

The 5th Recognizing Families in the Wild Data Challenge: Predicting Kinship from Faces

Joseph P. Robinson^{1,2}, Can Qin², Ming Shao³, Matthew A. Turk⁴,
Rama Chellappa⁵, and Yun Fu²

¹Vicarious Surgical ²Northeastern University ³UMass Dartmouth

⁴Toyota Technological Institute at Chicago (TTIC) ⁵Johns Hopkins University

Abstract—Recognizing Families In the Wild (RFIW), held as a data challenge in conjunction with the 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG), is a large-scale, multi-track visual kinship recognition evaluation. For the fifth edition of RFIW, we continue to attract scholars, bring together professionals, publish new work, and discuss prospects. In this paper, we summarize submissions for the three tasks of this year’s RFIW: specifically, we review the results for kinship verification, tri-subject verification, and family member search and retrieval. We look at the RFIW problem, share current efforts, and make recommendations for promising future directions.

I. INTRODUCTION

Automatic kinship recognition could be used for various uses, such as forensic analysis, automated photos app management, historical genealogy, multimedia analysis, missing kids and human trafficking tragedies, and immigration and border patrol concerns. Nonetheless, the challenges in such face-based tasks (*i.e.*, fine-grained classification in unconstrained settings) are only magnified in kin-based problem sets, as the data shows a high degree of variability in pose, illumination, background, and clarity, and soft biometric target labels, which only worsens the challenges with directional relationships consideration. As a result, the practical benefits of improving kinship-based technologies are counterbalanced by the difficulties posed by the problem of automated kinship comprehension. The *Recognizing Families In the Wild* (RFIW) challenge series was born out of this need: a large-scale data challenge with various tasks to advance kinship detection technology. RFIW is an open forum for researchers to present and discuss the state-of-the-art (SOTA).

In conjunction with the 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG), participants of the fifth RFIW¹ continue to push SOTA in each of the supported tasks (Fig. 1). In parallel to this effort is the focus on improving and extending the *Families In the Wild* (FIW) dataset [32], [34], [41]—a large-scale, multi-task image set for kinship recognition.² The size and scope of FIW have proven to match the demand of modern-day, data-hungry deep networks [4], [6], [9], [15], [44]. Nonetheless, there is room to grow in the quality, quantity, and even organization of the existing labels and evaluation protocols.

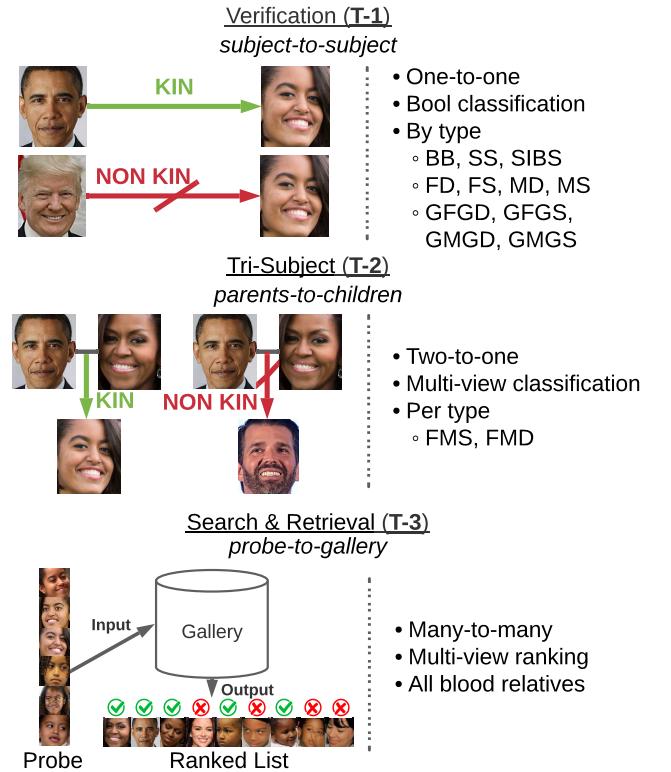


Fig. 1: An illustration of the three tasks supported in RFIW.

We propose modern technologies to work with the data, along with plans to improve the scopes of the experiments (*i.e.*, more realistic settings).

Specifically, the FIW dataset introduced imagery scraped *in the wild*, with unconstrained family photos from the web. Although it is the **largest** and **most comprehensive** dataset of its kind, there still exist several concerns. For starters, the diversity of the data: the near 1,000 families are vast compared to other datasets, but not compared to the real world. Furthermore, still-faces in images are not the only target in multimedia that can infer kinship, the facial dynamics in videos and speech signals were shown to complement still-faces (*i.e.*, *FIW in Multimedia* (FIW MM)). Hence, future RFIW will support the multimedia content. Lastly, the experimental settings should be tweaked and further bridge the gap between research and reality.

Like in [36], the 2021 RFIW comprised three tasks de-

¹RFIW2021 webpage, <https://medium.com/recognize-families>.

²FIW project page, <https://web.northeastern.edu/smilelab/fiw/>.

TABLE I: Counts for T-1: the number of unique pairs (**P**), families (**F**), and faces (**S**) [36].

