Supplementary material-Kinship Identification through Joint Learning Using Kinship Verification Ensemble

Anonymous ECCV submission

Paper ID 4125

1 Pretrained Models (in Anonymous)

The pretrained models can be download at https://drive.google.com/open?id=18rVOqfA6gg4_wpP661KflDgFG5Fk9ZJ_ (in anonymous)

2 Quantity results

The *Joint Learning* method can alleviate confusion and predict more precise results.

Table 1: Identification Results of Different Approaches on Example Images

Testing Images GT	Compared Kinship Verification Models				Ensemble Verification*	Multi-class Net	Proposed Joint Learning
	F-D model	*F-S model*	*M-D model*	*M-S model*	**predicted label**	**predicted label**	**predicted label**
F-D 1	1	1	1	1	4 (M-S)	2 (F-S)	1 (F-D)
F-D 2	1	1	1	1	2 (F-S)	2 (F-S)	1 (F-D)
F-S 1	1	1	1	1	3 (M-D)	1 (F-D)	2 (F-S)
F-S 2	1	1	1	1	1 (F-D)	1 (F-D)	2 (F-S)
M-D 1	1	0	0	1	1 (F-D)	1 (F-D)	3 (M-D)
M-D 2	1	1	1	1	2 (F-S)	4 (M-S)	3 (M-D)
M-S 1	1	1	1	1	3 (M-D)	2 (F-S)	4 (M-S)
M-S 2	1	0	1	1	3 (M-D)	3 (M-D)	4 (M-S)
Non-relation 1	0	1	1	0	3 (M-D)	2 (F-S)	0 (Non-relation)
Non-relation 2	0	0	0	1	4 (M-S)	2 (F-S)	0 (Non-relation)

We selected several representative samples. Table. 1 represents the identification results of different approaches. The abbreviations of image pairs represent the kinship type of these pairs, *e.g.* F-D 1('fd_039_1.jpg', 'fd_039_2.jpg')

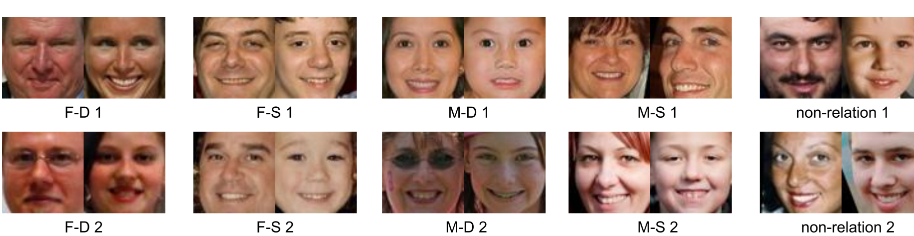


Fig. 1: Examples for Validation

represents this pair has a father-and-daughter relationship and non-relation 1('fs_188_1.jpg', 'fs_044_2.jpg') represents the pair does not have a kinship relation. The results show that when testing each trained verification model separately, models tend to get similar results (1 represent the positive output of a specific model). The final results show that the joint learning method yields a more precise result compared to other methods.

3 Details of Experiments

3.1 Experiments Results

The results of our experiments show that the joint learning method is reproducible, and it obtained the good performance on different image sets.

Table 2: Results of experiments on KinfaceWI (partial)

