IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 27 January 2024, accepted 24 March 2024, date of publication 8 April 2024, date of current version 24 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3385782

RESEARCH ARTICLE

DaliID: Distortion-Adaptive Learned Invariance for Identification–A Robust Technique for Face Recognition and Person Re-Identification

WES ROBBINS[1,2], GABRIEL BERTOCCO[2,3], AND TERRANCE E. BOULT[2], (Fellow, IEEE)
[1]Department of ECE, The University of Texas at Austin, Austin, TX 78712, USA
[2]Department of Computer Science, University of Colorado at Colorado Springs, Colorado Springs, CO 80918, USA
[3]Institute of Computing, Laboratory of Artificial Intelligence, University of Campinas, Campinas 13083-970, Brazil

Corresponding author: Terrance E. Boult (tboult@uccs.edu)

This work was supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), under Grant 2022-21102100003; and in part by São Paulo Research Foundation (FAPESP) under Grant 2022/02299-2.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board (IRB) under Protocol No. UCCS 2022-131-MAIN and No. UCCS 08-123-CNV.

ABSTRACT In real-world applications, face recognition, and person re-identification are subject to image degradations such as motion blur, atmospheric turbulence, or upsampling artifacts—which are known to lower performance. This work directly addresses challenges in low-quality scenarios with 1) practical, novel updates to training and inference, which improve robustness to realistic distortions in face recognition and person re-identification and 2) new datasets for long-distance recognition. We propose a method that progressively learns from images prone to soft and strong distortions caused mainly by atmospheric turbulence. The method has a novel distortion loss to improve robustness, which is empirically shown to be highly effective in low-quality scenarios. Two further strategies are proposed to integrate distortion augmentation while also retaining the highest performance in high-quality scenarios. First, during training, an adaptive weighting schedule, which leverages the construction of different levels of distortion augmentation, is used to train the model in an easy-to-hard manner. The second, at inference, is a magnitude-weighted fusion of features from the parallel models used to retain the highest robustness across both high-quality and low-quality imagery. Different from prior work, our model does not leverage any image restoration or style transfer technique, and we are the first to employ explicit distortion weighting during training and evaluation. Our model achieves the best performances compared to prior works on face recognition and person re-identification benchmarks, including IARPA Janus Benchmark-S (IJB-S), TinyFace, DeepChange, Multi-Scene Multi-Time 2017 (MSMT17), and our novel long-distance datasets.

INDEX TERMS Biometrics, face recognition, computer vision, person re-ID, deep learning.

I. INTRODUCTION

Humans can recognize faces or objects before and after considerable distortions. Consider Dali's renowned works *Persistence of Memory* and *Lincoln in Dalivision* shown in part I of Figure 1 where the reader will have no trouble recognizing multiple clocks or Lincoln, despite the distorted presentation. Comparatively, neural networks are brittle when presented with even mildly distorted images. Within the field of biometrics, the tasks of face recognition and person re-identification can be subject to distortions at inference time, such as atmospheric turbulence, motion blur, and artifacts from upsampling. Such distortions are common in security-sensitive settings such as energy infrastructure

The associate editor coordinating the review of this manuscript and approving it for publication was Zahid Akhtar.

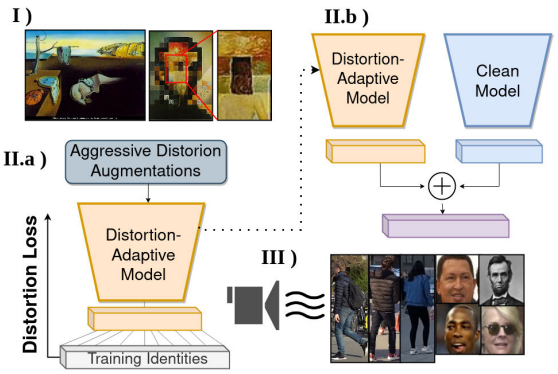


FIGURE 1. To overcome realistic distortions encountered by biometric models operating in unconstrained scenarios, we propose II.a) a novel training procedure for distortion robust models and II.b) magnitude-weighted feature-fusion from high- and low-quality training domains. To supplement evaluations on realistic distortions, III) we collect and provide an IRB-approved academic-use dataset at a range of 750+ meters.

security, surveillance systems, or counter-terrorism [27]. Thus, there is a significant social need for models that are robust in these conditions.

This work proposes practical, novel updates to training and inference to improve model performance in challenging test-time scenarios. Additionally, to aid evaluation in such scenarios, we collect and provide an IRB (Institutional Review Board)-approved long-distance recognition dataset from over 750+ meters. To demonstrate the generality of the proposed method, we perform experiments with benchmarks for both face recognition and person re-identification. Long-distance data is also collected for both face recognition and person re-identification.

The first contribution of this work is training with a mix of clean (without distortions) images and novel *atmospheric distortion augmentations*, which combines realistic spatial distortion and blur. While prior work [33] has used augmentations such as cropping and downsampling for face recognition, our augmentation contains a more complex transformation that is more closely matched to scenarios with motion blur, atmospheric turbulence, and even upsampling artifacts (see results on the TinyFace dataset in Table 4). Thus, by introducing our augmentation in training, the test-time domain shift is decreased in challenging scenarios as we improve invariance to distortions, see Fig. 8. This is especially true in the case of face recognition as the training data [92] is scraped from the web and is predominated by high-quality portrait images. By carefully tuning the strength of the distortion augmentation, DALIFace significantly improves performance on low-quality benchmarks such as IJB-S [32] and TinyFace [8]. Thus, an insight of this work is that carefully tuned augmentations are under-utilized in unconstrained scenarios. While our augmentation alone achieves the best performances compared to prior works on low-quality benchmarks, further adaptations discussed below allow us to achieve high performance on standard-quality datasets such as IJB-C [50]. Our augmentation is performed by leveraging the atmospheric turbulence image simulator proposed in [49]. It is important to note that the authors of the simulator proposed an algorithm to generate simulated data under different levels of atmospheric turbulence, however they have not employed it as augmentation in any model training. Conversely, we use the simulated atmospheric turbulence data for training our new architecture.

To integrate the augmentation during training, we propose an adaptive weighting mechanism that trains the model in an easy-to-hard manner. Each sample in every batch is reweighted as a function of the training iteration number and the strength of the augmentation. The augmentation's strength (meaning severity) on any given image is sampled from an empirically tuned distribution. In early training iterations, images with higher distortion are assigned lower weighting, and images with no distortion are assigned the greatest weighting. The weighting of distorted samples is increased throughout training such that by the end of training, all samples have equal weighting. The proposed weighting strategy is highly effective in combination with our distortion augmentation, as shown in Section V. We refer to a backbone trained with the distortion augmentation and the adaptive-weighting schedule as a *distortion-adaptive* model. For person re-identification, we additionally propose to use multiple class centers and class proxies that allow the model to better adapt to training distortions. The corresponding proxy loss (see Section III-C) also follows the adaptive-weighting schedule.

To further improve robustness at inference, two backbones are run in parallel: a distortion-adaptive backbone and a standard (or 'clean') backbone. The final distance between samples for open-set evaluations is calculated with a magnitude-weighted combination of feature distances from each backbone, respectively. Feature magnitude is used since it reflects the response of the learned features at the final layer, which is known to be correlated with sample quality [13], [33], [51]. Maybe surprisingly, this fusion approach is more robust than more complicated learned fusions such as an attention layer or full transformer encoder. Relative to a single distortion-adaptive backbone, the parallel backbone fusion improves performance on face recognition at low false-positive thresholds (e.g., IJB-C TAR@FAR=1e-4) and on all person re-identification benchmarks. The final result is a method that is highly robust across evaluation scenarios for both face recognition and person re-identification. We refer to the entirety of our proposed strategy as **DaliID**: **D**istortion-**A**daptive **L**earned **I**nvariance for **Id**entification. The effectiveness of DaliID is demonstrated empirically, showing it achieves the best performance compared to prior works on seven publicly available benchmarks: IJB-S, IJB-C, TinyFace, CFP-FP, Market1501, MSMT17, and DeepChange.

The final contribution of this work is the recapture of face recognition data over long distances with high-end imaging equipment and displays. At 750+ meters, our proposed datasets have the longest range of any academic-use dataset available. The collection process and hardware specifications are discussed in detail in Section IV-A. In Section V, prior

works are compared on our proposed evaluation datasets. The datasets will be made available for academic use.

In summary, the contribution of this work includes:

- Propose a novel distortion-adaptive training strategy in which we leverage the construction of distortion augmentation for an easy-to-hard weighting scheme with our novel distortion loss function to achieve improved distortion invariance, see Figure 8.
- Show our distortion-adaptive training is better than simply training with added atmospheric distorted images.
- Design a novel weighted combination strategy based on the feature magnitudes from both backbones from the training phase, allowing us to exploit complementary knowledge and reach the best performances compared to prior works across evaluation scenarios.
- Provide identification datasets through long-distance (750+ meters) to provide an assessment of the impact of significant atmospheric turbulence.

II. RELATED WORK

The problems of face recognition and person re-identification have been extensively studied. Most related to this paper are works that have studied low-quality conditions. For face recognition, Probabilistic Face Embeddings (PFE) [59] proposes representing faces with a Gaussian distribution in latent space to account for uncertainty. Data Uncertainty Learning (DUL) [3] builds on PFE by learning the mean and variance of the Gaussian distribution during training. URFace [60] uses data synthesis and a confidence-aware loss to learn universal representations. Several quality-aware face recognition loss functions have also been proposed. CurriculumFace [25] changes the margin of the loss throughout training, and the MagFace [51] and AdaFace [33] losses use adaptive margins that are a function of feature magnitude, which is a proxy for quality. Controllable Face Synthesis Model (CFSM) [45] is a method that learns the style of a test environment and uses a latent style model to modify training samples. CAFace is a clustering-based method for multi-frame face recognition [34]. In [56], the effects of atmospheric turbulence on face recognition are studied, where atmospheric distortions are found to significantly affect face recognition performance. Other works have developed upstream image restoration for atmospheric turbulence [39], [77], [78]. Image restoration methods focus on image-based metrics, such as PNSR, not recognition.

For person re-identification (PReID), CBDB-Net [64] proposes the Batch DropBlock to encourage the model to focus on complementary parts of the input image. CDNet [41] improves architecture search for PReID. FIDI [76] proposes a novel loss function to give different penalizations based on distances between images to encourage fine-grained feature learning. To deal with clothes-changing, Clothes-based Adversarial Loss(CAL) [17] regularizes the model learning with respect to the clothes labels to learn clothes-invariant features. There are many other prior works that leverage attention models [4], [6], [14], [15], [22], [24], [44], [55], [69], [79], [83], [84], [85], [89], neighborhood-based analysis [69], auxiliary data [20], [28], segmentation-based [31], semantics-based [30], [61], and part-based learning [61], [63], [67], [68], [73], [79], [84], [86], [90], [91]. To directly deal with different resolutions and points of view, some works leverage the camera information associated with each identity [93], super-resolution strategies [7], [29], and attention and multi-level mechanisms for cross-resolution feature alignment [54], [81]. There is insufficient space to compare orthogonally to all combinations of described methods above for PReID. We limit our scope comparison to the global feature representation learning model as described in the taxonomy of the recent survey from Ye et al. [79], in which we just perform global pooling operations over the last feature map of a CNN without further mechanisms. The core contributions of this paper are focused on learning distortion-invariant feature spaces and a methodology for dealing with distortion, which is demonstrated to be applicable to both face recognition and person re-identification. Future work should look at combining techniques such as image restoration, super-resolution, part-based mechanisms, or multi-frame aggregation with the improved feature spaces developed herein.