	1 st Generation			2 nd Generation			3 rd Generation					
	BB	SS	SIBS	FD	FS	MD	MS	GFGD	GFGS	GMGD	GMGS	Total
train	P	991	1,029	1,588	712	721	736	716	136	124	116	114
	F	303	304	286	401	404	399	402	81	73	71	66
	S	39,608	27,844	35,337	30,746	46,583	29,778	46,969	2,003	2,097	1,741	1,834
val	P	433	433	206	220	261	200	234	53	48	56	42
	F	74	57	90	134	135	124	130	32	29	36	27
	S	8,340	5,982	21,204	7,575	9,399	8,441	7,587	762	879	714	701
test	P	469	469	217	202	257	230	237	40	31	36	33
	F	149	150	89	126	133	136	132	22	21	20	22
	S	3,459	2,956	967	3,019	3,273	3,184	2,660	121	96	71	84
												39,743

picted in Fig. 1: (**T-1**) Kinship verification, (**T-2**) Tri-subject verification, and (**T-3**) Family member search & retrieval.

The rest of the paper is organized as follows. First, we review the related (Section II). A brief review of the data and task protocols follows (Section III). We then introduce the top methods of the challenge (Section IV). Lastly, we end with a discussion (Section V) and a conclusion (Section VI).

II. RELATED WORKS

Kinship understanding is a critical vision problem first published in the 2010 ICIP [8]. As a human-centered visual learning problem, earlier works focused on low-level features of facial imagery. To combat challenges inherent in visual kinship verification (*e.g.*, variations from age), researchers incorporated mainstream learning approaches like transfer learning [45], [46], and metric learning [18], [41]. More recently, advances in the evaluation protocols paved the way to pragmatic problem formulations (*e.g.*, tri-subject kinship verification [24] and family recognition [32], [34]). Furthermore, as a part of this RFIW data challenge series, with its debut in last year’s edition [36], search and retrieval of missing family problems mimic a variety of practical use-cases.³ Thus, they further closed the gap from research to reality. Robinson *et al.* analyze the evolution of the visual kinship problem domain over time, along with the various paradigms, SOTA, and promising future directions in a recent survey [31] and dissertation [25].

During the past decade, deep learning has been prominent across face-based vision systems [39]. Initial visual kinship benchmarks (*e.g.*, KinWild [18] and Family101 [7]) had a profound impact in organizing and promoting the problem. However, such minimal data were insufficient to match the capacity needed to train deep models, with few exceptions (*e.g.*, in [50], smaller, parts-based models gained an incremental boost in performance compared with the pre-existing low-level methods). Furthermore, these earlier datasets mostly used faces from the same photos, such as color features [17] and then with the *same photo* detectors [3] claimed SOTA, highlighting problems with constructing a verification set using face samples from the same photo.

The shortage of data quality and labels, and, hence, the absence of proper data distribution of the faces of families,

motivated the release of the large-scale dataset FIW [33]. Ever since, FIW has supported various deep learning approaches [42], [44], with some generative models to predict the appearances of family members [9], [22]. Even other multimodal views (*i.e.*, FIW MM [26]), and in tutorials at top-tier conferences (*i.e.*, ACM MM [30] and CVPR⁴).

As a series of workshops and data challenges based on FIW, RFIW has been held in different venues over the past four years [35], [36]. Plus, FIW premiered in a Kaggle competition that attracted over 500 teams to make submissions.⁵ As part of the FG, we hosted the fifth edition of RFIW (*i.e.*, 2021 RFIW). Hence, there is a continued effort to keep updating and promoting the FIW dataset to push SOTA and inspire researchers in the years to come.

III. TASK EVALUATIONS, PROTOCOLS, BENCHMARKS

We briefly introduce each task and refer readers to our earlier white paper for more details [36]. Also, see Fig. 1 for a visual depiction of each.

Historically, most focus has been on **T-1** [5], [15], [41], [44]. Then, we introduced **T-2** and **T-3** in 2020 [36], for which several participants engaged [13], [14], [21], [38], [48], [49], [53]. The three data splits are formed at the family level to enable multi-task solutions, along with the earlier need to keep the ground truth for a subset of the families from the public (*i.e.*, for blind testing). In other words, the three sets (*i.e.*, *train*, *val*, and *test*) have the same families across all tasks. Specifically, 60% of the families make up the *train* set, while the remaining 40% was split between *val* and *test*. Hence, the three sets are disjoint in family and identity, which remain consistent across the different tasks.

Since the ground truth was released after the last RFIW, it was fair to supply all data and labels. Teams were asked to only process the *test* set when generating submissions, and any attempt to analyze or understand the *test* pairs were prohibited. Also, outputs were scored on the server, for which we received and scored all submissions via Codalab (*i.e.*, **T-1**⁶, **T-2**⁷, and **T-3**⁸).