Training dataset	**Methods**	**KinfaceWI**									
		accuracy					F1				
		F-D	*F-S*	*M-D*	*M-S*	*mean*	*F-D*	*F-S*	*M-D*	*M-S*	*mean*
Independent Kin-type Dataset	Ensemble Verification* (time1)	0.6758	0.7538	0.6748	0.6457	0.6875	0.72	0.7772	0.6754	0.7064	0.7197
	Ensemble Verification* (time2)	0.5976	0.7413	0.693	0.581	0.6532	0.6409	0.7549	0.7414	0.6592	0.6991
	Ensemble Verification* (time3)	0.7017	0.7506	0.741	0.615	**0.7021**	0.6915	0.7472	0.7566	0.6648	**0.715**
Mixed-type Dataset	Multi-class Net (time1)	0.6306	0.6319	0.6154	0.5853	0.6158	0.554	0.5292	0.558	0.5158	0.5392
	Multi-class Net (time2)	0.6311	0.6634	0.6659	0.5687	0.6323	0.5709	0.6003	0.6347	0.537	0.5858
	Multi-class Net (time3)	0.6385	0.6443	0.645	0.5516	0.6198	0.5672	0.5637	0.6087	0.4897	0.5573
	Multi-class Net (time4)	0.6463	0.6797	0.665	0.577	**0.642**	0.6084	0.6563	0.6767	0.5766	**0.6295**
	Ensemble Verification	0.6425	0.6321	0.6382	0.577	0.6224	0.6639	0.6737	0.6735	0.6083	0.6548
	proposed Joint Learning** (time1)	0.635	0.6668	0.6305	0.5467	0.6198	0.6417	0.6573	0.6409	0.5711	0.6277
	proposed Joint Learning** (time2)	0.6534	0.6991	0.6539	0.5772	**0.6459**	0.6301	0.6952	0.6496	0.5816	**0.6391**
	proposed Joint Learning** (time3)	0.6348	0.6928	0.6419	0.56	0.6324	0.6183	0.6891	0.6571	0.5495	0.6285
	Joint Learning (time1)	0.6651	0.7503	0.7007	0.5815	0.6744	0.6352	0.7465	0.7145	0.6021	0.6746
	Joint Learning (time2)	0.6503	0.75	0.6847	0.577	0.6655	0.6191	0.7282	0.6852	0.583	0.6539
	Joint Learning (time3)	0.6947	0.7469	0.7004	0.6025	**0.6861**	0.684	0.7374	0.6902	0.6074	**0.6798**
	Joint Learning (time4)	0.6466	0.7535	0.6933	0.5899	0.6708	0.6074	0.7488	0.6491	0.6052	0.6526

Table. 2 and table. 3 show the results of different methods on the Independent Kin-type Image Set. From the results, we can conclude that:

- The *Joint Learning* method provides reproducible results.
- When trained on the same dataset (Mixed-type dataset), the *Joint Learning* method yields the best results.

Table 3: Results of experiments on KinfaceWII (partial)

Training dataset	Methods	KinfaceWII									
		accuracy					F1				
		F-D	*F-S*	*M-D*	*M-S*	*mean*	*F-D*	*F-S*	*M-D*	*M-S*	*mean*
Independent Kin-type Dataset	**Ensemble Verification*** (time1)	0.746	0.744	0.752	0.732	**0.7435**	0.7671	0.7589	0.769	0.7607	**0.7639**
	Ensemble Verification* (time2)	0.728	0.700	0.74	0.744	0.728	0.7523	0.7139	0.7549	0.7692	0.7476
Mixed-type Dataset	Multi-class Net (time1)	0.598	0.628	0.618	0.564	0.602	0.5728	0.6031	0.5827	0.4762	0.5587
	Multi-class Net (time2)	0.588	0.624	0.62	0.592	**0.606**	0.5629	0.600	0.6143	0.5062	**0.5709**
	Ensemble Verification	0.606	0.600	0.586	0.626	0.6045	0.6213	0.6439	0.6051	0.6399	0.6276
	proposed Joint Learning** (time1)	0.616	0.61	0.600	0.65	**0.619**	0.6396	0.6166	0.6061	0.6191	**0.6203**
	proposed Joint Learning** (time2)	0.634	0.61	0.600	0.61	0.6135	0.6444	0.6408	0.6286	0.6029	0.6292
	Joint Learning (time1)	0.69	0.76	0.714	0.714	0.7195	0.7045	0.7622	0.7196	0.7525	0.7347
	Joint Learning (time2)	0.662	0.71	0.718	0.704	0.6985	0.6953	0.7377	0.7484	0.7528	0.7336
	Joint Learning (time3)	0.700	0.744	0.718	0.728	**0.7225**	0.7189	0.7588	0.7275	0.7526	**0.7394**
	Joint Learning (time4)	0.698	0.732	0.73	0.736	**0.724**	0.7354	0.7439	0.746	0.7612	**0.7466**

