k} = s(W^{(1)}x_{2,k} + b^{(1)})$, where $W^{(1)} \in \Re^{p^{(1)}\times d}$ and $b^{(1)} \in \Re^{p^{(1)}\times 1}$ are weights matrix and bias vector, respectively, and $s : \Re \longrightarrow \Re$ is a nonlinear operator such as tanh function. Then respectively $h^{(1)}_{1,k}$ and $h^{(1)}_{2,k}$ build $h^{(2)}_{1,k} = s(W^{(2)}_k h^{(1)}_{1,k} + b^{(2)}_k)$ and $h^{(2)}_{2,k} = s(W^{(2)}_k h^{(1)}_{2,k} + b^{(2)}_k)$, where $W^{(2)}_k \in \Re^{p^{(2)}\times p^{(1)}}$ and $b^{(2)}_k \in \Re^{p^{(2)}\times 1}$ are k-th weights matrix and k-th bias vector, and s is a nonlinear operator again. Finally, distance metric is calculated by squared Euclidean distance for each pair of signature samples $(x_{1,\,k}, x_{2,\,k})$ as:

$$d^2_f(x_{1,k}, x_{2,k}) = \|h^{(2)}_{1,k} - h^{(2)}_{2,k}\|^2_2 \tag{1}$$

DMML is formulated as the bellow optimization problem:

$$\begin{aligned}\arg\min_{min} J = &\frac{1}{2}\sum_{k=1}^{P}\sum_{i,j} g\Big(1 - l_{i,j}(\tau - d^2_f(x_{i,k}, x_{j,k}))\Big)\\ &+ \frac{\lambda}{2}(\|W^{(1)}\|^2_F + \|b^{(1)}\|^2_2)\\ &+ \frac{\lambda}{2}\sum_{k=1}^{P}(\|W^{(2)}_k\|^2_F + \|b^{(2)}_k\|^2_2)\end{aligned} \tag{2}$$

in that $g(z) = \frac{1}{\beta}\log(1 + \exp(\beta z))$ is the smoothed approximation for $[z]_+ = \max(z, 0)$, β determines sharpness parameter, $\|A\|_F$ is Frobenius norm, and $\lambda \geq 0$ is a regularization parameter to avoid over-fitting. The difference between DMML and DDML cost function is two summations for P distinct authors that involve writer-dependent' weights and biases.

Similar to the technique proposed by [17], τ is a threshold that determines output labels as bellow:

$$l_{i,j} = \begin{cases} 1 & d^2_f(x_{i,k}, x_{j,k}) < \tau - 1 \\ -1 & d^2_f(x_{i,k}, x_{j,k}) > \tau + 1 \end{cases} \tag{3}$$

To solve the optimization problem in Eq. 2 and update weights and biases, similar to [17] we use sub-gradient descent and need the gradient of the objective function J with respect to weights and biases:

$$\frac{\partial J}{\partial W^{(1)}} = \sum_{i,j}(\Delta^{(1)}_{i,j}x_i^T + \Delta^{(1)}_{j,i}x_j^T) + \lambda W^{(1)} \tag{4}$$

$$\frac{\partial J}{\partial W^{(2)}_k} = \sum_{i,j}(\Delta^{(2)}_{i,j,k}h^{(1)T}_{i,k} + \Delta^{(2)}_{j,i,k}h^{(1)T}_{j,k}) + \lambda W^{(2)}_k \tag{5}$$

$$\frac{\partial J}{\partial b^{(1)}} = \sum_{i,j}(\Delta^{(1)}_{i,j} + \Delta^{(1)}_{j,i}) + \lambda b^{(1)} \tag{6}$$

$$\frac{\partial J}{\partial b^{(2)}_k} = \sum_{i,j}(\Delta^{(2)}_{i,j} + \Delta^{(2)}_{j,i}) + \lambda b^{(2)}_k \tag{7}$$

where:

$$\Delta^{(2)}_{i,j,k} = g'(c)l_{i,j}(h^{(2)}_{i,k} - h^{(2)}_{j,k}) \odot s'(z^{(2)}_i) \tag{8}$$

$$\Delta^{(2)}_{j,i,k} = g'(c)l_{i,j}(h^{(2)}_{j,k} - h^{(2)}_{i,k}) \odot s'(z^{(2)}_j) \tag{9}$$

Input: P training sets $X_k = \{(x_{i,k}, x_{j,k}, l_{i,j})\}, k = 1, 2, ..., \mathrm{P}$, learning rate μ, number of iterations $\mathrm{I_t}$, regularization parameter λ, and convergence error ϵ.