⁴https://web.northeastern.edu/smilelab/cvpr19_tutorial/

⁵<https://www.kaggle.com/c/recognizing-faces-in-the-wild>

⁶<https://competitions.codalab.org/competitions/21843>

⁷<https://competitions.codalab.org/competitions/22117>

⁸<https://competitions.codalab.org/competitions/22152>

³<https://towardsdatascience.com/b5651999385e>

TABLE II: Averaged verification accuracy scores for **T-1**.

Team	1 st Generation			2 nd Generation			3 rd Generation				Average	
	BB	SS	SIBS	FD	FS	MD	MS	GFGD	GFGS	GMGD	GMGS	
Baseline [36]	0.61	0.66	0.69	0.62	0.66	0.71	0.73	0.68	0.57	0.64	0.50	0.64
TeamCNU [51]	0.82	0.84	0.80	0.76	0.82	0.75	0.77	0.76	0.71	0.75	0.59	0.80
vuvko [38]	0.75	0.81	0.78	0.74	0.78	0.69	0.76	0.60	0.80	0.80	0.77	0.78
nc2893 [12]	0.76	0.78	0.75	0.74	0.70	0.67	0.70	0.59	0.79	0.79	0.75	0.77
jh3450 [12]	0.76	0.78	0.75	0.74	0.70	0.67	0.70	0.59	0.79	0.79	0.75	0.77
paw2140 [12]	0.75	0.78	0.76	0.74	0.68	0.69	0.72	0.59	0.78	0.79	0.75	0.77
DeepBlueAI [53]	0.74	0.81	0.75	0.74	0.72	0.73	0.67	0.68	0.77	0.77	0.75	0.76
ustc-nelslip [47], [48]	0.76	0.82	0.75	0.75	0.79	0.69	0.76	0.67	0.75	0.74	0.72	0.76
stefhoer [13]	0.77	0.80	0.77	0.78	0.70	0.73	0.64	0.60	0.66	0.65	0.76	0.74

 TABLE III: Verification accuracy scores for **T-2**.

Team	FMS	FMD	Avg.
Baseline [36]	0.68	0.68	0.68
TeamCNU [51]	0.86	0.82	0.84
stefhoer [13]	0.74	0.72	0.73
DeepBlueAI [53]	0.77	0.76	0.77
ustc-nelslip [47], [48]	0.80	0.78	0.79

All faces were encoded via Sphereface Convolutional Neural Network (CNN) [16] (*i.e.*, 512 D), with the pre-processing and training from the original work.⁹ Cosine similarity determined the closeness of pairing faces by comparing features p_1 and p_2 , which is defined as $\text{CS}(p_1, p_2) = \frac{p_1 \cdot p_2}{\|p_1\| \cdot \|p_2\|}$ [20].

A. Kinship Verification (**T-1**)

To verify kinship is predicting whether a pair of individuals are blood relatives. In computer vision, we compare faces to classify the pairs as KIN or NON-KIN as true or false, respectively. This one-to-one view of kinship recognition typically assumes prior knowledge in the relationship type [30]. Hence, relationship types are evaluated independently. With the introduction of FIW, the number of face pairs and relationship types for kinship verification (*i.e.*, **T-1**) has significantly increased. Three sets of the data (*i.e.*, *train*, *val*, and *test*) are partitioned for RFIW (Table I). The test set had an equal number of positive and negative pairs, and no family (and, hence, subject identity) overlaps between sets. As of 2020, the challenge began to support grandparent-grandchild types, *i.e.*, grandfather-granddaughter (GFGD), grandfather-grandson (GFGS), grandmother-granddaughter (GMGD), grandmother-grandson (GMGS). Due to insufficient counts across folds, along with an incredible bias skewed away from the few families that make up pairs across three generations, the great grandparent-great grandchild pairs of FIW are omitted from **T-1** of RFIW.

Verification accuracy is used to evaluate the performance. Specifically, $\text{Accuracy}_j = \frac{\text{TP}_j + \text{TN}_j}{N_j}$, where $j^{th} \in \{\text{all 11 relationship types}\}$. Then, the overall accuracy is the weighted sum. The threshold for positive and negative pairs

was found by the value that maximizes the accuracy on the *val* set. Results are listed in Table II.

B. Tri-Subject Verification (**T-2**)

Tri-Subject Verification focuses on a different view of kinship verification—the goal is to decide if a child is related to a pair of parents. First introduced in [24], it makes a more realistic assumption, as knowing one parent often means the other potential parent(s) can be easily inferred.

Triplet pairs consist of Father (F) / Mother (M) - Child (C) (FMC) pairs, where the child C could be either a Son (S) or a Daughter (D) (*i.e.*, triplet pairs are FMS and FMD).

Triplets were formed by first matching each mother-father pair with their biological children to form the list of positives. Then, negative triplets were generated by shuffling



Fig. 2: Sample family with mixed-race parents from **T-2** test set. Top row: parents of the three daughters (bottom row). Our baseline incorrectly labeled all triplets (*i.e.*, parent-child pairs) with the daughter shown left-most, while correctly classifying all pairs with the other two daughters (shown in the middle and right columns): the former scored less when compared to the mother, while the latter two daughters scored higher (*i.e.*, more similar) to their mother such that it outweighed the low similarity compared with the father (*i.e.*, all three daughters score low with their father).