We compute each approach multiple times and report the results of the comprehensive optimal models for comparison. As shown in Table. 2, since *Ensemble Verification* (time3)* and *Multi-class Net (time4)* provides good results on the Independent Kin-type Image Set(KinfaceWI dataset), the results of these two models are selected as the final results. The models of final results are: *Ensemble Verification*(time3), Ensemble Verification, Multi-class Net (time4), proposed Joint Learning** (time2)* and *Joint Learning (time3)*. The *Joint Learning (full)* is the same model as *Joint Learning (time3)* but predicts a combination of binary outputs and multi-class outputs.

Table 4: Selected Optimal Models of Each Appraoch

Ensemble Verification	**Multi-class Net**	**Joint Learning**
Ensemble Verification*(time3)	Multi-class Net(time4)	proposed Joint Learning**(time2)
Ensemble Verification		Joint Learning(time3)

The results of different methods on the Mixed-type Image Set and Real-scenario Image Set are listed in Table. 5 and Table 6 respectively. Since the F1 score balances the recall and precision rate, the F1 score is taken. Table. 5 shows the results of the different experiments on the Mixed-type Image Set for the kinship identification task. *Joint learning* methods outperform other methods and provides the best performance on both macro F1 and accuracy measurement, which is also reproducible.

Table 5: Macro F1 score and Accurac of experiments (partial)

Training Dataset	Methods	KinfaceWI		Methods	KinfaceWII	
		macro F1	*acc*		*macro F1*	*acc*
Independent Kin-type Dataset	Ensemble Verification* (time3)	0.3240	0.3723	Ensemble Verification* (time1)	0.2846	0.3319
Mixed-type Dataset	Multi-class Net (time4)	0.5291	0.5494	Multi-class Net (time2)	0.4861	0.5225
	Ensemble Verification	0.4837	0.4887	Ensemble Verification	0.4464	0.4564
	proposed Joint Learning** (time2)	0.5139	0.5467	proposed Joint Learning** (time1)	0.46114	0.485
	Joint Learning (time1)	0.5202	0.5646	Joint Learning (time1)	0.4877	0.5155
	Joint Learning (full)) (time1)	0.5246	0.5796	Joint Learning (full)) (time1)	0.5048	0.5645
	Joint Learning (time2)	0.5317	0.579	Joint Learning (time2)	0.47538	0.490
	Joint Learning (full) (time2)	0.5416	**0.6028**	Joint Learning (full) (time2)	0.49496	0.5469
	Joint Learning (time3)	0.5377	0.5898	Joint Learning (time3)	0.5003	0.5285
	Joint Learning (full) (time3)	0.5392	**0.6002**	Joint Learning (full) (time3)	**0.5143**	**0.5685**
	Joint Learning (time4)	**0.5440**	0.5861	Joint Learning (time4)	0.4794	0.5109
	Joint Learning (full) (time4)	0.5256	0.5880	Joint Learning (full) (time4)	0.5064	0.56

Table. 6 shows the results of different approaches to the Real-scenario image set. We mainly focus on the results of the F10 score since, in reality, the goal is obtain a high recall. Mean represents the average value of the F10 score of different kin-types and excludes the negative types (non-relationship). F10 all represents the average value of the F10 score of different kin-types and includes the negative types. The results show that Joint Learning obtains higher results compared to other approaches.