III. APPROACH

We propose DaliID for learning models robust to realistic test-time distortions such as motion blur, upsampling artifacts, and atmospheric turbulence. We use strong levels of distortion augmentation (Section III-A), which serves the purpose of supervising the model to learn a feature space that is invariant to distortions that have been shown to considerably degrade model performance [56], [77]. To allow the model to adapt to strong levels of augmentation, we propose an adaptive-weighting distortion-aware strategy (Section III-C) where we dynamically change the weights of different distortion levels throughout training. To get the highest performance across the range of evaluation scenarios, we train two models in parallel: one with clean images and the other with clean and distorted images (Section III-D). Then we perform a weighted combination of the feature spaces from both models based on the magnitude of the feature vectors from each, which yields the highest performance. DaliID methodology is designed for general identification scenarios such as face recognition and person re-identification tasks. Figure 2 shows an overview of the approach.

A. DISTORTION AUGMENTATIONS

Image augmentations allow better generalization by adding variance to training data. A vast space of augmentations can be performed on an image; many have been successful for computer vision tasks. However, there is a bias-variance trade-off. In this work, we leverage a new augmentation for face recognition and person re-identification training based on atmospheric turbulence to generate the different distortion levels for the images. Atmospheric turbulence contains

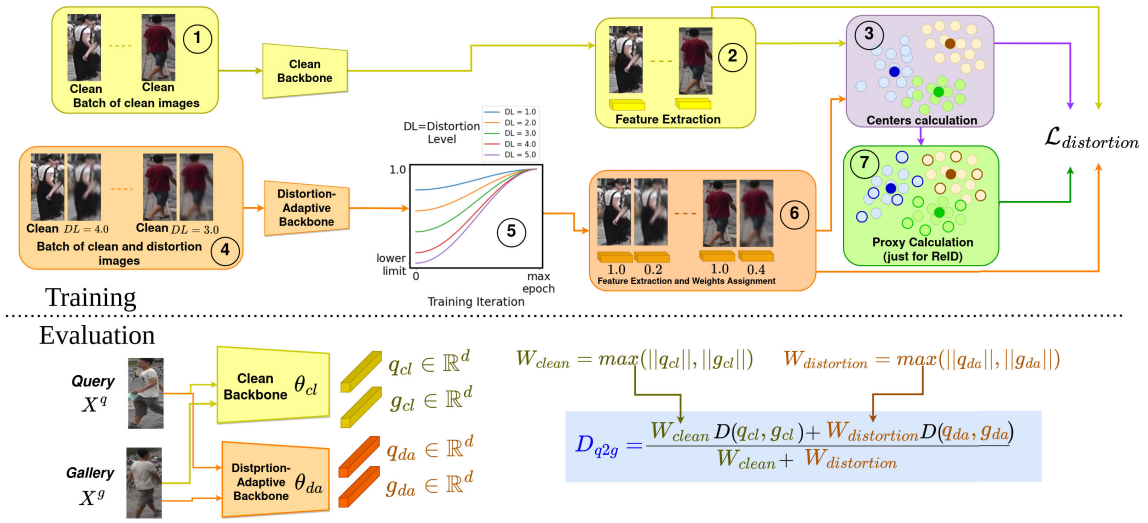


FIGURE 2. An overview of the DaliID pipeline for face recognition (DaliFace) and person re-identification (DaliReID). Steps 1, 2, and 3 are performed for training without distorted images, while steps 3, 4, 5, and 6 are distortion-adaptive training. In Step 4, we create a batch of clean and distorted images; then a dynamically varied weight is assigned as a function of the distortion level (*DL*) (Step 5). Then we extract the features and optimize the distortion loss ($\mathcal{L}_{distortion}$). Step 7 is applied just for DaliReID training due to the high intra-class variation to sample count ratio faced in the whole-body recognition task. On evaluation, both clean and distortion-adaptive backbone decisions are weighted and combined based on the magnitudes of the query and gallery feature vectors to obtain the final decision (distance) for retrieval.

random temporally and spatially variable distortions, which are absent in Gaussian blur or down-sampling augmentations. Atmospheric turbulence simulation code [49] is used to implement the augmentation, which generates physically realistic distortions. It is important to note that the author of [49] proposed an algorithm to generate simulated data under different levels of atmospheric turbulence, however they have not employed it on any training or evaluation. Conversely, we propose to use the simulated atmospheric turbulence data for training in order to achieve distortion-invariant feature representation, which has not been done before by any prior work. Our approach of simulated distortions is of practical interest because it is not tractable to collect real labeled data through atmospherics at a scale suitable for training deep learning models. Experimentally, we find training with our distortion augmentation yields the best performances compared to prior works on long-distance and low-resolution test sets. Distortion levels used herein are based on different atmospheric turbulence conditions to train our models. Since our base model is AdaFace [33] for face recognition, we provide a brief background about it in the next section.

B. BACKGROUND OF FACE RECOGNITION WITH ADAPTIVE MARGIN

For face recognition, the AdaFace [33] loss is used, which uses an adaptive margin as a function of the feature norm. The adaptive margin includes both an angular margin g_{angle} and an additive margin g_{add} calculated as

$$g_{angle} = -m \cdot \widehat{||x_i||}, \quad g_{add} = m \cdot \widehat{||x_i||} + m, \tag{1}$$

where $\widehat{||x_i||}$ is the feature magnitude after normalizing the magnitudes with batch statistics. m is a margin hyperparameter. The penalty for each sample can be represented with the piece-wise function f:

$$f(\theta_j, m) = \begin{cases} s\cos(\theta_j + g_{angle}) - g_{add} & j = y_i \\ s\cos\theta_j & j \neq y_i, \end{cases} \tag{2}$$

where θ_j is the angle between the feature vector from the backbone proxy class-center of the j^{th} class. Scalar s is a hyperparameter, and y_i is the ground-truth. The final AdaFace loss $\mathcal{L}_{adaface}$ is then calculated as follows:

$$\mathcal{L}_{adaface} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{f(\theta_j, m))}}{e^{f(\theta_j, m))} + \sum_{j \neq y_i} e^{f(\theta_j, m)}}. \tag{3}$$

The loss function 3 is closer to zero when the logarithmic argument gets closer to 1. Since the numerator is also present in the denominator, just the term $\sum_{j \neq y_i} e^{f(\theta_j, m)}$ needs to be zero. More specifically, θ_j is the angle between the feature vector of the input image and the proxy of the j^{th} class, as explained in the last paragraph. Then the mentioned summation will be zero when $f(\theta_j, m)$ assumes the lowest possible value, which means that the feature vector will be encouraged to be the farthest possible from all class proxies except its own class proxy. It turns out that, after optimization, the feature vector will be close to its own class proxy and farther away from the other class proxies, consequently $f(\theta_j, m))$ with $j \neq y_i$ will be lower, and the summation $\sum_{j \neq y_i} e^{f(\theta_j, m)}$ will be closer to zero, turning Equation 3 also closer to zero. In summary, Equation 3 encourages an input image feature vector to be close to its own proxy class and apart from the other proxy classes, encouraging the class's separability and increasing model discrimination ability.

C. ADAPTIVE WEIGHTING

Different levels of distortion compress different degrees of difficulty during training. Randomly sampling images from different distortion levels can result in sub-optimal performance since higher distortion levels (i.e., lower-quality samples) dominate the gradient during training. In other words, the mere use of atmospheric turbulence data as an augmentation might deteriorate the performance in standard (high-quality) datasets and cross-quality datasets (Table 9). Therefore, a strategy needs to be designed to effectively employ the distorted data based on simulated atmospheric turbulence. In counterpart, hard-training mining strategies have shown promising performance in PReID models [21], [58] and face recognition [25]. In this context, we propose an easy-to-hard training regime in which we start by assigning higher weights for lower levels of distortion and lower weights for higher levels of distortion. *Different than prior works, we directly leverage the construction of the augmentation to assign weights.* Weighting the loss as a function of the distortion level allows the model to focus on easier examples (by giving them higher weights). By lowering the weighting of high-distortion samples, the model becomes distortion-aware without allowing them to dominate the loss in early epochs. As the training progresses, the weights for all distortion levels increase according to a cosine schedule. Step 5 of Figure 2 illustrates the weighting for each distortion level and is formally described below.

The distortion-aware training considers a batch of images $B = \{X^i\}_{i=1}^{N_b}$ that is composed by a mix of clean (without distortions) images X^i_{cl} and distorted images X^i_{dl} with distortion levels randomly sampled from five possible values ($\text{dl} \in \{1, 2, 3, 4, 5\}$), where N_b is the batch size. A higher dl value indicates a stronger distortion. We keep the same number of clean and distorted images in the batch. Then features f^i_t, with $t \in \{\text{cl}, \text{dl}\}$ are extracted from the backbone (θ_{da}). During the loss calculation, the respective weight w^i_t is assigned to each image according to the cosine weighting schedule. These steps are shown in Steps 4, 5, and 6 in Figure 2. For the same distortion level, the weights increase along the training following a cosine schedule (Step 5). After that, the centers are obtained for each class (Step 3), and if we are performing PReID training, we also take the classes' proxies in Step 7. The distortion loss is calculated as follows:

$$\mathcal{L}_{ce}(f, q, P) = -log \frac{e^{cos((\omega_{fq}+m_1)/\tau)+m_2}}{e^{cos((\omega_{fq}+m_1)/\tau)+m_2} + \sum_{p\in P, p\neq q} e^{(cos\omega_{fp})/\tau}} \tag{4}$$

$$\mathcal{L}_{distortion} = \frac{1}{W} \sum_{i=1}^{N_b} \sum_{t\in\{\text{cl},\text{dl}\}} w^i_t \mathcal{L}_{ce}(f^i_t, p_+, P), \tag{5}$$

where p_+ is the positive class-center (i.e., proxy), P is the set of all class-centers, ω_{fq} is the angle between vectors f and q (same definition for ω_{fp}), and $W = \sum_{i=1}^{|B|} \sum_{t\in\{\text{cl},\text{dl}\}} w^i_t$. For hyperparameters, τ is temperature, m_1 is angular margin, and m_2 is the additive margin. For face recognition $\tau = 1$ and m_1, m_2 are adaptive as proposed in AdaFace [33]. For person re-identification, $m_1, m_2 = 0$ and $\tau = 0.05$. We highlight that, different from the previous margin-based loss function (Eq. 2), we do not employ the hyperparameter s relying just on the margin and temperature parameters.

1) THE CLASS PROXIES FOR PREID

In this subsection, we present how we calculate the class proxies for PReID. To better adapt to distortions, we extend the use of multiples proxies [71] to the supervised case. This is necessary due to limited training samples and high intra-class variance [67], [79], which occurs since the whole-body images are captured from different cameras resulting in views of the same person in different poses, illumination conditions, backgrounds, occlusions, and resolutions. Step 7 of Figure 2 shows the multiple proxies with the circles with dark outlines.