⁹<https://github.com/wyliu/sphereface>

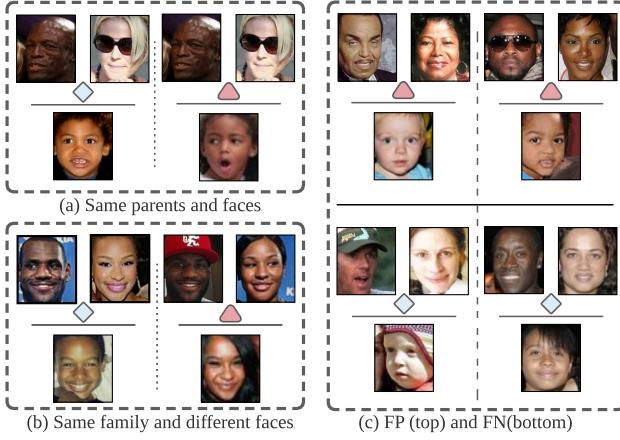


Fig. 3: Triplets of **T-2** incorrectly predicted, with the parents above the child separated by a shape standing for the type of error, where false negatives (FN) and false positives (FP) are marked by \diamond and \triangle , respectively. Exemplified are both error types for the same parent face pair, with wrong classifications marking the true son as NON-KIN and a random daughter as KIN (a). Besides the mother’s images with sunglasses occlude the periocular region, other challenges seem to be in the mixed race, for the FP and FN samples shown, although on the wrong side of the decision boundary, are not far from the other (*i.e.*, similar scores). The next triplet pairs incorrectly classified are of the same parents’ pairs (*i.e.*, different faces). Again, the son is falsely classified as NON-KIN, while a random daughter is an FN (b). The last set of sample triplets incorrectly classified show challenges posed by challenges in age (*i.e.*, FN triplet with a baby face), while the others are of mixed-race parents (c).

the children in the positive list such that pairs of parents remained constant to yield the same number of negatives. Note that the number of possible negatives is far more than positives, so a pair of faces of a parent pair was used once and only once to produce a balanced list. Again, no family or subject identity overlaps between sets: the same families make up the *train*, *val*, and *test* sets across the tasks.

Verification accuracy WAS first calculated per triplet-pair type (*i.e.*, FMD and FMS), and then averaged via the weighted sum. A score was assigned to each triplet (F_i, M_i, C_i) in the *val* and *test* sets using the formula $\text{score}_i = \text{mean}(\cos(F_i, C_i), \cos(M_i, C_i))$, where F_i , M_i , and C_i are the feature vectors of the i -th triplet. Scores were compared to a threshold γ to infer a label (*i.e.*, KIN if the score surpasses the threshold; else, NON-KIN). The threshold was found experimentally on the *val* set. The threshold was applied to the *test* (Table III).

C. Search and Retrieval (**T-3**)

As a search cue, kinship information can improve conventional FR search systems as prior knowledge for mining social or ancestral relationships in industries like *Ancestry.com*.

TABLE IV: Performance ratings for **T-3**.

Team	Average	mAP	Rank@5
ustc-nelslip [47], [48]	0.35	0.15	0.54
Baseline [36]	0.28	0.11	0.45
TeamCNU [51]	0.40	0.21	0.60
HCMUS notweeb [21]	0.17	0.07	0.28
DeepBlueAI [53]	0.19	0.06	0.32
vuvko [38]	0.39	0.18	0.60

However, the task is most related to missing persons. Thus, we pose **T-3** as a set-based paradigm. For this, we imitate template-based evaluations on the *probe* side but with a *gallery* of faces [43]. Furthermore, the goal is to find relatives of search subjects (*i.e.*, subject-level *probes*) in a search pool (*i.e.*, face-level *gallery*).

The protocol of **T-3** could be used to find parents and other relatives of unknown, missing children. The *gallery* has 31,787 facial images from 190 families in the *test* set. The inputs are media for a subject (*i.e.*, *probes*), and outputs are ranked lists of all faces in the *gallery*. The number of relatives varies for each subject, ranging anywhere from 1 to 20+. Furthermore, probes have one-to-many samples—the means of fusing feature sets (*i.e.*, a *probe*’s media) is an open research question [52]. This *many-to-many* task is currently set up in closed form (*i.e.*, every *probe* has relative(s) in the *gallery*).