Table 6: F10 score of different approaches on real-scenario image Set

methods	KinfaceWI							KinfaceWII						
	F-D	*F-S*	*M-D*	*M-S*	*mean*	*f10_all*	*acc*	*F-D*	*F-S*	*M-D*	*M-S*	*mean*	*f10_all*	*acc*
Ensemble Verification* (time1)	0.1074	0.1596	0.0655	0.0843	0.1042	0.1550	0.3557	0.0520	0.0607	0.0708	0.0763	0.0650	0.1432	0.4533
Ensemble Verification* (time3/2)	0.0886	0.1179	0.1236	0.1003	0.1076	0.1830	0.4807	0.0469	0.0713	0.0726	0.0904	0.0703	0.1498	0.4647
Multi-class Net (time2/1)	0.1582	0.2920	0.3046	0.1645	0.2298	0.3258	0.7060	0.1380	0.1828	0.1758	0.0990	0.1489	0.2381	0.5920
Multi-class Net (time4/2)	0.1548	0.2951	0.3047	0.1539	0.2271	0.2947	0.5618	0.1468	0.1972	0.1853	0.1076	0.1592	0.2528	0.6240
Ensemble Verification	0.1508	0.2791	0.2740	0.1378	0.2104	0.2596	0.4537	0.1399	0.1681	0.1496	0.0900	0.1369	0.2075	0.4874
Joint Learning** (time2/1)	0.1510	0.2951	0.2899	0.1593	0.2238	0.2980	0.5916	0.1409	0.1743	0.1595	0.0974	0.1430	0.2297	0.5738
Joint Learning 3	0.167	0.3169	0.2986	0.1590	0.2354	0.3280	0.6949	0.1496	0.1930	0.2030	0.1158	0.1654	0.2561	0.6161
Joint Learning 3(full)	0.1762	0.3272	0.2935	0.1710	**0.2420**	0.3464	0.7606	0.1708	0.2111	0.2345	0.1226	**0.1847**	0.2937	0.7267
Joint Learning 4	0.1825	0.3064	0.3165	0.1604	0.2415	0.3322	0.6916	0.1531	0.1933	0.1944	0.1040	0.1612	0.2512	0.6084
Joint Learning 4(full)	0.1773	0.3028	0.2929	0.1552	0.2320	0.3372	0.7543	0.1584	0.2119	0.2459	0.1178	0.1835	0.2930	0.7279

3.2 Parameters Adjustment on *Joint Learning* Method

General Parameters The general training parameters for the Joint Learning Method are shown in Table. 7. The parameters of data augmentation, training parameters, and training methods affect the final performance of the model of the Joint Learning method. Considering the efficiency, we do not exhaust the possibility of all matches one by one. We believe there will be better parameters to be used that help improve the performance in the future.

Table 7: General Training Parameters for *Joint Learning* Method

Training Parameters								
Data Augmentations	*ColorJitter*	*RandomGrayscale*	*RandomHorizontalFlip*	*RandomPerspective*	*RandomResizedCrop*	*Normalize*	*RandomErasing*	*ToTensor*
	brightness=0.3, contrast=0.3, saturation=0.3, hue=0.3	p = 0.1	p = 0.4	distortion_scale = 0.1 p = 0.3	size=(64,64) scale=(0.9, 1.05) ratio=(0.97, 1.05) interpolation=2	mean= 0.485, 0.456, 0.406) srd= (0.229, 0.224, 0.225)	p=0.5	/
Training Parameters	*Imsize*	*Epoch Number*	*Training batch*	*Learning Rate*	*lr_decay*	*lr_milestones*	*weight_decay*	*Opimizer*
	(6,64,64)	∼200	64	0.0001	0.5	[80,150,350]	5e-3	Adam

Training Algorithm The algorithm of updating parameters while training can be described in Algorithm 1.