Without loss of generality, consider a class $C = \{c_1, \ldots, c_{N_C}\}$ in the dataset with N_C examples. To calculate the proxies set, we start by randomly selecting a sample $c_i \in C$ ($1 \leq i \leq N_C$) to be the first proxy, and we calculate the distance between c_i and each element in C and store these distances in a cumulative vector $V_C \in R^{N_C}$. We call the first proxy $p^1_C = c_i$. To calculate the second proxy, we consider the element with the furthest distance to the first proxy (the sample with maximum distance value in V_C). Formally:

$$p^2_C := \arg\max V_C. \tag{6}$$

After that, we calculate the distance of p^2_C to all samples in C to obtain the distance vector $D(p^2_C) \in R^{N_C}$. Then we update V_C considering its current values (the distances of the class samples to the first proxy) and $D(p^2_C)$ (the distance of the class samples to the second proxy) following the formulation:

$$V_C := min(V_C, D(p^2_C)), \tag{7}$$

where $min(.,.)$ is the element-wise minimum operation between two vectors. More specifically, the j^{th} position of V_C will hold the minimum distance of the sample $c_j \in C$ considering the first and second proxies. So the j^{th} position holds the distance of c_j to the closest proxy, and the maximum value in V_C is from the farthest sample from both proxies. We consider this sample as the next proxy p^3_C. To obtain p^3_C, we apply again Eq. 6 but considering the updated V_C calculated from Eq. 7, and repeat the whole process again for the new proxy. We write both equations in their general formats:

$$p^t_C := \arg\max V^{t-1}_C \tag{8}$$

$$V^t_C := min(V^{t-1}_C, D(p^t_C)). \tag{9}$$

As explained before, we initialize $V^1_C := D(p^1_C)$ where p^1_C has been randomly selected from C to be the first proxy. We keep alternating between Eq. 8 and 9 until $t = 5$ to get five proxies per class. During training, for sample $X_i \in B$ (where B is the batch), we call P_i by the proxies set of its class and N_i

by the set of the top-50 closest negative proxies and use them to calculate $\mathcal{L}_{proxy}$ in Eq. 10.

$$\mathcal{L}_{proxy} = \frac{1}{W}\sum_{i=1}^{N_b}\sum_{t\in\{\text{cl},\text{dl}\}} w_t^i \frac{1}{|P_i|}\sum_{q\in P_i} L_{ce}(f_t^i, q, P_i \cup N_i). \tag{10}$$

After that, $\mathcal{L}_{proxy}$ loss is added in Eq. 5, obtaining the final loss function in Eq. 11 for PReID:

$$\mathcal{L}_{distortion} = \frac{1}{W}\sum_{i=1}^{N_b}\sum_{t\in\{\text{cl},\text{dl}\}} w_t^i \mathcal{L}_{ce}(f_t^i, p_+, P) + \lambda\mathcal{L}_{proxy}, \tag{11}$$

where λ controls the contribution of $\mathcal{L}_{proxy}$ to the final loss. $\mathcal{L}_{distortion}$ is applied for both distortion-adaptive and clean backbones training. To train the clean backbone, we have $w_i = 1$ for all samples because no distortion augmentations are applied. The class proxy calculation is used just for person re-identification training.

To train the clean and distortion models, we employ the Adam [37] optimizer with weight decay of $5 \cdot 10^{-4}$ and initial learning rate of $3.5 \cdot 10^{-4}$. We train both models for 250 epochs and divide the learning rate by 10 every 100 epochs. As explained, the number of proxies per class is fixed in 5 (i.e., $\forall_i |P_i| = 5$) for all datasets. To create the batch to optimize the clean model, we adopt a similar approach to the PK batch strategy [21] in which we randomly choose P identities and, for each identity, K clean images (without distortion). To train the distortion model, we sample K clean images and K distorted images randomly sampled from five different levels of distortion strength. We also apply random crop, random horizontal flipping, random erasing, and random changes in brightness, contrast, and saturation as data augmentation.

To improve the performance, we adopt the Mean-Teacher [65] to self-ensemble the weights of the backbones along the training. Considering both clean and distortion-adaptive backbones with parameters θ_{cl} and θ_{da} (which are initialized with weights pre-trained on Imagenet), respectively, we keep another backbone for each one with parameters Θ_{cl} and Θ_{da} with the same architecture to self-ensemble their weights along training through the following formula:

$$\Theta_s^{t+1} := \beta\Theta_s^t + (1-\beta)\theta_s^t, \tag{12}$$

where $s \in \{\text{cl}, \text{da}\}$, β is a hyper-parameter to control the inertia of the weights, and t is the instant of time. We set $\beta = 0.999$ for all models following prior person re-identification works [16], [80]. We use the backbones Θ_{cl} and Θ_{da} for the final evaluation.

D. CROSS-DOMAIN FUSION

The distortion-adaptive backbone improves performance on face recognition benchmarks and for person re-identification but not on all high-quality scenarios for face recognition at low false-positive thresholds. In practice, we do not know the test-time distortion level; thus, a good model should be robust across all scenarios. Inspired by recent studies in magnitude-based training [33] and the effects of atmospheric turbulence in face recognition [56], we propose to employ the feature magnitude also on evaluation to combine knowledge from different backbones. More specifically, we train a backbone without distortion augmentations, denoted θ_{cl} (which we called "clean model"), in parallel to the distortion-adaptive backbone, denoted by θ_{da}. To leverage knowledge from both backbones, we apply magnitude-weighted fusion between the backbones as shown in Figure 2. We call this cross-domain fusion since the backbones were trained on different training distributions. The main rationale is that since we do not know the distortion level of images in testing scenarios, we can use the magnitude of the feature vectors generated by each backbone as a proxy for it. In other words, the stronger the distortion the lower the magnitude of the output feature vector from the clean model (since it has not been trained with distorted data), and the higher the magnitude of the output feature vector from the distortion-adaptive backbone. Then we use the magnitude of the feature vectors to weigh the decision from each backbone. The advantage of this approach is evident in Table 9. At inference, for a query and gallery image pair, we extract both feature vectors $q_{\text{cl}} = \theta_{\text{cl}}(X^q)$ and $g_{\text{cl}} = \theta_{\text{cl}}(X^g)$ considering the clean model with parameters θ_{cl}, and the feature vectors $q_{\text{da}} = \theta_{\text{da}}(X^q)$ and $g_{\text{da}} = \theta_{\text{da}}(X^g)$ considering the distortion-adaptive model with parameters θ_{da}. For person re-identification we use the respective self-ensembled weights Θ_{cl} and Θ_{da} calculated in Eq. 12. We calculate the distance between the query and gallery considering each backbone to obtain distances $D(q_{\text{cl}}, g_{\text{cl}})$ and $D(q_{\text{da}}, g_{\text{da}})$, which are weighted combined considering the maximum feature magnitude for each pair before L2 normalization as shown in the equation on the lower half of Figure 2.

IV. DATASETS

Many datasets from two different modalities are used in our evaluations. Table 1 is provided as a reference for the different characteristics of the datasets, and well-known datasets are only briefly described. Figure 6 shows samples from the LD datasets and Figure 7 shows samples from the government-use long-range-dataset.

A. LONG DISTANCE RECAPTURE DATA

As discussed in Section I, long-range recognition is relevant in many applications. However, the collection of biometric data is extremely expensive and time consuming. Currently, the most related dataset, IJB-S [32], is not available for common academic use, and an earlier dataset at 100M [18] was withdrawn from public use. Furthermore, IJB-S is not a strictly long-range dataset. To overcome the lack of available long-range data, some prior works have used simulated atmospheric turbulence as a proxy for real data [56], [77], [78].

TABLE 1. Reference table of datasets. Six well known face datasets are used for direct comparison with prior work. The CFP-LD and LFW-LD datasets are novel contributions of this work; see Sec.IV-A to support controlled long-distance evaluation. The LRD (Long-Range Dataset) is a government-use dataset. For CFP-LD and LFW-LD, we use the same evaluation as standard LFW and CFP. For LRD, we use the same metrics as for IJB-S because the dataset has the same gallery/query format. "PReID" holds for "person re-identification". All other evaluation metrics follow standard practice from prior works. mean Average Precision (mAP), Rank-1 (R-1), and Rank-5 (R-5) are retrieval metrics.

Dataset Reference

Dataset	Modality	Evaluation Metric	Characteristics
CFP-FP	face	1:1 verification	Relatively high-quality; frontal-profile pairs
LFW	face	1:1 verification	Relatively high-quality
AgeDB-30	face	1:1 verification	Relatively high-quality; pairs with 30 year difference
IJB-C	face	TAR@FAR=1e-4	Mixed-quality
IJB-S	face	R-1, R-5, TPIR@FPIR=1e-1,1e-2	High-quality gallery; low spatial resolution faces in probe video
TinyFace	face	R-1, R-5	Low spatial resolution probe and gallery
CFP-LD	face	1:1 verification	Recapture dataset at 770m; strong atmospheric turbulence
LFW-LD	face	1:1 verification	Recapture dataset at 770m; strong atmospheric turbulence
LRD	face	R-1, R-5, TPIR@FPIR=1e-1,1e-2	HQ gallery images; query images up to 500m. Government-use.
DeepChange	PReID	mAP, R-1	16-cameras low-resolution with 450 clothes-changing identities
Market	PReID	mAP, R-1	6-cameras low/high-resolution with 751 same-clothes identities
MSMT17	PReID	mAP, R-1	15-cameras low/high-resolution with 1041 same-clothes identities

However, the effectiveness of simulated atmospherics for face recognition has not been validated because, as mentioned before, there is no real data for validation.

We recapture datasets through the atmosphere to facilitate academic research on biometric recognition over long distances. To perform the capture, we use three 4k outdoor televisions, a 4k Basler camera, and an 800 mm lens with a 1.4x adapter. Custom capture and display software are developed for the collection, and custom mounting hardware is built for stable capture. The displays are mounted to avoid direct sunlight on the screens. The camera is directed at the displays from a structure at a distance of 770 meters, and videos of the displays are captured at 30 frames per second. A video of the displays running is provided on the GitHub site of the data, where considerable atmospheric effects can be seen. Our collection setup yields significant atmospheric distortions, which can be noticed between sequential frames. Figure 3 shows two examples. The data collection process went through IRB approval and is being distributed for non-commercial use.

We refer to our recapture datasets as the original dataset name followed by "-LD" ("the LD datasets"), where LD stands for long-distance. The evaluation datasets provided are LFW-LD and CFP-LD. Twelve recaptured samples are provided for each image in the original dataset because atmospheric turbulence is temporally variable. For CFP-LD and LFW-LD, two protocols are proposed: clean-to-long-distance (C-to-LD) and long-distance-to-long-distance (LD-to-LD). C-to-LD uses verification pairs where one image is standard and thus higher quality. For LD-to-LD, all samples are recaptured over long distances. The LD datasets expand the evaluation of our methods in the following section, where evaluations are made with a single frame for each image. However, future work should consider new protocols allowing frame fusion or frame selection across frames. Previous unconstrained evaluation datasets (e.g., IJB-S) have been distributed over a terabyte of raw video, which is burdensome to process. In contrast, the LD datasets are pre-processed and pre-aligned in the same format as the original datasets, which streamlines evaluation and comparison. The final release will include the recapture of person re-identification datasets, plus a WebFace4M recapture for training.