Outputs: weights and biases: $W^{(1)}, W_k^{(2)}, b^{(1)}, b_k^{(2)}$, $(k = 1, 2, ..., \mathrm{P})$.
Initialization initialize weights and biases.
Optimization by back propagation

for $t = 1, 2, ..., \mathrm{I_t}$
  for $k \in \{1, 2, ..., \mathrm{P}\}$ in random order
    randomly select a pair X_k
    // Forward Propagation //
      Calculate $h_{i,k}^{(1)}, h_{j,k}^{(1)}, h_{i,k}^{(2)}, h_{j,k}^{(2)}$.
    // Calculating gradients //
      Calculate gradients by Eq. 4-7
    // Back Propagation //
      Update weights and biases by Eq. 15-18
  end
  Calculate J by Eq. 2.
  if $t > 1$ and $|J_t - J_{t-1}| < \epsilon$, **Return**
end
Return: $W^{(1)}, W_k^{(2)}, b^{(1)}, b_k^{(2)}$, $(k = 1, 2, ..., \mathrm{P})$.

Fig. 3. DMML learning algorithm.

$$\Delta_{i,j}^{(1)} = (W_k^{(2)^T} \Delta_{i,j,k}^{(2)}) \odot s'(z_i^{(1)}) \tag{10}$$

$$\Delta_{j,i}^{(1)} = (W_k^{(2)^T} \Delta_{j,i,k}^{(2)}) \odot s'(z_j^{(1)}) \tag{11}$$

In above equations $\odot$ means element-wise multiplication. c, $z_u^{(2)}$, and $z_u^{(1)}$ $(u = i, j)$ are calculated by Eq. 16–14:

$$c \triangleq 1 - l_{i,j}(\tau - d_f^2(x_{i,k}, x_{j,k})) \tag{12}$$

$$z_u^{(2)} \triangleq W_k^{(2)} h_{u,k}^{(1)} + b_k^{(2)} \tag{13}$$

$$z_u^{(1)} \triangleq W^{(1)} x_{u,k} + b^{(1)} \tag{14}$$

Finally weights and biases are updated by steepest descent method (Eq. 15)–(18) with learning rate μ:

$$W^{(1)} = W^{(1)} - \mu \frac{\partial J}{\partial W^{(1)}} \tag{15}$$

$$W_k^{(2)} = W_k^{(2)} - \mu \frac{\partial J}{\partial W_k^{(2)}} \tag{16}$$

$$b^{(1)} = b^{(1)} - \mu \frac{\partial J}{\partial b^{(1)}} \tag{17}$$

$$b_k^{(2)} = b_k^{(2)} - \mu \frac{\partial J}{\partial b_k^{(2)}} \tag{18}$$

Fig. 3 describes proposed DMML learning algorithm.

4. Experiments

In this section, we compare the performance of proposed DMML with DDML (both WI and WD approach) and support vector machine (SVM) classifier. For WI we train one DDML for all authentic individuals but in WD, each person has its own weights and biases.

4.1. Datasets and Experimental Settings

We used four offline signature datasets, UTSig[1] [34], MCYT-75 [13], GPDSsynthetic [11], and GPDS960GraySignatures [12]. UTSig is a new Persian offline signature, which consists of totally 8280 images and 115 sets. Each set has 27 genuine signatures of an authentic person, 3 opposite-hand signed and 42 skilled forged samples (42 + 3 = 45 forgeries for each person). MCYT-75 is a Spanish offline signature dataset with 2250 images and 75 sets. Each set consists of 15 genuine and 15 skilled forgery. GPDSsynthetic is a computer-generated dataset that was produced following the previous attempt to automatically generate signatures of new identities [10]. This offline dataset contains 4000 sets and each set has 24 genuine and 30 forged samples [11]. GPDS960GraySignatures is the gray version of GPDS960Signature dataset [36] that has 881 sets and in average, each set consists of 24 genuine and 30 forged samples [12].