Algorithm 1: Parameter Updating Druing Training

initialzation;
lr of Optimizer during **Step 1** = 0.8* lr of Optimizer during **Step 2**
while *epoch* < *epoch numbers* **do**
  if *epoch in matched epoch_lambda_milestone((epoch_start, epochstop])* **then**
    choose matched $\lambda_i; i \in [1, 2, 3, 4, 5]$
  end
  During each epoch;
  step 1: parameter update based on binary outputs losses;
    $\lambda_i \mathcal{L}oss_{kvi}$ parameter update;
  step 2: parameter update based on multi loss;
    $\mathcal{L}oss = \sum_{i=1}^{4} \lambda_i \mathcal{L}oss_{kvi} + \lambda_5 \mathcal{L}oss_{kI}$,parameter update;
end

Detail Parameters of each *Joint Learning* Experiment The parameters of *Joint Learning* Experiments are listed in Table. 8. Different parameters can influence the performance of experiments. There will be better parameters to be used to help *Joint Learning* method improve performance in the future.

Table 8: Training Parameters of each *Joint Learning* Experiment

Training Dataset	Training Times	Parameters				
		Weighted Cross Entropy	***Epoch_num***	***lr_milestones***	***Epoch_lambdas***	
					epoch_milestone	*matched_lambdas*
KinfaceWI	Joint Learning (time 1)	[0.1,2,2,2,2]	150	[80, 130]	[1,150]	[1,1,1,1,10]
	Joint Learning (time 2)	[0.1,2,2,2,2]	200	[80,150]	[1,130)	[1,1,1,1,10]
					[130,150)	[0.3,0.3,0.3,0.3,10]
					[150,200]	[0,0,0,0,10]
	Joint Learning (time 3)	[0.1,2,2,2,2]	220	[80,150]	[1,220]	[1,1,1,1,10]
	Joint Learning (time 4)	[0.1,2,2,2,2]	220	[80,150]	[1,220]	[1,1,1,1,10]
KinfaceWII	Joint Learning (time 1)	[0.1,2.2,2.2,2.2,2.2]	300	[80,150,280]	[1,130)	[1,1,1,1,10]
					[130,180)	[0.3,0.3,0.3,0.3,10]
					[150,300]	[0,0,0,0,10]
	Joint Learning (time 2)	[0.1,2,2,2,2]	220	[80,150]	[1,220]	[1,1,1,1,10]
	Joint Learning (time 3)	[0.1,2.2,2.2,2.2,2.2]	350	[80,150,280]	[1,130)	1,1,1,1,10]
					[130,180)	[0.3,0.3,0.3,0.3,10]
					[150,350]	[0,0,0,0,10]
	Joint Learning (time 4)	[0.1,2.1,2.1,2.1,2.1]	350	[80,150,280]	[1,130)	1,1,1,1,10]
					[130,180)	[0.3,0.3,0.3,0.3,10]
					[150,350]	[0,0,0,0,10]

3.3 Reimplementation of the Backbone (Attention Network)

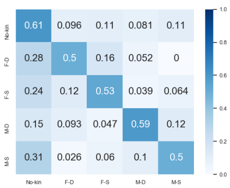
Since there is no public code available for the kinship verification model, we chose a recently proposed model, Attention Network, and asked the authors for the script but get no response. We reimplemented the structure of the attention network from scratch. The results of our reimplemented model and reported results on KinfaceWI and KinfaceWII are listed in Table. 9. Considering that all our experiments are conducted using the same backbone (Attention Network), the reimplementing process does not affect the comparison of our experiments. Since the *Joint Learning* method is not restricted to a specific backbone, it is promising that a better backbone can improve the results of the *Joint Learning* method.

Table 9: Results of attention network on each specific kin-type dataset of KinfaceWI and KinfaceWII based on kinship verification method.