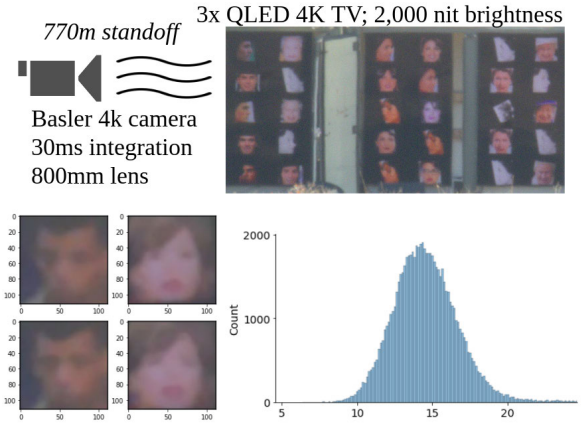


FIGURE 3. Top. Recapture specifications and a raw frame from our recollection. Lower Left. Two consecutive frames (33.3 ms apart) for two different identities from LFW-LD. Differences can be observed between sequential frames, such as around the eyes or face outline. Lower Right. Distribution of feature distances in degrees between sequential frames of the same display image from LFW-LD and CFP-LD. Surprisingly, the distances are not 0 – the effects of atmospherics from frame to frame are considerable!.

The collection setup went through IRB approval, and both the LFW and CFP dataset licenses allow redistribution. Specifications of imaging equipment and collection conditions are shown in Table 2. Figure 4 shows the display and Figure 5 shows the camera used for recapture. The LD datasets contain 12 recaptured face chips for each original face as the capture occurs continuously over time, and atmospheric turbulence is temporally variable (atmospheric effects are shown in Figure 6 and at https://youtu.be/cBcik5U7kfM). The nature of the data allows for research uses such as frame selection, frame aggregation, distortion robustness, quality prediction, and direct feature comparisons to the same image with and without real atmospheric turbulence.

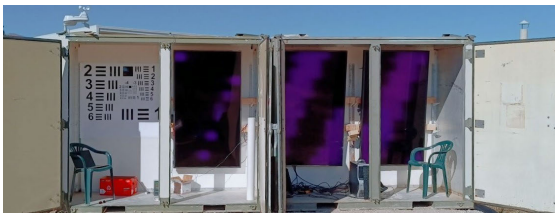


FIGURE 4. 3 × 75″ 4k OLED 2,000 nit outdoor displays mounted in containers for recapture.

FIGURE 5. Lens and camera with custom mounting hardware for recapture.

To post-process the images, fixed regions from the screens are cropped, and then RetinaFace [10] face detector is used to detect landmarks and realign the images. A non-local mean denoising algorithm is used to reduce noise in the recaptured images. Figure 6 shows samples from the LD datasets.

V. RESULTS

Our experiments are performed on face recognition and person re-identification tasks with an emphasis on low-image-quality scenarios. Common training and evaluation procedures are followed for each task, respectively. We start the evaluation of the face recognition models on five low-image-quality datasets and four standard image-quality datasets and then on recaptured and real-long distance data. The low-resolution TinyFace [8] dataset has 2,569 probe identities and 157,871 gallery images. Following previous work [33], 1:N Rank-1 and Rank-5 are presented for TinyFace. The IJB-S [32] contains gallery images for 201 identities and over 30 hours of probe video. For IJB-S, we report the surveillance-to-booking and surveillance-to-surveillance protocols. Additional details on IJB-S evaluation can be found in Sec V-B. Our LFW-LD and CFP-LD datasets (see Section IV-A) are evaluated with 1:1 accuracy. Representing standard image quality scenarios, we report on LFW [23], CFP [57], AgeDB [53], and IJB-C [50] with standard metrics. Among them, we report the True Acceptance Rate (TAR) at False Acceptance Rate (FAR) in 1% and 10% (TAR@FAR = 1% or 10%). Training is done on the WebFace4M and WebFace12M datasets [92]. Training is not performed on MS1Mv* datasets due to redaction.

For person re-identification (PReID), we used two same-clothes datasets: Market1501 and MSMT17, and one clothes-changing dataset: DeepChange. For PReID evaluation, following prior work, experiments are run with predefined train-test splits, and mAP and CMC metrics are reported. Market1501 [87] has 12,936 images of 751 identities in the training set. The test set is divided into 3,368 images for the query set and 15,913 images for the gallery set. MSMT17 [74] is the most challenging same-clothes ReID dataset. It comprises 32,621 images of 1,401 identities in the training and validation sets and 93,820 images of 3,060 identities in the test set. DeepChange [75] has lower-quality images than MSMT17 and Market. It has 75,083 images of 450 identities on the training set. The validation and test sets are divided into query and gallery sets.

TABLE 2. Camera, weather condition, and display settings for the collection of LFW-LD and CFP-LD.

Parameter	Value
Camera	Basler acA2440-35uc
Lens focal length	800 mm +1.4x Extender
Capture distance	770 meters
Integration time	30 μs
Capture rate	30 fps
Wind speed	5-15 mph
Temperature	15 °C

In addition to the previous public data, a non-public government-owned long-distance identification dataset is used for added validation. For evaluation, we use a gallery of 375 subjects with one high-quality image for enrollment. These are compared with 1,219 probe images captured at multiple distances up to 500m. Performance on the Real Long Distance (RLD) provides another comparison under real atmospheric turbulence and a comparison to our LFW-LD and CFP-LD.

We also would like to point out that, despite our proposed real long-range dataset, it is hard to get enough real atmospheric turbulence data for training, and for this reason we employ the simulated distortion data during feature learning. We employ the real one just on evaluation as we show in Table 4 and Table 5. We argue that the turbulence effects on images can be diverse; it depends on the location, temperature, weather, distance to the camera, specs of the acquisition equipment, and so on, which can be hard to quantify and qualitatively compare the simulated and real imaging under all conditions. Besides, visual analysis is subjective and subjected to the background and context of the viewer. Even if the simulated data is not exactly the same as real data, it can effectively be employed as a type of distortion that, along with the other contributions, encourages the models to learn distortion-invariant representation and achieve top-tier performance outperforming prior work in well-known face recognition and person re-identification benchmarks as we will show in the next sections.

A. EXPERIMENTAL SETTINGS

Common experimental settings are used for face recognition and person re-identification, respectively. For face recognition, a ResNet100 [19] is used as the backbone model with an embedding size of 512. Mixed precision floating point training [52] is used, and the total batch size is 1,024. Stochastic Gradient Descent (SGD) is used as the optimizer with polynomial weight decay of $5 \cdot 10^{-4}$ and momentum of 0.9. A base learning rate of 0.1 is used with

FIGURE 6. Sample images from the LD datasets. It can be seen that our recapture setup yielded significant atmospheric turbulence effects (*also see video at https://youtu.be/cBcik5U7kfM*). These datasets can facilitate research into 1) quality/confidence-aware models, 2) models that are robust to face-feature distortion, and 3) frame aggregation under atmospheric turbulence (12 frames are provided per display image).

FIGURE 7. Samples from the gallery set (top row) and query set (bottom two rows) from the long-range dataset (LRD). Query images are taken at distances between 100-500 meters. All subjects consented to image use in publication.

a polynomial learning rate scheduler. In addition to distortion augmentations discussed in Section III-A, a horizontal flip and crop are used as augmentations.

For fair comparison to the prior person re-identification work, we adopt the ResNet50 [19] and OSNet [88] as the model backbone. In ResNet50, following previous works [48], [54], we change the stride of the last residual block to 1 to increase the feature map size. Then we insert a global average pooling and global max pooling layer after the last feature map and sum their outputs element-wise [54]. After that, we add batch normalization and perform the L2 normalization to project them to the unit hyper-sphere.

B. COMPARISON TO STATE-OF-THE-ART METHODS

1) FACE RECOGNITION

The IJB-S [32] is a surveillance dataset that is distributed as a set of gallery images for 202 identities and over 30 hours of query videos. The dataset has 15 million face bounding-box annotations. To process the data, we follow the following steps:

1) Extract all 15 million annotated face regions from all images and videos.
2) Run all extracted regions through MTCNN [82] face detector. MTCNN detected 7.28M/15M face regions.
3) Use face landmarks from MTCNN for an affine transformation to fixed positions on 112 × 112 image — zero-padding is added if necessary.

Evaluation is performed with the surveillance-to-booking and surveillance-to-surveillance protocols. Surveillance-to-booking protocols use videos with thousands of frames for a query and a template of multi-view high-quality gallery images. Surveillance-to-surveillance uses surveillance video for both the probe and the gallery. Seven gallery images are used for each of the 202 identities and 7,287,724 query face detections are used.

In Table 4, DaliFace is compared to prior works on low-image-quality benchmarks. In the WebFace4M regime, DaliFace improves over the prior state-of-the-art on TinyFace by 1.96% on Rank-1 and 2.55% on Rank-5. On IJB-S, DaliFace achieves state-of-the-art on six out of eight metrics by an average of 1.53%. On the long-distance datasets (LFW-LD and CFP-LD), DaliFace averages 3.04% higher accuracy than prior work. In the WebFace12M regime, the prior state-of-the-art is CFSM [45], which uses a latent-style model to learn the domain of the testing data. In contrast, our work does not use the testing data but still improves over CFSM on 8/8 IJB-S metrics and 2/2 TinyFace metrics. DaliFace achieves state-of-the-art in 12/14 metrics in the WebFace4M regime and 11/14 metrics in the WebFace12M regime. Those results show our model is able to learn discriminative and distortion-invariant features from low-quality data and achieve state-of-the-art performance in most metrics in different face recognition benchmarks.

Table 3 shows results on high-quality datasets IJB-C, LFW, AgeDB, and CFP. Despite using significant distortions during training, our DaliFace methodology achieves comparable or higher performance on high-quality benchmarks. State-of-the-art is reached on IJB-C with an accuracy of 97.40% and on CFP-FP with an accuracy of 97.27%.