DDML and DMML need some parameters that must be determined before the training, we consider $\tau = 3$, $\mu = 10^{-3}$, $\lambda = 10^{-2}$ and tan as nonlinear operation that have better results in [17]. Moreover, for the number of nodes in each layer we empirically set 250 nodes for UTSig, which has more complex signatures, and 200 nodes for other datasets.

To train DMML, DDML, and SVM we used similar experimental setups. For UTSig we build 115 training sets from 12 genuine sample and $570 = 5 \times (115 - 1)$ random forgeries, this produces $78 = \binom{12}{2} + 12$ positive and $6840 = 12 \times 5 \times (115 - 1)$ negative pairs. For MCYT-75, 10 genuine and $74 = 1 \times (75 - 1)$ random forgeries are used for each 75 sets. Similarly, that produces $55 = \binom{10}{2} + 10$ positive and $740 = 10 \times 1 \times (75 - 1)$ negative pairs. We repeated experiments for MCYT-75 with 5 genuine samples too. For each set of GPDSsynthetic our default was using 10 genuine signatures, 74 and 149 random forgeries for experiments on 75 and 150 users, respectively, and 299 random forgeries for test on 881, 1500, 2500 and 4000 users that similarly made their corresponding pairs. Furthermore, for GPDS960GraySignatures sets, we considered 10 genuine samples and 74 random forgeries for experiments on 75 users, and 299 random forgeries for test on 300 and 881 users.

To evaluate systems, for each set in training set, we build pairs $(x_{i,\,k}, x_{j,\,k})$ as testing pairs, where $x_{i,\,k}$ is genuine samples used in training phase (reference samples) and $x_{j,\,k}$ is remaining genuine samples from the k-th authentic person, his skilled forged samples and remaining random forged samples (genuine and skilled forged samples of the others). In UTSig k-th testing set, there are $180 = 12 \times (27 - 12)$ similar pairs, $540 = 12 \times 45$ dissimilar pairs as skilled forgery, and $82080 = 12 \times 114 \times (72 - 12)$ dissimilar pairs as random forgery. For MCYT-75 in k-th testing set, there are $50 = 10 \times (15 - 10)$ similar pairs, $150 = 10 \times 15$ dissimilar pairs as skilled forgery, and $14800 = 10 \times 74 \times (30 - 10)$ dissimilar pairs as random forgery. Similarly, for testing with 5 genuine samples, there are 50, 75, and 9250 pairs, respectively. In the case of GPDSsynthetic and GPDS960GraySignatures, we followed the above rules, except for random forgery that similar to [12], we defined it as just genuine samples of the others.

[1] UTSig dataset is available in Machine Learning and Computational Modeling (MLCM) lab: http://mlcm.ut.ac.ir/Datasets.html

Table 1
Results on UTSig dataset. Training with 12 genuine samples.

Feature	Method	FRR(%)	RF*:FAR(%)	SF#:FAR(%)	SF#:EER(%)
	SVM	26.09	0.03	15.85	20.63
HOG	DDML (WI)	23.71	0.01	16.73	19.78
	DDML (WD)	29.16	0.00	10.28	18.84
	DMML	18.96	0.00	16.15	17.45
	SVM	32.35	0.01	23.44	27.64
DRT	DDML (WI)	36.32	0.00	13.14	23.03
	DDML (WD)	35.78	0.00	10.02	22.11
	DMML	32.42	0.00	9.09	20.28

* RF–Random Forgery
# SF–Skilled Forgery

Table 2
Results on MCYT-75 dataset. Training with 10 genuine samples.