For each of the N test *probes* (*i.e.*, family f), the average precision (AP) is calculated:

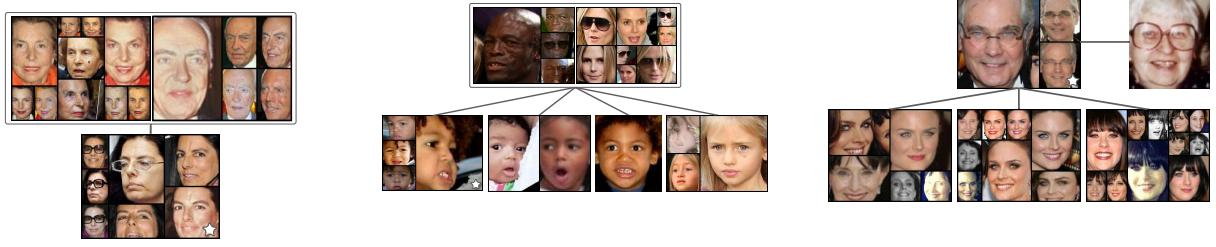
$$\text{AP}(f) = \frac{1}{P_F} \sum_{tp=1}^{P_F} \text{Prec}(tp) = \frac{1}{P_F} \sum_{tp=1}^{P_F} \frac{tp}{\text{rank}(tp)},$$

where P_F is the number of true-positives (TPs) for the f -th family. The average AP (*i.e.*, mean average precision (mAP)) is then reported: $\text{mAP} = \frac{1}{N} \sum_{f=1}^N \text{AP}(f)$. Finally, Rank@5 is reported: the average number of *probes* returned at least one TP in the top five gallery faces. The baseline, along with others on the scoreboard, is shown in Table IV.

IV. SUMMARY OF SUBMISSIONS

New solutions published as part of the 2021 RFIW challenge in the FG proceedings are introduced. Readers are referred to the paper references for additional details.

a) *TeamCNU*: proposed a contrastive learning framework to tackle and lead all three tasks [51]. The core idea of their solution is that self-supervised contrastive learning can help to learn powerful representations for different downstream tasks such as three tracks in RFIW. Following the 2020-RFIW’s winning team Vuvko [38], ArcFace (*i.e.*, a pre-trained ResNet-101) was used to encode raw faces, and then a multi-layer perceptron (MLP) was attached to obtain the low-dimensional feature pair for computing the contrastive loss [2]. To output the kinship verification results for all three tasks, it only needs to remove the MLP layers and take the mid-level features extracted by the Siamese backbone to compute the similarity score. Then, a predefined threshold is naturally applied to select the positive pairs. It



(a) Top scoring.

(b) Runner-up (*i.e.*, 2nd scoring).

(c) Lowest scoring.

Fig. 4: The top-ranking (a), runner-up (b), and last-place (c) families, on average, in **T-3**. The probe (*i.e.*, search query) displays a white star in the lower-left corners of the respective montage. The top two families (*i.e.*, (a) and (b)) are queries of children with both parents in the gallery, while the family with the lowest accuracy was a subject with children and a sibling (*i.e.*, a sister) present. Note that connections indicate blood relatives, where (a) and (b) the parents that share children (*i.e.*, the only blood connection shared is in the children), while the topmost of (c) does not include the mother of the children, for she is not a blood relative of the query (*i.e.*, who is a father and brother in the types of relationships in focus here).

surprisingly outperforms all the former SOTA methods on most of the tracks in the RFIW challenge.

b) paw2140: used ensemble learning on both data and networks levels [12]. Specifically, this team applied data augmentation to obtain more representative input data, like random rotation in angle, minor crops, horizontal flips, and channel-wise transformation. Moreover, the authors used multiple duplicate data samples with different augmentations as the ensemble inputs for testing. The network ensemble has employed multiple structures such as ResNet50 [10], FaceNet [37], VGGFace [23], and SENet50 [11] as the backbones where the features from different backbones fused in multiple ways. The Hadamard product, squared, and absolute value difference of the pairs of features are concatenated for the high-level similarity quantization. Moreover, there is the ensemble across different splits of the training data with k folds. For each instance, the final $4*k$ ensemble member networks, trained with k different splits and four Siamese networks for each split, are applied for the prediction. To further boost the performance, they take the program synthesis based on the OpenAI’s Codex [1] to automatically generate variants of networks for the ensemble.

c) ustc-nelslip: used a Siamese neural network designed for all three tasks [47]. For the *one-vs-one* kinship verification (**T-1**), the team takes a two-branch deep Siamese neural network with the enhanced feature fusion (*i.e.*, concatenation of squared difference, the difference of squared features, and dot product). To address the *two-vs-one* verification (**T-2**), they proposed a pair of deep Siamese neural networks, each included four branches. There is only one branch network for the child data shared by the two Siamese networks. Moreover, the other two branches are for mother and father images, respectively. Team ustc-nelslip introduced the feature fusion similarity and cosine similarity for measurement to obtain the rank of similar images of a query (**T-3**). Finally, the obtained similarity scores of mother-children and father-children are weighted to generate the parent-child similarity score for kinship verification.

V. DISCUSSION

Focusing on **T-2**, by not only depicting a shortcoming of the existing methods but also the potential for more advanced systems compared to models of **T-1**, let us use the sample family in Fig. 2 to examine the common challenge of mixed-race pairs. As described, the *test* set is split by family, which stays constant for all tasks. Hence, the pair-list of *T-2* is a subset of *T-1* (*i.e.*, *parent-child* pairs with children with both parents present). For instance, if son $S \in FS$ & $MS- > FMS$. Hence, the family shown is of type *FMD*, where all face pairs of *FD* were falsely predicted as *NON-KIN*, while all *MD* pairs were correctly inferred. As for the *FMD* triplets, all pairs with one of their daughters were incorrect, while all others involving the other two daughters were correct. Hence, the information is there for the correct decision on all triplets, unlike with pairs of **T-1** (*i.e.*, all *FD* were incorrect), but only two of three daughters look enough like the mother to offset the dissimilarity between the children and father. A robust fusion scheme is a promising direction that could improve such cases (*i.e.*, mixed-race parents), especially considering how more common that is in modern-day families.