TABLE 3. Performance comparisons between DaliFace and prior works on relatively high-image-quality scenario benchmarks. Following the most common protocols, 1:1 verification accuracy is reported for LFW [23], CFP-FP [57], and AgeDB [53]; and TAR@FAR=1e-4 is reported for IJB-C [50]. Despite our focus on low-quality scenarios (see Table 4, Table 5), it can be seen that our models are competitive with or better than state-of-the-art models on popular high-image-quality benchmarks. ***Due to redaction, we do not perform training with MS1Mv1,2,3 datasets.***

Method	LFW	CFP-FP	AgeDB	IJB-C
MS1Mv* Training				
CosFace [70] (CVPR18)	99.81	98.12	98.11	96.37
ArcFace [11] (CVPR19)	99.83	98.27	98.28	96.03
GroupFace [35] (CVPR20)	99.85	98.63	98.28	96.26
CircleLoss [62] (CVPR20)	99.73	96.02	-	93.95
DUL [3] (CVPR20)	99.83	98.78	-	94.61
CF [25] (CVPR20)	99.80	98.37	98.32	96.10
URFace [60] (CVPR20)	99.78	98.64	-	96.60
DB [2] (CVPR20)	99.78	-	97.90	-
Sub-center [9] (ECCV20)	99.80	98.80	98.31	96.28
BroadFace [36] (ECCV20)	99.85	98.63	98.38	96.38
VPL [12] (CVPR21)	99.83	99.11	98.60	96.76
VirFace [43] (CVPR21)	99.56	97.15	-	90.54
DCQ [40] (CVPR21)	99.80	98.44	98.23	-
MagFace [51] (CVPR21)	99.83	98.46	98.17	95.97
Virtual FC [42] (CVPR21)	99.38	95.55	-	71.47
CFSM [45] (ECCV22)	-	-	-	95.90
WebFace4M Training				
ArcFace [11] (CVPR19)	**99.83**	99.19	**97.95**	97.16
AdaFace [33] (CVPR22)	99.80	99.17	97.90	**97.39**
Partial FC [1] (CVPR22)	**99.85**	**99.23**	**98.01**	97.22
DaliFace (ours)	**99.83**	**99.27**	97.85	**97.40**

To show that the improvements hold for actual long-distance data, we also compared DaliFace to various algorithms on the Real Long Distance (RLD) dataset, with the results in Table 5. It can be seen that our algorithm significantly improves over prior works across metrics on RLD data. DaliFace achieves TPR@FPR of 63.7% @ 1%; the next best algorithm is AdaFace at 58.3%.

The above results show our model is able to learn discriminative and robust features to varied quality data, as it achieves state-of-the-art performance in low-quality data scenarios (Table 4), keep the top-tier performance in standard-quality scenarios (Table 3), and achieve the best or second-best performance in real long-distance datasets (Table 5).

2) PERSON RE-IDENTIFICATION

DaliReID is compared with state-of-the-art methods in PReID for both the same-clothes scenario and the clothes-changing scenario. We provide the main advances and limitations from prior PReID works compared to ours in Table 6. For the same-clothes scenario, results are reported in Table 7. Our method is orthogonal to the backbone, and we show results with two backbones used in prior works: ResNet50 and OSNet [88]. DaliReID achieves the highest performance on the Market dataset, outperforming FIDI [76] by 0.8 in mAP, and the second position (along with FIDI) with R1 = 94.5. In MSMT17, the most challenging PReID benchmark, we reach the best performance by outperforming CDNet by a margin of 5.9 and 3.2 in mAP and R1, respectively, with ResNet50. With OSNet, we achieve the best performance in both datasets for both metrics. Our method is able to rank ground-truth gallery images closer to the query and outperforms prior work in mAP in all setups. To show our model generalization ability, we trained DaliReID for DeepChange, in which subjects' clothes differ among views. The results in Table 8 show our method also outperforms the state-of-the-art methods. We outperformed the recent CAL [17] by 2.9 and 6.8 in mAP and R1, respectively. Besides the clothes-changing, DeepChange has more distortions and low-quality data than Market and MSMT17. We obtain the highest gain on it for R1 and the second highest gain for mAP (after MSMT17), showing our method can better improve performance in low-quality datasets. We do not employ any kind of part-based, alignment, segmentation mask, or pose variation strategies, in order to verify the performance improvement brought just by our DaliReID model.

Following the conclusions from the previous section, our model is also able to achieve state-of-the-art performance in three cross-quality person re-identification datasets, which shows our model learns features robust to different levels of distortions in two different biometrics tasks.

C. VISUALIZATION

In Figure 8, we visualize the feature activation obtained from the clean and distortion-adaptive models.

Each of the five images in the middle depicts the same person. The first one is a clean image, i.e. without distortions, the second one was generated by employing the atmospheric turbulence (AT) simulator used in our work with level 1 of turbulence (AT = 1.0); the third one was generated by employing the simulator with level 2 of turbulence (AT = 2.0) and so on, until the last image with level 4 of turbulence (AT = 4.0). The upper left bars in the image show the feature magnitudes sorted based on the average magnitude feature value of the representations. For instance, the bar "ID: G02058 Model: CL AT-Level: Clean" (the first one in the top left) has 25 rows, and each one is a feature vector from clean images from identity G02058. The columns are their features sorted based on the average absolute value of the magnitudes of the features over the 25 clean images. Those feature vectors were obtained using the clean model trained in our solution. The second bar, "ID: G02058 Model: CL AT-Level: AT 1.0" also has 25 rows where each row is a feature vector from images of identity G02058 under level 1 of atmospheric turbulence (AT = 1.0), and the columns are also the features sorted based on the average absolute value of the magnitudes of the features over the 25 clean images. The same rationale applies to the other three bars (AT = 2.0, AT = 3.0 and AT = 4.0). We see that the features with stronger activations for the clean images (first bar) have smaller and smaller activations when they are from images under stronger atmospheric turbulence (from the second to the fifth bar), and the features with weaker activations for the clean images are stronger for images under different AT levels. This is illustrated by the red arrows in the left part

TABLE 4. Comparison of DaliFace to prior work on benchmarks containing distortions. Common metrics are reported for TinyFace [8] and for IJB-S [32] protocols surveillance-to-booking and surveillance-to-surveillance. For LFW-LD and CFP-LD (Section IV-A), 1:1 verification accuracy with clean-to-long-distance (C-to-LD) pairs and long-distance-to-long-distance (LD-to-LD) pairs. KEYS. Bold: First; Blue: Second. ** The IJB-S dataset contains over three million raw video frames and 15 million face annotations. Face recognition results are subject to detection and pre-processing steps. Using official code and pre-trained model, all WebFace{4M,12M} comparisons are run with our pre-processing to ensure fair comparison.**

Method	Dataset	TinyFace [8]		IJB-S S-to-B [32]				IJB-S S-to-S [32]				LFW-LD (Sec. IV-A)		CFP-LD (Sec. IV-A)	
		Rank-1	Rank-5	Rank-1	Rank-5	1%	10%	Rank-1	Rank-5	1%	10%	C-to-LD	LD-to-LD	C-to-LD	LD-to-LD
PFE [59]	MS1Mv2 [11]	-	-	53.60	61.75	35.99	39.82	9.20	20.82	0.84	2.83	-	-	-	-
ArcFace [11]	MS1Mv2 [11]	-	-	57.36	64.95	41.23	-	-	-	-	-	-	-	-	-
URFace [60]	MS1Mv2 [11]	63.89	68.67	61.98	67.12	42.73	-	-	-	-	-	-	-		
CF [25]	MS1Mv2 [11]	63.68	67.65	63.81	69.74	47.57	-	19.54	32.80	2.53	-	-	-	-	-
AdaFace [33]	MS1Mv2 [11]	68.21	71.54	66.27	71.61	50.87	-	23.74	37.47	2.50	-	-	-	-	-
ArcFace [11]	WF4M [92]	71.11	74.38	68.38	73.64	52.47	60.69	27.20	38.36	**4.30**	**15.95**	87.80	84.65	78.12	71.17
AdaFace [33]	WF4M [92]	**72.02**	**74.52**	**69.52**	**74.41**	**54.92**	**62.82**	**27.90**	**40.11**	4.20	14.44	**89.20**	**86.10**	**79.87**	**72.57**
DaliFace (ours)	WF4M [92]	**73.98**	**77.07**	**72.21**	**76.77**	**54.07**	**63.10**	**30.65**	**42.33**	**4.21**	**16.73**	**93.91**	**88.15**	**83.21**	**74.61**
AdaFace [33]	WF12M [92]	72.29	74.52	69.73	74.49	**56.86**	**63.98**	28.83	40.99	**4.04**	15.11	89.89	86.32	80.57	71.71
CFSM [45]	WF12M [92]	**73.87**	**76.77**	**70.36**	**75.89**	55.92	63.63	**30.44**	**41.57**	3.78	**15.88**	**90.88**	**86.62**	**83.13**	**75.10**
DaliFace (ours)	WF12M [92]	**74.76**	**77.36**	**72.19**	**76.66**	**56.04**	**64.37**	**32.25**	**43.03**	**3.81**	**16.97**	**94.00**	**89.10**	**83.98**	**74.96**

TABLE 5. Performance on the Real Long Distance (RLD) dataset, which contains real images captured at up to 500 meters (see Section V for details). The methods in italics have one or more of our contributions. All models are trained on WebFace4M [92]. The improvement of DaliFace over other prior state-of-the-art algorithms is more than the gaps between previous algorithms and is consistent with other experiments.

Real Long-Distance (RLD) dataset				
Method	Rank-1	Rank-5	1%	10%
ArcFace [11]	47.42	57.26	55.95	69.32
MagFace [51]	45.69	57.67	56.52	69.73
AdaFace [33]	49.96	59.15	58.24	71.21
AdaFace + Distortion Aug	56.52	66.28	62.10	75.72
Distortion-Adaptive	56.77	**67.27**	63.17	**76.13**
DaliFace (Ours)	**56.93**	67.02	**63.67**	75.98

TABLE 6. Comparison of the prior art and our method in terms of advantages and limitations.

Method	Advances	Limitations
OSNet [88]	It proposes a data-driven feature scale weighting and lightweight architecture to gather information from varied scales.	It is not designed for cross-resolution and long-range re-identification.
GCS [5]	It calculates local and global similarities at the batch level and employs them to improve performance.	It needs to calculate the Conditional Random Field (CRF) for each batch and estimate similarities.
SFT [46]	It explores in-batch relations through group-wise learning and spectral clustering.	Spectral clustering is done for each batch, and big batches might increase complexity.
CBN [93]	It effectively reduces the learning gap based on camera distributions.	It requires a camera label for cross-domain alignment.
STNReID [47]	It deals with partially-visible identities through 2D affine transformations.	It requires a two-stage training.
CBDB-Net [64]	It employs a continuous drop block of features that regularizes training.	It considers just horizontal occlusions in the training data.
BAT-Net [15]	It proposes a novel bilinear attention block to improve feature learning.	It trains two coupled networks, which increase number of parameters and does not allow parallel training.
CDNet [41]	It employs a Neural Architecture Search space and strategy along with a local part description for feature learning.	It requires setting the search space and strategy parameters, which brings more hyperparameters for model optimization.
FIDI [76]	It proposes a new fine-grained aware loss function to highlight fine-grained features.	The proposed loss function brings two more hyper-parameters to tune.
Ours (DaliReID)	**It deals with long-range Face Recognition and Person Re-Identification. It does not rely on batch clustering, camera labels, attention models, or architecture searching.**	**It requires the training of two backbones, but it can be done in parallel.**

of the image. The bottom left bars follow the same idea as the top left bars. However, the feature vectors were obtained using the distortion-adaptive model (the same set of images are used for comparison) of our DaliID solution. We see that the DA features maintain more important high-magnitude feature distortion and have far fewer incorrect low-magnitude features introduced, as illustrated by the green arrows. The bars on the right are non-matched images, i.e. images from identities different from G02058, so we see that different features are activated for recognition across different levels of AT.