Features	Method	FRR(%)	RF*:FAR(%)	SF#:FAR(%)	SF#:EER(%)
	SVM	12.53	0.03	13.16	12.80
HOG	DDML (WI)	4.53	0.14	22.67	12.23
	DDML (WD)	21.93	0.00	4.18	11.58
	DMML	6.13	0.00	12.71	9.86
	SVM	12.27	0.01	19.11	15.78
DRT	DDML (WI)	14.82	0.01	8.27	12.01
	DDML (WD)	8.06	0.00	12.74	11.65
	DMML	10.73	0.00	10.02	10.06

* RF–Random Forgery
# SF–Skilled Forgery

Note for determining the authenticity of a questioned signature, 12 distances correspond to 12 reference samples are calculated for UTSig, 10 or 5 distances for MCYT-75, and 10 distances for two other datasets. So many fusion techniques can be applied to obtain one distance such as $\min$, $\max$, and *mean*. In this study we empirically choose $\min$ as the fusion technique, so if a questioned signatures has at least a metric distance below the threshold ($d_f^2(x_{references}, x_{questioned}) < \tau - 1$), it is genuine.

To extract features, images are divided to 5×5 blocks and then 16-bin Histogram of Oriented Gradients (HOG) features [8] are calculated. It builds 400-dimensional features ($5 \times 5 \times 16$). In order to achieve better results, we use $\log$ function to make the distribution of features more uniform and then PCA is applied. Moreover, we extracted Discrete Radon Transform (DRT) features similar to the methods presented by [26]. For GPDSsynthetic, because of large number of users, we just used HOG that builds features with lower dimension. Furthermore for GPDS960GraySignatures dataset, which is not accessible to us, we employed the exact Local Binary Pattern (LBP) feature vector used by [12].

To compare systems, we report False Rejection Rate (FRR), which is when genuine signature is rejected, and False Acceptance Rate (FAR) separately for random forgery and skilled forgery, which are respectively when random forged and skilled forged sample are accepted. Furthermore, Equal Error Rate (EER) by changing the threshold is reported too. Results are the average of 10 different experiments.

4.1.1. Results

Tables 1–4 show the results of 4 different classifiers using HOG and DRT features on UTSig, MCYT-75 with 10 and 5 genuine samples, and GPDSsynthetic datasets, respectively. In these datasets, DMML has the best results for EER, simultaneously FRR and skilled forgery FAR, and FAR for random forgery. Two structures of DDML have relatively better accuracy than SVM, and in all experiments, WD approach overtakes WI.

Table 5–8 compare results of DMML and other published methods in the literature with similar number of genuine samples in training. Table 5 shows that for UTSig, DMML surpasses the only

Table 3
Results on MCYT-75 dataset. Training with 5 genuine samples.

Features	Method	FRR(%)	RF*:FAR(%)	SF#:FAR(%)	SF#:EER(%)
	SVM	15.47	0.03	13.42	14.67
HOG	DDML (WI)	5.09	0.24	24.87	14.56
	DDML (WD)	25.15	0.00	4.89	14.01
	DMML	14.80	0.00	12.44	13.44
	SVM	25.20	0.01	13.96	17.33
DRT	DDML (WI)	9.10	0.00	25.01	17.01
	DDML (WD)	24.17	0.00	8.02	16.50
	DMML	17.23	0.00	14.31	15.89

* RF–Random Forgery
# SF–Skilled Forgery

Table 4
Results on GPDSsynthetic dataset using HOG features and training with 10 Genuine samples for 2500 signers.

Method	FRR(%)	RF*:FAR(%)	SF#:FAR(%)	SF#:EER(%)
SVM	5.80	0.58	29.49	15.64
DDML (WI)	5.50	0.31	26.24	15.03
DDML (WD)	6.10	0.17	24.76	14.89
DMML	6.51	0.11	18.23	12.80

* RF–Random Forgery
# SF–Skilled Forgery

Table 5
Comparison between proposed and previous method for UTSig dataset.