Additionally, as depicted in Fig. 3, we visually explore the common errors of **T-2**. For true and false pairs marked incorrectly classified, hard samples often include mixed-race parents. On average, the SOA *parent-child* verification system is best with both parents present (*i.e.*, $mAP(FD, FS, MS, MD) < mAP(FMD, FMS)$). Hence, there is great promise in improving the fusion. We also can see, as usual, that a hard sample is of a baby’s face (Fig. ??).

To explore easy and challenging cases of **T-3**, we took the average across the baseline and two submissions of this RFIW. The top two queries in mAP are of children, for which both parents are present in the gallery. The last-place query is a subject with no parents present in the query (Fig. 4). It is interesting to note that the parents of the second family are inter-racial: as expected, children tend to inherit features

```

import fiftyone as fo
import fiftyone.zoo as foz

"""
download and prepare test data and labels
"""

dataset = \
    foz.load_zoo_dataset("fiw", split="test")

"""
launch desktop app to explore test faces
"""

session = fo.launch_app(dataset)

```

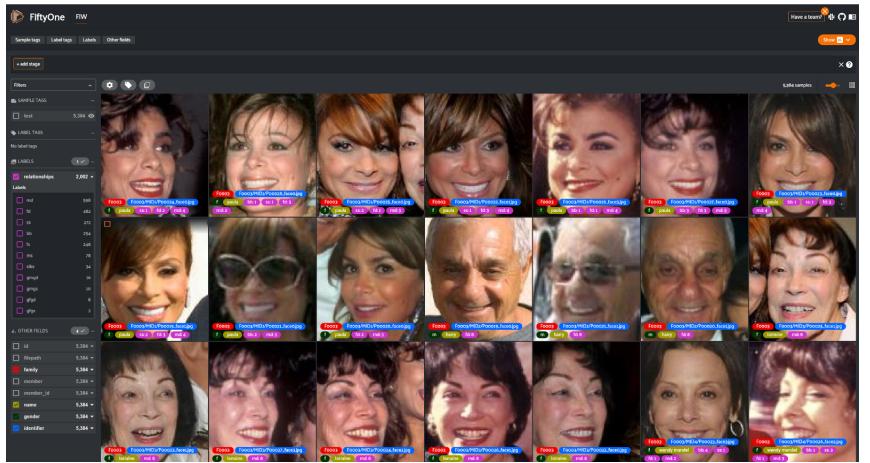


Fig. 5: FIW is now available as a part of fiftyone’s datazoo. Fiftyone is an open-source tool for dataset curating and model analysis via an easy-to-use Python interface and application (<https://fiftyone.ai>). We aim for simpler access, built-in data exploration features, and to inspire community contributions through engaging. Hence, the small code snippet (left) launches the application, as shown (right). The left-side of the application window are filters, allowing the user to view different relationship types easily (e.g., BB, SS, FD, MD) and filter families. Below the filters are check-boxes corresponding to label displays (e.g., gender, image name, family label) - check to toggle metadata shown as colored labels at the bottom of each face. We chose fiftyone for various reasons, with it being extendable an essential spec: different labels and metadata are easy to add, manage, and extend. This incentive will support all tasks and, eventually, the multimedia variant FIW MM [26]. FIW sets (i.e., *train*, *val*, *test*) are obtainable independently or all together. For instance, we expect analysis capabilities to be enhanced and the tracking of data bugs for an improved version rollout. Best viewed electronically.

from both parents, on average.

A challenge in facial recognition problems is age: it is typically more difficult when compared to a baby’s face or an elder. Age variations are especially challenging in face-based visual kinship recognition and to the degree earlier works base the underlying motivation (*i.e.*, the story) about age. For instance, Wang *et al.* viewed the problem as imbalanced data solved by augmenting data by transforming faces at age intervals [40]. Note that the faces of the last place query are all at an older age (Fig. 4c). All blood relatives present in the gallery are the opposite sex (*i.e.*, one sister and three daughters), while both the top-scoring (Fig. 4a) and the runner-up (Fig. 4b) have at least half of their known (*i.e.*, present) relatives of the same sex with most of the faces.

FIW was added to fiftyone’s datazoo [19] - accessing and exploring the data made efficient. Additionally, we plan to release evaluations per Voxel51’s Python API (Fig. 5). FIW is more accessible and extendable as we work to incorporate improved protocols, data quality, and MM in future RFIWs - whether raw faces, faces with predicted landmark (*i.e.*, generated via [27]), aligned faces, or corresponding face encodings, a variety of versioned data will be easy to access). We also plan to incorporate our *Balanced Faces In the Wild* (BFW) [28], [29] similarly, which will aid much-needed studies of bias in the FIW dataset.