TABLE 7. Comparison to the state-of-the-art models in same-clothes Person Re-Identification setup. Bold and Blue indicate the best and second-best values. *CD-Net is not based on ResNet50, but the authors of that paper mostly compared to ResNet50-based models, so we leave it here for a fair comparison.

		Market		MSMT17	
Method	Venue	mAP	R1	mAP	R1
OSNet-based models					
OSNet [88]	ICCV19	84.9	94.8	52.9	78.7
DaliReID (OSNet)	**This work**	**87.2**	**95.0**	**59.5**	**82.6**
ResNet50-based models					
GCS [5]	CVPR18	81.6	93.5	-	-
SFT [46]	ICCV19	82.7	93.4	47.6	73.6
CBN [93]	ECCV20	83.6	94.3	-	-
STNReID [47]	TMM20	84.9	93.8	-	-
CBDB-Net [64]	TCSVT21	85.0	94.4	-	-
BAT-Net [15]	ICCV19	85.5	94.1	50.4	74.1
CDNet(*) [41]	CVPR21	86.0	**95.1**	**54.7**	**78.9**
FIDI [76]	TMM21	**86.8**	**94.5**	-	-
DaliReID (R50)	**This work**	**87.6**	**94.5**	**60.6**	**82.1**

VI. ABLATION STUDIES

A. FACE RECOGNITION

To demonstrate the improvements of the respective components of DaliFace, Table 9 shows an ablation with datasets representing three different evaluation scenarios: IJB-S for standard quality, CFP-LD for long distance, and TinyFace for low spatial resolution. It can be seen that aggressive distortion augmentations create a significant performance improvement in low-quality datasets CFP-LD and TinyFace; however, performance drops significantly on IJB-C, which is a dataset with relatively higher-quality images, showing that the use of distorted data without any further strategy might harm the performance in high-quality datasets. We argue that this is due to the inability of the model to effectively learn from distorted data, allowing it to dominate the optimization since the beginning of the training. After adding adaptive weighing and magnitude-based fusion, it can be seen that the final model (i.e., DaliFace) is the best performing. In Table 5, which also ablates different components of our model, a similar pattern can be observed. An additional ablation of our distortion

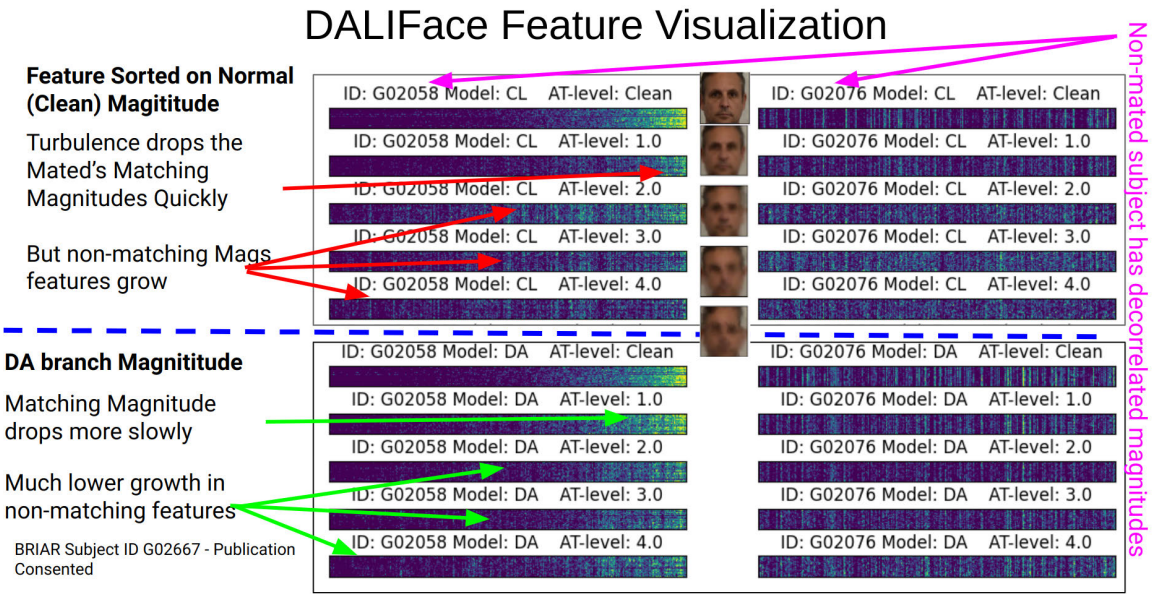


FIGURE 8. Visualization of features to highlight the learned invariance, best viewed in color. Each color bar has a feature in a column, and the bar has 25 rows where each is an image. The features are absolute values mapped to colors with blue small, green medium, and yellow high, with features sorted on the average absolute magnitude on the clean image. The top five bars are the features from the "clean" branch, with different simulated atmospheric turbulence (AT) levels added, with a sample image shown for each level. As highlighted by the red arrows, as AT increases, the originally large features lose magnitude while originally low features see an increase – both of which reduce the matching score. The right column of bars shows that a non-matching person is not well correlated. The bottom five bars are from the Distortion-Adapted (DA) trained branch. As the AT level increases, the DA features maintain more important high-magnitude feature distortion and have far fewer incorrect low-magnitude features introduced. These examples highlight that our DALI training approach produces greater invariance to distortion.

TABLE 8. Comparison to the state-of-the-art models in clothes-changing Person Re-Identification setup. Bold and Blue indicates the best second-best values. All methods, except ViT, consider ResNet50 as the backbone.

		DeepChange	
Method	Venue	mAP	R1
ReIDCaps [26]	TCSVT20	11.3	39.5
ViT [75]	ArXiv20	15.0	49.8
ViT (with Grayscale) [75]	ArXiv20	15.2	48.0
CAL [17]	CVPR22	**19.0**	**54.0**
DaliReID (R50)	**This work**	**21.9**	**60.8**

augmentation compared to Gaussian blur and downsampling is also provided in Table 11 and Section VI-C.

B. PERSON RE-IDENTIFICATION

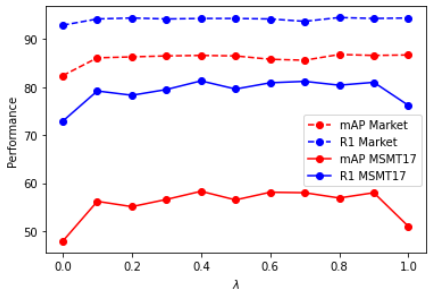
We perform a set of ablation studies over the PReID datasets to measure the impact of different components. The results are shown in Table 9. When we use distorted images as augmentations without our adaptive-weighting strategy (second line), we see a performance drop for both metrics in MSMT17 and DeepChange, and for mAP in Market when compared to the distortion-adaptive and DaliReID models. As the varied distortion images have the same importance for training, the model does not effectively learn from distorted data. For MSMT17 and DeepChange, the results are also worse than the baseline showing that just employing distortion as augmentations hinders model performance. We face the same performance dropping when we take out our proxy loss ($\lambda = 0$ in Eq. 11), showing it is an essential contribution (third line of Table 9). In contrast, just the distortion-adaptive backbone (fourth line) yields performance improvements for both MSMT17 and DeepChange and for mAP in Market, showing that it can learn a distortion-invariant feature space to some extent. Our final DaliReID model combines both clean and distortion-adaptive backbones (first and fourth lines), which leads to the best performance for MSMT17 (an increase of 5.2 and 3.6 for mAP and R1, respectively) and DeepChange (an increase of 1.7 and 2.2 for mAP and R1 respectively). This shows that DaliReID can effectively combine knowledge from both backbones.

In the future, we aim to apply our methodology in PReID datasets considering moving cameras (e.g., UAV) with distortion levels caused by distance and altitude [38].

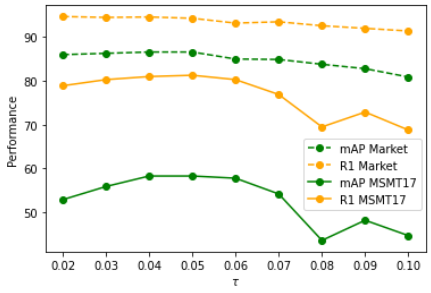
1) PREID PARAMETER ANALYSIS

There are two hyper-parameters on the final loss function (Eq. 11) for person re-identification: τ value to control the sharpening of the probability distribution in both terms and λ value to weight the contribution of $\mathcal{L}_{proxy}$ term. The impact of these parameters on the performance of the Distortion-Adaptive backbone is shown in Figure 9.

For λ in Figure 9a, we see stable performance for Market along different values after $\lambda = 0.1$, while for MSMT17 we see a peak at $\lambda = 0.4$, then a suitable decrease after this value. For both datasets, we see a performance drop for $\lambda = 0.0$ (no $\mathcal{L}_{proxy}$), showing the proxy-based loss term has a positive impact on training. In contrast, an equal contribution of both



(a) Impact of different λ values



(b) Impact of different τ values

FIGURE 9. Analysis of the impact of the parameters τ and λ on the final loss function considering the training of the distortion-adaptive backbone for PReID.

TABLE 9. Ablation study for both Face and PReID datasets. The respective first lines show the performance of the baseline (clean) model (trained without simulated distorted data). The second and third lines are for backbones trained with distortion as augmentation and our adaptive weighting strategy, respectively. For PReID, line 4 ablates the proxy loss (all other PReID lines contain $\mathcal{L}_{proxy}$), and the final line is the proposed DaliReID model. CFP-LD is reported as an average of the two protocols shown in Table 4.

Face Ablation	IJB-C	CFP-LD	TinyFace	Average
Baseline (θ_{cl})	97.38	75.22	72.18	81.59
Distortion Aug	96.91	78.16	74.11	83.06
Distortion-Adaptive (θ_{da})	96.92	78.37	**74.22**	83.17
DaliFace	**97.40**	**78.91**	73.98	**83.43**

	Market		MSMT17		DeepChange	
ReID Ablation	mAP	R1	mAP	R1	mAP	R1
Baseline (θ_{cl})	86.6	94.2	57.6	80.3	20.5	59.3
Distortion Aug	86.3	**94.7**	55.4	78.5	20.2	58.6
Distortion-Adaptive (θ_{da})	86.6	94.3	58.3	81.3	20.7	59.2
Distortion-Adaptive w/o $\mathcal{L}_{proxy}$	82.4	92.9	47.9	72.9	19.2	55.6
DaliReID	**87.6**	94.5	**60.6**	**82.1**	**21.9**	**60.8**

terms $\lambda = 1.0$ hurts the performance mainly for MSMT17. Since MSMT17 is more challenging, we select $\lambda = 0.4$ as the operational value. Further analysis of the impact of $\mathcal{L}_{proxy}$ is presented in Table 9.

The impact of τ is shown on Figure 9b. The performance drops when τ is lower than 0.04 for MSMT17 but has a stable behavior for Market, while values greater than 0.06 deteriorate the performance for both datasets. To achieve a good trade-off considering the dataset complexities, we choose $\tau = 0.05$.