Method	FRR(%)	RF*: FAR(%)	SF#: FAR(%)	SF#: EER(%)
[34]	39.27	0.08	21.29	29.71
DMML	18.96	0.00	16.15	17.45

* RF–Random Forgery
# SF–Skilled Forgery

method on this dataset in which we used fixed-point arithmetic feature and SVM [34]. According to Table 6, for MCYT-75, while DMML overtakes previous published results in terms of random forgery EER, its skilled forgery performance is in competition with others. In other words, the proposed method has better skilled forgery EER than the methods in [2,39], and [26], but that of [13,14], and [37] surpass DMML. Table 7 indicates the results of DMML using HOG features and those reported in GPDSsynthetic website [15] using method employed by [12], for different number of users. While DMML random forgery ERR is acceptable, its skilled forgery EER overtakes previous ones. Moreover, Table 8 compares the results of method used by [12] and DMML using the exact similar features for 75, 300 and 881 users of GPDS960GraySignatures. Similarly while DMML random forgery EER is in competition with the previous method, its skilled forgery performance surpass that.

5. Conclusion

In this paper, by employing the idea of multitask and transfer learning, we mix WI and WD approaches in signature verification to learn a distance metric that measures the similarity between pairs of signatures. This structure helps to use the knowledge form the similarity and dissimilarity of genuine and forged samples of others, to achieve better results. In other words, DMML is a multitask leaning verison of DDML, in which we assume one shared layer as a first layer that is followed by separated layers for all signers. Shared layer helps the network to learn a representation that discovers underlying shared factors among signatures of different individuals, while separated layers try to learn writer-specific factors.

We used four different offline signature datasets (UTSig, MCYT-75, GPDSsynthetic, and GPDS960GraySignature) to evaluate our

Table 6
Comparison between proposed and other published methods for MCYT-75 dataset.

Method	Training with 10 genuine samples		Training with 5 genuine samples	
	RF* : EER(%)	SF#: ERR(%)	RF* : EER(%)	SF#: ERR(%)
[2]	5.85	20	9.34	22.4
[39]	NA	NA	NA	15.02
[26]	NA	9.87	NA	13.86
[13]	1.14	9.28	2.69	11
[37]	0.57	7.08	2.43	11.28
[14]	1.18	6.44	2.18	10.18
DMML	0.37	9.86	1.73	13.44

* RF–Random Forgery
# SF–Skilled Forgery

Table 7
Comparison between proposed and other published method for GPDSsynthetic dataset.

Method	Reported in [15] website		DMML	
Number of users	RF* : EER(%)	SF#: ERR(%)	RF* : EER(%)	SF#: ERR(%)
75	0.76	16.01	1.08	12.83
150	0.75	15.08	1.13	12.67
881	0.63	15.19	1.23	12.14
1500	0.79	15.13	1.31	12.43
2500	0.74	15.89	1.60	12.80
4000	0.79	16.44	1.98	13.30

* RF–Random Forgery
# SF–Skilled Forgery

Table 8
Comparison between proposed and other published method for GPDS960GraySignatures dataset using exactly the same features.

Method	Ferrer et al. [12]		DMML	
Number of users	RF* : EER(%)	SF#: ERR(%)	RF* : EER(%)	SF#: ERR(%)
75	0.72	15.56	1.30	15.08
300	2.00	23.03	2.15	20.94
881	3.08	23.58	3.20	22.76

* RF–Random Forgery
# SF–Skilled Forgery

proposed method. Results show that using similar features extraction method and experimental setups, DMML outperforms SVM and DDML whether WI or WD. Moreover, in comparison with other published methods with similar number of reference samples, DDML has better skilled forgery EER for UTSig, GPDSsynthetic, and GPDS960GraySignature datasets. Meanwhile, our method achieves competitive random forgery EER for all used datasets.

Considering more than one reference signature in training phase, DMML produces more than one number as distance metric for each questioned signature, so the study of different fusion techniques can be done in the future. Furthermore, for future works the effect of employing different features can be studied in two ways, since DMML structure let us whether mix all features and train one DMML or train different DMML for different features and finally use another fusion at the output.