VI. CONCLUSION

Another year of *Recognizing Families In the Wild* (RFIW) in conjunction with the 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG),

state-of-the-art (SOTA) in kin-based vision models continue to improve across all three tasks currently supported by the *Families In the Wild* (FIW) dataset. *TeamCNU* topped the score charts via a contrastive learning framework geared to learn better representation for comparing faces in all three kinship recognition tasks. We are improving the existing data and settings by working to make FIW more accessible while also bringing in the multimedia data in the benchmarks. Baseline code at github.com/visionjo/pykinship.

REFERENCES

- [1] M. Chen, J. Tworek, H. Jun, Q. Yuan, H. Ponde, J. Kaplan, H. Edwards, Y. Burda, N. Joseph, G. Brockman, et al. Evaluating large language models trained on code. *CoRR arXiv:2107.03374*, 2021.
- [2] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.
- [3] M. Dawson, A. Zisserman, and C. Nelläker. From same photo: Cheating on visual kinship challenges. In *Asian Conference on Computer Vision (ACCV)*, 2018.
- [4] Q. Duan and L. Zhang. Advnet: Adversarial contrastive residual net for 1 million kinship recognition. In *ACM MM RFIW Workshop*, 2017.
- [5] Q. Duan and L. Zhang. Advnet: Adversarial contrastive residual net for 1 million kinship recognition. In *ACM MM RFIW Workshop*, 2017.
- [6] I. Ö. Ertugrul and H. Dibeklioglu. What will your future child look like? modeling and synthesis of hereditary patterns of facial dynamics. In *Conference on FG*, 2017.
- [7] R. Fang, A. Gallagher, T. Chen, and A. Loui. Kinship classification by modeling facial feature heredity. In *Inter. Conference on Image Processing (ICIP)*, 2013.
- [8] R. Fang, K. D. Tang, N. Snavely, and T. Chen. Towards computational models of kinship verification. In *Inter. Conference on Image Processing (ICIP)*. IEEE, 2010.
- [9] P. Gao, S. Xia, J. Robinson, J. Zhang, C. Xia, M. Shao, and Y. Fu. What will your child look like? DNA-Net: Age and gender aware kin face synthesizer. *arXiv:1911.07014*, 2019.