2) IMPACT OF POOLING OPERATIONS IN EVALUATION (PREID)

As shown in Figure 2, the inference is performed by a weighted combination of the decisions from clean and distortion-adaptive backbones. The weights W_{clean} and $W_{distortion}$ are the maximum magnitudes of the feature vectors for each query and gallery image pair for each backbone. Among the different pooling strategies to get the final feature representation, we choose Global Average Pooling (GAP), Global Max Pooling (GMP), and a combination of both (GAP+GMP) to check the impact on final performance. The performances are reported in Table 10. Note that in this case, the pooling operations are **just to calculate the magnitudes**, since the final representation is always obtained by the element-wise sum of the output of the GAP and GMP layers for PReID.

TABLE 10. Ablation of the pooling operation to calculate the magnitudes for fusion in PReID.

	Market		MSMT17		DeepChange	
Setup	mAP	R1	mAP	R1	mAP	R1
GMP	87.6	94.4	60.5	82.1	21.9	60.7
GMP+GAP	87.6	94.4	60.6	82.1	21.8	60.8
DaliReID (GAP)	**87.6**	94.5	**60.6**	**82.1**	**21.9**	**60.8**

TABLE 11. A comparison between training augmentations. The distortion augmentation performs better than using Gaussian blur and down-sampling.

	IJB-C	CFP-LD	TinyFace
DS+GB	96.48	77.13	73.39
Distortion Aug (ours)	96.91	78.16	**74.11**
DaliFace (ours)	**97.40**	**78.91**	73.98

	Market		MSMT17		DeepChange	
	mAP	R1	mAP	R1	mAP	R1
DS+GB	78.0	91.2	44.7	69.5	16.2	51.5
Distortion Aug (ours)	86.3	**94.7**	55.4	78.5	20.2	58.6
DaliReID (ours)	**87.6**	94.5	**60.6**	**82.1**	**21.9**	**60.8**

DS+GB=down-sampling + Gaussian blur

We see among GAP, GMP, and GAP+GMP, we have a similar performance in evaluation, with a slight improvement for GAP. All of them have similar performances over the final result showing our proposed fusion strategy is robust to different pooling operations.

C. DISTORTION AUGMENTATION

A key contribution of the paper is the use of distortion augmentation inspired by atmospheric turbulence. This ablation shows that the gains are not simply from data augmentation. Table 11 shows a comparison to a combination of other similar augmentations used in computer vision: down-sampling and Gaussian blur. Gaussian blur and down-sampling are applied at equally challenging levels as distortion augmentation (as measured by the loss). In Table 11, it can be seen that distortion augmentation performs better than data augmentation Gaussian blur and down-sampling on both face recognition and person re-identification benchmarks, but not near as well as DALI.

D. FEATURE FUSION METHODS.

Our DaliID method uses a magnitude-weighted fusion of features from two backbones (see Figure 2). We also performed experiments with learned fusion layers. The magnitude-weighted fusion outperformed learned fusions as shown in Table 12.

TABLE 12. Experiments with three different learning methods to combine the feature vectors from the clean and distortion-adaptive backbones. Perhaps surprisingly, we get the best results without learning a final representation but rather performing magnitude-weighted fusion.

Fusion	IJB-C	CFP-LD	TinyFace
magnitude weighted fusion	**97.40**	**78.97**	**73.98**
linear layer	97.07	78.19	73.87
attention layer	97.20	78.26	73.84
transformer decoder	97.22	77.98	73.82

VII. CONCLUSION

In this work, DaliID is presented as a methodology for improving robustness to distortions common in real-world applications. The proposed components include distortion augmentation, distortion-adaptive weighting, and a parallel-backbone magnitude-weighted feature fusion. While face recognition and person re-identification have considerable differences, DaliID is shown to be applicable in both tasks with state-of-the-art performance on seven datasets. The proposed LD datasets, captured over the longest distance of any academic dataset, allow for further evaluation of realistic distortions.

In the future we aim to explore self-paced curriculum learning [72] to search for automatic curriculum criteria for the loss of weighting. Also, we intend to explore different backbone combinations (e.g., ResNet and Transformer [66]) to bring a diversity of learned knowledge by different models. Finally, our solution can be extended to further image recognition tasks operating in unconstrained scenarios where distortion can affect image features.

CODE AND DATA RELEASE

The code and datasets can be found in this https://github.com/Gabrielcb/DaliID repository.

ACKNOWLEDGMENT

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

REFERENCES

[1] X. An, J. Deng, J. Guo, Z. Feng, X. Zhu, J. Yang, and T. Liu, "Killing two birds with one stone: Efficient and robust training of face recognition CNNs by partial FC," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 4032–4041.

[2] D. Cao, X. Zhu, X. Huang, J. Guo, and Z. Lei, "Domain balancing: Face recognition on long-tailed domains," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5670–5678.

[3] J. Chang, Z. Lan, C. Cheng, and Y. Wei, "Data uncertainty learning in face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5709–5718.

[4] B. Chen, W. Deng, and J. Hu, "Mixed high-order attention network for person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 371–381.

[5] D. Chen, D. Xu, H. Li, N. Sebe, and X. Wang, "Group consistent similarity learning via deep CRF for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8649–8658.

[6] T. Chen, S. Ding, J. Xie, Y. Yuan, W. Chen, Y. Yang, Z. Ren, and Z. Wang, "ABD-net: Attentive but diverse person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8350–8360.

[7] Z. Cheng, Q. Dong, S. Gong, and X. Zhu, "Inter-task association critic for cross-resolution person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2602–2612.

[8] Z. Cheng, X. Zhu, and S. Gong, "Low-resolution face recognition," in *Proc. 14th Asian Conf. Comput. Vis. (ACCV)*. Perth, WA, Australia: Springer, Aug. 2018, pp. 605–621.

[9] J. Deng, J. Guo, T. Liu, M. Gong, and S. Zafeiriou, "Sub-center ArcFace: Boosting face recognition by large-scale noisy web faces," in *Proc. Eur. Conf. Comput. Vis.* Glasgow, U.K.: Springer, 2020, pp. 741–757.

[10] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou, "RetinaFace: Single-shot multi-level face localisation in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5202–5211.

[11] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4685–4694.

[12] J. Deng, J. Guo, J. Yang, A. Lattas, and S. Zafeiriou, "Variational prototype learning for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 11901–11910.

[13] A. R. Dhamija, M. Günther, and T. Boult, "Reducing network agnostophobia," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–12.

[14] H. Dong, Y. Yang, X. Sun, L. Zhang, and L. Fang, "Cascaded attention-guided multi-granularity feature learning for person re-identification," *Mach. Vis. Appl.*, vol. 34, no. 1, pp. 1–16, Jan. 2023.

[15] P. Fang, J. Zhou, S. Roy, L. Petersson, and M. Harandi, "Bilinear attention networks for person retrieval," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8029–8038.

[16] Y. Ge, D. Chen, and H. Li, "Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification," 2020, *arXiv:2001.01526*.

[17] X. Gu, H. Chang, B. Ma, S. Bai, S. Shan, and X. Chen, "Clothes-changing person re-identification with RGB modality only," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 1050–1059.

[18] M. Günther, P. Hu, C. Herrmann, C. H. Chan, M. Jiang, S. Yang, A. R. Dhamija, D. Ramanan, J. Beyerer, J. Kittler, M. A. Jazaery, M. I. Nouyed, G. Guo, C. Stankiewicz, and T. E. Boult, "Unconstrained face detection and open-set face recognition challenge," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Oct. 2017, pp. 697–706.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[20] S. He, H. Luo, P. Wang, F. Wang, H. Li, and W. Jiang, "TransReID: Transformer-based object re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 14993–15002.

[21] A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," 2017, *arXiv:1703.07737*.

[22] R. Hou, B. Ma, H. Chang, X. Gu, S. Shan, and X. Chen, "Interaction-and-aggregation network for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9309–9318.

[23] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," in *Proc. Workshop Faces in'Real-Life'Images: Detection, Alignment, Recognit.*, 2008, pp. 1–12.

[24] Y. Huang, S. Lian, and H. Hu, "AVPL: Augmented visual perception learning for person re-identification and beyond," *Pattern Recognit.*, vol. 129, Sep. 2022, Art. no. 108736.

[25] Y. Huang, Y. Wang, Y. Tai, X. Liu, P. Shen, S. Li, J. Li, and F. Huang, "CurricularFace: Adaptive curriculum learning loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5900–5909.

[26] Y. Huang, J. Xu, Q. Wu, Y. Zhong, P. Zhang, and Z. Zhang, "Beyond scalar neuron: Adopting vector-neuron capsules for long-term person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 10, pp. 3459–3471, Oct. 2020.

[27] *Biometric Recognition and Identification At Altitude and Range (Briar) Program*, IARPA Broad Agency Announcement: IARPA-BAA-20-04, Bethesda, MD, USA, 2020.

[28] M. Jia, X. Cheng, S. Lu, and J. Zhang, "Learning disentangled representation implicitly via transformer for occluded person re-identification," *IEEE Trans. Multimedia*, vol. 25, pp. 1294–1305, 2023.

[29] J. Jiao, W.-S. Zheng, A. Wu, X. Zhu, and S. Gong, "Deep low-resolution person re-identification," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, 2018, pp. 6967–6974.

[30] X. Jin, C. Lan, W. Zeng, G. Wei, and Z. Chen, "Semantics-aligned representation learning for person re-identification," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 11173–11180.

[31] M. M. Kalayeh, E. Basaran, M. Gökmen, M. E. Kamasak, and M. Shah, "Human semantic parsing for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1062–1071.

[32] N. D. Kalka, B. Maze, J. A. Duncan, K. O'Connor, S. Elliott, K. Hebert, J. Bryan, and A. K. Jain, "IJB–S: IARPA Janus surveillance video benchmark," in *Proc. IEEE 9th Int. Conf. Biometrics Theory, Appl. Syst. (BTAS)*, Oct. 2018, pp. 1–9.

[33] M. Kim, A. K. Jain, and X. Liu, "AdaFace: Quality adaptive margin for face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*. Jun. 2022, pp. 18729–18738.

[34] M. Kim, F. Liu, A. Jain, and X. Liu, "Cluster and aggregate: Face recognition with large probe set," in *Proc. Adv. Neural Inf. Process. Syst.*, 2022, pp. 1–13.

[35] Y. Kim, W. Park, M.-C. Roh, and J. Shin, "GroupFace: Learning latent groups and constructing group-based representations for face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5620–5629.

[36] Y. Kim, W. Park, and J. Shin, "BroadFace: Looking at tens of thousands of people at once for face recognition," in *Proc. Comput. Vis. ECCV*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham, Switzerland: Springer, 2020, pp. 536–552.

[37] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[38] S. V. A. Kumar, E. Yaghoubi, A. Das, B. S. Harish, and H. Proença, "The P-DESTRE: A fully annotated dataset for pedestrian detection, tracking, and short/long-term re-identification from aerial devices," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1696–1708, 2021.

[39] C. P. Lau, H. Souri, and R. Chellappa, "ATFaceGAN: Single face image restoration and recognition from atmospheric turbulence," in *Proc. 15th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Nov. 2020, pp. 32–39.

[40] B. Li, T. Xi, G. Zhang, H. Feng, J. Han, J. Liu, E. Ding, and W. Liu, "Dynamic class queue for large scale face recognition in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 3762–3771.