References

[1] Z. Al-Halah, L. Rybok, R. Stiefelhagen, Transfer metric learning for action similarity using high-level semantics, Pattern Recog. Lett. (2015).
[2] F. Alonso-Fernandez, M.C. Fairhurst, J. Fierrez, J. Ortega-Garcia, Automatic measures for predicting performance in off-line signature, in: Proceedings of IEEE International Conference on Image Processing, 2007. ICIP 2007, vol. 1, IEEE, 2007, pp. I–369.
[3] J. Attenberg, K. Weinberger, A. Dasgupta, A. Smola, M. Zinkevich, Collaborative email-spam filtering with the hashing trick, in: Proceedings of the Sixth Conference on Email and Anti-Spam, 2009.
[4] Y. Bengio, A. Courville, P. Vincent, Representation learning: A review and new perspectives, Pattern Analysis and Machine Intelligence, IEEE Transactions on 35 (8) (2013) 1798–1828.
[5] J. Blitzer, R. McDonald, F. Pereira, Domain adaptation with structural correspondence learning, in: Proceedings of the 2006 conference on empirical methods in natural language processing, Association for Computational Linguistics, 2006, pp. 120–128.
[6] Y. Chang, J. Bai, K. Zhou, G.-R. Xue, H. Zha, Z. Zheng, Multi-task learning to rank for web search, Pattern Recognition Letters 33 (2) (2012) 173–181.
[7] J. Chen, X. Liu, Transfer learning with one-class data, Pattern Recognition Letters 37 (2014) 32–40.
[8] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 1, IEEE, 2005, pp. 886–893.
[9] M. Fang, Y. Guo, X. Zhang, X. Li, Multi-source transfer learning based on label shared subspace, Pattern Recognition Letters 51 (2015) 101–106.
[10] M.A. Ferrer, M. Diaz-Cabrera, A. Morales, Synthetic off-line signature image generation, in: Biometrics (ICB), 2013 International Conference on, IEEE, 2013, pp. 1–7.
[11] M.A. Ferrer, M. Diaz-Cabrera, A. Morales, Static signature synthesis: A neuromotor inspired approach for biometrics, Pattern Analysis and Machine Intelligence, IEEE Transactions on 37 (3) (2015) 667–680.
[12] M.A. Ferrer, J. Vargas, A. Morales, A. Ordóñez, Robustness of offline signature verification based on gray level features, Information Forensics and Security, IEEE Transactions on 7 (3) (2012) 966–977.
[13] J. Fiérrez-Aguilar, N. Alonso-Hermira, G. Moreno-Marquez, J. Ortega-Garcia, An off-line signature verification system based on fusion of local and global information, in: Biometric Authentication, Springer, 2004, pp. 295–306.
[14] A. Gilperez, F. Alonso-Fernandez, S. Pecharroman, J. Fierrez, J. Ortega-Garcia, Off-line signature verification using contour features (2008).
[15] GPDS, GPDS Synthetic Signature Database Website, Accessed: 2016.03.05.
[16] L.G. Hafemann, R. Sabourin, L.S. Oliveira, Offline handwritten signature verification - literature review, CoRR abs/1507.07909 (2015).
[17] J. Hu, J. Lu, Y.-P. Tan, Discriminative deep metric learning for face verification in the wild, in: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, IEEE, 2014, pp. 1875–1882.
[18] J. Hu, J. Lu, Y.-P. Tan, Deep transfer metric learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 325–333.
[19] P. Huang, G. Wang, S. Qin, A novel learning approach to multiple tasks based on boosting methodology, Pattern Recognition Letters 31 (12) (2010) 1693–1700.
[20] P. Huang, G. Wang, S. Qin, Boosting for transfer learning from multiple data sources, Pattern Recognition Letters 33 (5) (2012) 568–579.
[21] D. Impedovo, G. Pirlo, Automatic signature verification: the state of the art, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 38 (5) (2008) 609–635.
[22] X. Ling, G.-R. Xue, W. Dai, Y. Jiang, Q. Yang, Y. Yu, Can chinese web pages be classified with english data source? in: Proceedings of the 17th international conference on World Wide Web, ACM, 2008, pp. 969–978.
[23] V.E. Liong, J. Lu, Y. Ge, Regularized local metric learning for person re-identification, Pattern Recognition Letters (2015).
[24] Y. Luo, D. Tao, B. Geng, C. Xu, S.J. Maybank, Manifold regularized multitask learning for semi-supervised multilabel image classification, Image Processing, IEEE Transactions on 22 (2) (2013) 523–536.
[25] M. Norouzi, D.M. Blei, R.R. Salakhutdinov, Hamming distance metric learning, in: Advances in neural information processing systems, 2012, pp. 1061–1069.
[26] S.Y. Ooi, A.B.J. Teoh, Y.H. Pang, B.Y. Hiew, Image-based handwritten signature verification using hybrid methods of discrete radon transform, principal component analysis and probabilistic neural network, Applied Soft Computing 40 (2016) 274–282.
[27] S.J. Pan, D. Shen, Q. Yang, J.T. Kwok, Transferring localization models across space., in: AAAI, 2008, pp. 1383–1388.
[28] S.J. Pan, Q. Yang, A survey on transfer learning, Knowledge and Data Engineering, IEEE Transactions on 22 (10) (2010) 1345–1359.
[29] R. Plamondon, G. Lorette, Automatic signature verification and writer identificationthe state of the art, Pattern recognition 22 (2) (1989) 107–131.