- [10] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [11] J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.
- [12] J. Huang, M. Strome, I. Jenkins, P. Williams, B. Feng, Y. Wang, R. Wang, V. Bagri, N. Cheng, and I. Drori. Top 3 in FG 2021 families in the wild kinship verification challenge. In *Conference on Automatic Face and Gesture Recognition (FG)*, 2021.
- [13] S. Hörmann, M. Knoche, and G. Rigoll. A multi-task comparator framework for kinship verification. In *FG Challenge on RFIW*, 2020.
- [14] O. Laiadi, A. Oumame, A. Benakcha, A. Taleb-Ahmed, and A. Hadid. Multi-view deep features for robust facial kinship verification. In *FG Challenge on RFIW*, 2020.
- [15] Y. Li, J. Zeng, J. Zhang, A. Dai, M. Kan, S. Shan, and X. Chen. Kinnet: Fine-to-coarse deep metric learning for kinship verification. In *ACM MM RFIW Workshop*, 2017.
- [16] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song. Sphereface: Deep hypersphere embedding for face recognition. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [17] M. B. López, E. Boutellaa, and A. Hadid. Comments on the “kinship face in the wild” data sets. *Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, 2016.
- [18] J. Lu, X. Zhou, Y.-P. Tan, Y. Shang, and J. Zhou. Neighborhood repulsed metric learning for kinship verification. *Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, 2014.
- [19] B. E. Moore and J. J. Corso. Fiftyone. *Github. Note: https://github.com/voxel51/fiftyone*, 2020.
- [20] H. V. Nguyen and L. Bai. Cosine similarity metric learning for face verification. In *Asian Conference on Computer Vision (ACCV)*, 2010.
- [21] T.-D. H. Nguyen, H.-N. Nguyen, and H. Dao. Recognizing families through images. In *FG Challenge on RFIW*, 2020.
- [22] S. Ozkan and A. Ozkan. Kinshipgan: Synthesizing of kinship faces from family photos by regularizing a deep face network. In *Inter. Conference on Image Processing (ICIP)*. IEEE, 2018.
- [23] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *British Machine Vision Conference (BMVC)*, 2015.
- [24] X. Qin, X. Tan, and S. Chen. Tri-subject kinship verification: Understanding the core of a family. *arXiv:1501.02555*, 2015.
- [25] J. P. Robinson. *Automatic Face Understanding: Recognizing Families in Photos*. PhD thesis, Northeastern University, 2020.
- [26] J. P. Robinson, Z. Khan, Y. Yin, M. Shao, and Y. Fu. Families in wild multimedia (FIW MM): A multimodal database for recognizing kinship. *IEEE Transactions on Multimedia*, 2021.
- [27] J. P. Robinson, Y. Li, N. Zhang, Y. Fu, and S. Tulyakov. Laplace landmark localization. In *Inter. Conference on Computer Vision (ICCV)*, 2019.
- [28] J. P. Robinson, G. Livitz, Y. Henon, C. Qin, Y. Fu, and S. Timoner. Face recognition: too bias, or not too bias? In *CVPR Workshop*, pages 0–1, 2020.
- [29] J. P. Robinson, C. Qin, Y. Henon, S. Timoner, and Y. Fu. Balancing biases and preserving privacy on balanced faces in the wild. *arXiv preprint arXiv:2103.09118*, 2021.
- [30] J. P. Robinson, M. Shao, and Y. Fu. To recognize families in the wild: A machine vision tutorial. In *ACM on Conference on MM*, 2018.
- [31] J. P. Robinson, M. Shao, and Y. Fu. Survey on the analysis and modeling of visual kinship: A decade in the making. *Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, (01):1–1, 2021.
- [32] J. P. Robinson, M. Shao, Y. Wu, and Y. Fu. Families in the wild (FIW): Large-scale kinship image database and benchmarks. In *ACM on Conference on Multimedia (MM)*, 2016.
- [33] J. P. Robinson, M. Shao, Y. Wu, and Y. Fu. Families in the wild (fiw): Large-scale kinship image database and benchmarks. In *ACM on Conference on Multimedia (MM)*, 2016.
- [34] J. P. Robinson, M. Shao, Y. Wu, H. Liu, T. Gillis, and Y. Fu. Visual kinship recognition of families in the wild. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2018.
- [35] J. P. Robinson, M. Shao, H. Zhao, Y. Wu, T. Gillis, and Y. Fu. Recognizing families in the wild (RFIW). In *ACM MM RFIW Workshop*, 2017.
- [36] J. P. Robinson, Y. Yin, Z. Khan, M. Shao, S. Xia, M. Stopa, S. Timoner, M. A. Turk, R. Chellappa, and Y. Fu. Recognizing families in the wild (RFIW): The 4th edition. In *Conference on Automatic Face and Gesture Recognition (FG)*. IEEE Computer Society, 2020.
- [37] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [38] A. Shadrakov. Achieving better kinship recognition through better baseline. In *FG Challenge on RFIW*, 2020.
- [39] M. Wang and W. Deng. Deep face recognition: A survey. *Neurocomputing*, 2020.
- [40] S. Wang, Z. Ding, and Y. Fu. Cross-generation kinship verification with sparse discriminative metric. *Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, 41(11):2783–2790, 2018.
- [41] S. Wang, J. P. Robinson, and Y. Fu. Kinship verification on families in the wild with marginalized denoising metric learning. In *Conference on Automatic Face and Gesture Recognition (FG)*, 2017.
- [42] Z. Wei, M. Xu, L. Geng, H. Liu, and H. Yin. Adversarial similarity metric learning for kinship verification. *IEEE Access*, 7:100029–100035, 2019.
- [43] C. Whitelam, E. Taborsky, A. Blanton, B. Maze, J. Adams, T. Miller, N. Kalka, A. K. Jain, J. A. Duncan, K. Allen, et al. Iarpa janus benchmark-b face dataset. In *CVPR Workshop*, 2017.
- [44] Y. Wu, Z. Ding, H. Liu, J. Robinson, and Y. Fu. Kinship classification through latent adaptive subspace. In *Conference on Automatic Face and Gesture Recognition*. IEEE, 2018.
- [45] S. Xia, M. Shao, and Y. Fu. Kinship verification through transfer learning. In *Inter. Joint Conferences on AI (IJCAI)*, 2011.
- [46] S. Xia, M. Shao, J. Luo, and Y. Fu. Understanding kin relationships in a photo. *Trans. on Multimedia*, 2012.
- [47] J. Yu, G. Xie, X. Hao, Z. Cui, L. Zhang, and Z. C. Cai. Deep kinship verification and retrieval based on fusion siamese neural network. In *Conference on Automatic Face and Gesture Recognition (FG)*, 2021.
- [48] J. Yu, G. Xie, M. Li, and X. Hao. Deep fusion siamese network for automatic kinship verification. In *FG Challenge on RFIW*, 2020.
- [49] J. Yu, G. Xie, M. Li, and X. Hao. Retrieval of family members using siamese neural network. In *FG Challenge on RFIW*, 2020.
- [50] K. Zhang, Y. Huang, C. Song, H. Wu, and L. Wang. Kinship verification with deep convolutional neural networks. In *British Machine Vision Conference (BMVC)*, 2015.
- [51] X. Zhang, M. Xu, X. Zhou, and G. Guo. Supervised contrastive learning for facial kinship recognition. In *Conference on Automatic Face and Gesture Recognition (FG)*, 2021.
- [52] J. Zhao, J. Li, X. Tu, F. Zhao, Y. Xin, J. Xing, H. Liu, S. Yan, and J. Feng. Multi-prototype networks for unconstrained set-based face recognition. In *Inter. Joint Conferences on AI (IJCAI)*, 2019.
- [53] L. Zhipeng, Z. Zhiguang, X. Zhenyu, and C. Lixuan. Challenge report-recognizing families in the wild data challenge. In *FG Challenge on RFIW*, 2020.