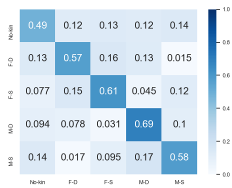
Dataset	attention network	F-D	F-S	M-D	M-S	Mean
KinfaceWI	reported	0.701	0.689	0.827	0.729	0.739
	re-implement time 1	0.676	0.754	0.675	0.646	0.687
	re-implement time 2	0.598	0.741	0.693	0.581	0.653
	re-implement time 3	0.702	0.751	0.741	0.615	0.702
KinfaceWII	reported	0.814	0.866	0.902	0.888	0.868
	re-implement time 1	0.746	0.744	0.752	0.732	0.744
	re-implement time 2	0.728	0.700	0.74	0.744	0.728

4 Confusion Matrix of different experiments based on Mixed-type Iamge Set of KinfaceWI

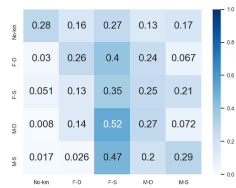
Fig. 2 and Fig. 3 depict the confusion matrix of different approaches on the Mixed-type Image Set for kinship identification. The results show that the *Ensemble verification** method can cause confusion. Compared to it, the *Joint Learning* method alleviate this problem.



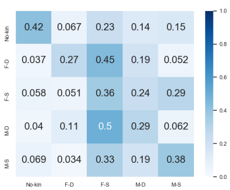
(a) Multi-class Net



(b) Multi-class Net

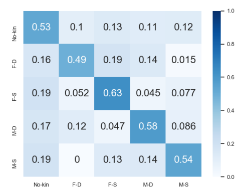


(c) Ensemble Verification*

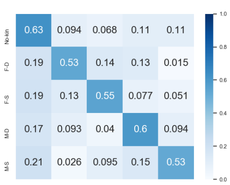


(d) Ensemble Verification*

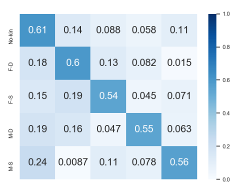
(e) Ensemble Verification



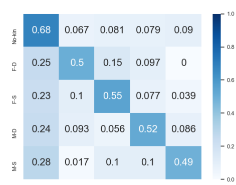
(f) Joint Learning**



(g) Joint Learning

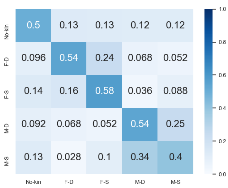


(h) Joint Learning

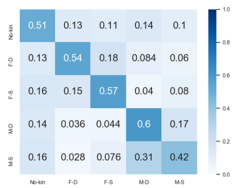


(i) Joint Learning full

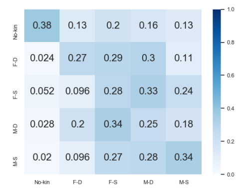
Fig. 2: Confusion Matrix of different experiments on Mixed-type Image Set based on KinfaceWI dataset.



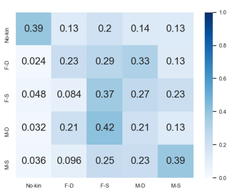
(a) Multi-class Net



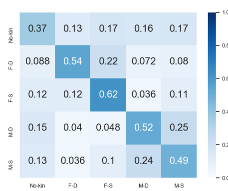
(b) Multi-class Net



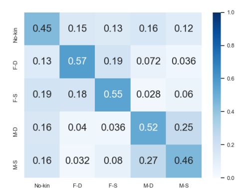
(c) Ensemble Verification*



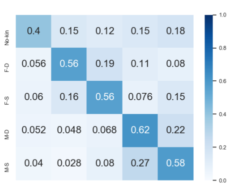
(d) Ensemble Verification*



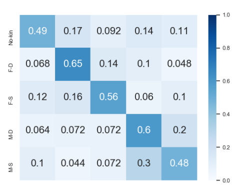
(e) Ensemble Verification



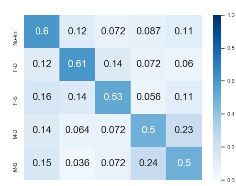
(f) Joint Learning**



(g) Joint Learning



(h) Joint Learning



(i) Joint Learning full

Fig. 3: Confusion Matrix of different experiments on Mixed-type Image Set based on KinfaceWII dataset.