[30] K. Radhika, M. Venkatesha, G. Sekhar, An approach for on-line signature authentication using zernike moments, Pattern Recognition Letters 32 (5) (2011) 749–760.

[31] V.C. Raykar, B. Krishnapuram, J. Bi, M. Dundar, R.B. Rao, Bayesian multiple instance learning: automatic feature selection and inductive transfer, in: Proceedings of the 25th international conference on Machine learning, ACM, 2008, pp. 808–815.

[32] D. Rivard, E. Granger, R. Sabourin, Multi-feature extraction and selection in writer-independent off-line signature verification, International Journal on Document Analysis and Recognition (IJDAR) 16 (1) (2013) 83–103.

[33] M.L. Seltzer, J. Droppo, Multi-task learning in deep neural networks for improved phoneme recognition, in: Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, IEEE, 2013, pp. 6965–6969.

[34] A. Soleimani, K. Fouladi, B.N. Araabi, Utsig: A persian offline signature database, Submitted Manuscript, Submitted at IET Biometrics in Aug 2015.

[35] S.N. Srihari, A. Xu, M.K. Kalera, Learning strategies and classification methods for off-line signature verification, in: Frontiers in Handwriting Recognition, 2004. IWFHR-9 2004. Ninth International Workshop on, IEEE, 2004, pp. 161–166.

[36] F. Vargas, M. Ferrer, C. Travieso, J. Alonso, Off-line handwritten signature gpds-960 corpus, in: icdar, IEEE, 2007, pp. 764–768.

[37] J.F. Vargas, M.A. Ferrer, C. Travieso, J.B. Alonso, Off-line signature verification based on grey level information using texture features, Pattern Recognition 44 (2) (2011) 375–385.

[38] M. Wang, D. Ji, Q. Tian, X.-S. Hua, Intelligent photo clustering with user interaction and distance metric learning, Pattern Recognition Letters 33 (4) (2012) 462–470.

[39] J. Wen, B. Fang, Y.Y. Tang, T. Zhang, Model-based signature verification with rotation invariant features, Pattern Recognition 42 (7) (2009) 1458–1466.

[40] P. Wu, T.G. Dietterich, Improving svm accuracy by training on auxiliary data sources, in: Proceedings of the twenty-first international conference on Machine learning, ACM, 2004, p. 110.

[41] Y. Zhang, X. Hu, P. Li, L. Li, X. Wu, Cross-domain sentiment classification-feature divergence, polarity divergence or both? Pattern Recognition Letters 65 (2015) 44–50.

[42] V.W. Zheng, S.J. Pan, Q. Yang, J.J. Pan, Transferring multi-device localization models using latent multi-task learning., in: AAAI, vol. 8, 2008, pp. 1427–1432.