[41] H. Li, G. Wu, and W.-S. Zheng, "Combined depth space based architecture search for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 6725–6734.

[42] P. Li, B. Wang, and L. Zhang, "Virtual fully-connected layer: Training a large-scale face recognition dataset with limited computational resources," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13310–13319.

[43] W. Li, T. Guo, P. Li, B. Chen, B. Wang, W. Zuo, and L. Zhang, "VirFace: Enhancing face recognition via unlabeled shallow data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14724–14733.

[44] Y. Li, J. He, T. Zhang, X. Liu, Y. Zhang, and F. Wu, "Diverse part discovery: Occluded person re-identification with part-aware transformer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2897–2906.

[45] F. Liu, M. Kim, A. Jain, and X. Liu, "Controllable and guided face synthesis for unconstrained face recognition," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Tel Aviv, Israel. Cham, Switzerland: Springer, 2022, pp. 701–719.

[46] C. Luo, Y. Chen, N. Wang, and Z.-X. Zhang, "Spectral feature transformation for person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4975–4984.

[47] H. Luo, W. Jiang, X. Fan, and C. Zhang, "STNReID: Deep convolutional networks with pairwise spatial transformer networks for partial person re-identification," *IEEE Trans. Multimedia*, vol. 22, no. 11, pp. 2905–2913, Nov. 2020.

[48] H. Luo, W. Jiang, Y. Gu, F. Liu, X. Liao, S. Lai, and J. Gu, "A strong baseline and batch normalization neck for deep person re-identification," *IEEE Trans. Multimedia*, vol. 22, no. 10, pp. 2597–2609, Oct. 2020.

[49] Z. Mao, N. Chimitt, and S. H. Chan, "Accelerating atmospheric turbulence simulation via learned Phase-to-Space transform," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 14739–14748.

[50] B. Maze, J. Adams, J. A. Duncan, N. Kalka, T. Miller, C. Otto, A. K. Jain, W. T. Niggel, J. Anderson, J. Cheney, and P. Grother, "IARPA Janus benchmark–C: Face dataset and protocol," in *Proc. Int. Conf. Biometrics (ICB)*, Feb. 2018, pp. 158–165.

[51] Q. Meng, S. Zhao, Z. Huang, and F. Zhou, "MagFace: A universal representation for face recognition and quality assessment," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14220–14229.

[52] P. Micikevicius, S. Narang, J. Alben, G. Diamos, E. Elsen, D. Garcia, B. Ginsburg, M. Houston, O. Kuchaiev, G. Venkatesh, and H. Wu, "Mixed precision training," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–12.

[53] S. Moschoglou, A. Papaioannou, C. Sagonas, J. Deng, I. Kotsia, and S. Zafeiriou, "AgeDB: The first manually collected, in-the-wild age database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1997–2005.

[54] A. Munir, C. Lyu, B. Goossens, W. Philips, and C. Micheloni, "Resolution based feature distillation for cross resolution person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 281–289.

[55] Y. Rao, G. Chen, J. Lu, and J. Zhou, "Counterfactual attention learning for fine-grained visual categorization and re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 1005–1014.

[56] W. Robbins and T. Boult, "On the effect of atmospheric turbulence in the feature space of deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 1617–1625.

[57] S. Sengupta, J.-C. Chen, C. Castillo, V. M. Patel, R. Chellappa, and D. W. Jacobs, "Frontal to profile face verification in the wild," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2016, pp. 1–9.

[58] H. Sheng, Y. Zheng, W. Ke, D. Yu, X. Cheng, W. Lyu, and Z. Xiong, "Mining hard samples globally and efficiently for person reidentification," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 9611–9622, Oct. 2020.

[59] Y. Shi and A. Jain, "Probabilistic face embeddings," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6901–6910.

[60] Y. Shi, X. Yu, K. Sohn, M. Chandraker, and A. K. Jain, "Towards universal representation learning for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 6816–6825.

[61] V. Somers, C. D. Vleeschouwer, and A. Alahi, "Body part-based representation learning for occluded person re-identification," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 1613–1623.

[62] Y. Sun, C. Cheng, Y. Zhang, C. Zhang, L. Zheng, Z. Wang, and Y. Wei, "Circle loss: A unified perspective of pair similarity optimization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 6397–6406.

[63] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, "Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 480–496.

[64] H. Tan, X. Liu, Y. Bian, H. Wang, and B. Yin, "Incomplete descriptor mining with elastic loss for person re-identification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 1, pp. 160–171, Jan. 2022.

[65] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–16.

[66] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–11.

[67] G. Wang, S. Yang, H. Liu, Z. Wang, Y. Yang, S. Wang, G. Yu, E. Zhou, and J. Sun, "High-order information matters: Learning relation and topology for occluded person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 6448–6457.

[68] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou, "Learning discriminative features with multiple granularities for person re-identification," in *Proc. 26th ACM Int. Conf. Multimedia*, Oct. 2018, pp. 274–282.

[69] H. Wang, J. Shen, Y. Liu, Y. Gao, and E. Gavves, "NFormer: Robust person re-identification with neighbor transformer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 7287–7297.

[70] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, "CosFace: Large margin cosine loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5265–5274.

[71] M. Wang, B. Lai, J. Huang, X. Gong, and X.-S. Hua, "Camera-aware proxies for unsupervised person re-identification," in *Proc. AAAI Conf. Artif. Intell.*, May 2021, vol. 35, no. 4, pp. 2764–2772.

[72] X. Wang, Y. Chen, and W. Zhu, "A survey on curriculum learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 4555–4576, Sep. 2022.

[73] Z. Wang, F. Zhu, S. Tang, R. Zhao, L. He, and J. Song, "Feature erasing and diffusion network for occluded person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 4744–4753.

[74] L. Wei, S. Zhang, W. Gao, and Q. Tian, "Person transfer GAN to bridge domain gap for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 79–88.

[75] P. Xu and X. Zhu, "DeepChange: A large long-term person re-identification benchmark with clothes change," 2021, *arXiv:2105.14685*.

[76] C. Yan, G. Pang, X. Bai, C. Liu, X. Ning, L. Gu, and J. Zhou, "Beyond triplet loss: Person re-identification with fine-grained difference-aware pairwise loss," *IEEE Trans. Multimedia*, vol. 24, pp. 1665–1677, 2022.

[77] R. Yasarla and V. M. Patel, "Learning to restore images degraded by atmospheric turbulence using uncertainty," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2021, pp. 1694–1698.

[78] R. Yasarla and V. M. Patel, "CNN-based restoration of a single face image degraded by atmospheric turbulence," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 4, no. 2, pp. 222–233, Apr. 2022.

[79] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. C. H. Hoi, "Deep learning for person re-identification: A survey and outlook," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 6, pp. 2872–2893, Jun. 2022.

[80] Y. Zhai, Q. Ye, S. Lu, M. Jia, R. Ji, and Y. Tian, "Multiple expert brainstorming for domain adaptive person re-identification," 2020, *arXiv:2007.01546*.

[81] G. Zhang, Y. Ge, Z. Dong, H. Wang, Y. Zheng, and S. Chen, "Deep high-resolution representation learning for cross-resolution person re-identification," *IEEE Trans. Image Process.*, vol. 30, pp. 8913–8925, 2021.

[82] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016.

[83] Z. Zhang, C. Lan, W. Zeng, X. Jin, and Z. Chen, "Relation-aware global attention for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3183–3192.

[84] Z. Zhang, H. Zhang, and S. Liu, "Person re-identification using heterogeneous local graph attention networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 12131–12140.

[85] S. Zhao, C. Gao, J. Zhang, H. Cheng, C. Han, X. Jiang, X. Guo, W.-S. Zheng, N. Sang, and X. Sun, "Do not disturb me: Person re-identification under the interference of other pedestrians," in *Proc. 16th Eur. Conf. Comput. Vis. (ECCV)*. Glasgow, U.K.: Springer, Aug. 2020, pp. 647–663.

[86] F. Zheng, C. Deng, X. Sun, X. Jiang, X. Guo, Z. Yu, F. Huang, and R. Ji, "Pyramidal person re-IDentification via multi-loss dynamic training," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8506–8514.

[87] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1116–1124.

[88] K. Zhou, Y. Yang, A. Cavallaro, and T. Xiang, "Omni-scale feature learning for person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3701–3711.

[89] H. Zhu, W. Ke, D. Li, J. Liu, L. Tian, and Y. Shan, "Dual cross-attention learning for fine-grained visual categorization and object re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 4682–4692.

[90] K. Zhu, H. Guo, S. Liu, J. Wang, and M. Tang, "Learning semantics-consistent stripes with self-refinement for person re-identification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 11, pp. 8531–8542, Nov. 2022.

[91] K. Zhu, H. Guo, Z. Liu, M. Tang, and J. Wang, "Identity-guided human semantic parsing for person re-identification," in *Proc. 16th Eur. Conf. Comput. Vis (ECCV)*. Glasgow, U.K.: Springer, Aug. 2020, pp. 346–363.

[92] Z. Zhu, G. Huang, J. Deng, Y. Ye, J. Huang, X. Chen, J. Zhu, T. Yang, J. Lu, D. Du, and J. Zhou, "WebFace260M: A benchmark unveiling the power of million-scale deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10487–10497.

[93] Z. Zhuang, L. Wei, L. Xie, T. Zhang, H. Zhang, H. Wu, H. Ai, and Q. Tian, "Rethinking the distribution gap of person re-identification with camera-based batch normalization," in *Proc. 16th Eur. Conf. Comput. Vis. (ECCV)*. Glasgow, U.K.: Springer, Aug. 2020, pp. 140–157.

WES ROBBINS received the Bachelor of Science degree in computer science from Montana State University and the M.S. degree in computer science from the Vision and Security Technology Laboratory, University of Colorado at Colorado Springs, under the supervision of Prof. Terrance Boult. He is currently pursuing the Ph.D. degree in electrical and computer engineering with The University of Texas at Austin. He received the Outstanding Graduate Student Award.

GABRIEL BERTOCCO received the B.Sc. degree in computing engineering with the Artificial Intelligence Laboratory (**Recod.ai**), Institute of Computing, University of Campinas, Brazil, in 2019, where he is currently pursuing the Ph.D. degree in computer science, with a focus on digital forensics and machine learning. From 2022 to 2023, he was a Visiting Scholar with the Vision and Security Laboratory (VAST), University of Colorado at Colorado Springs, USA. He has publications in top venues, such as IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, IEEE Security and Privacy, and the IEEE International Joint Conference on Biometrics. His research interests include machine learning, computer vision, digital forensics, and biometrics. nn

TERRANCE E. BOULT (Fellow, IEEE) received the B.S. degree in applied mathematics, the M.S. degree in computer science, and the Ph.D. degree in computer science from Columbia University, New York, NY, USA, in 1983, 1984, and 1986, respectively. He is currently a Distinguished Professor and the El Pomar Endowed Professor of Innovation and Security with the University of Colorado at Colorado Springs, Colorado Springs, CO, USA, a Serial Entrepreneur, and an Internationally Acknowledged Researcher in machine learning, computer vision, biometrics, and cybersecurity. He has issued 15 patents and more than 400 articles. He spent six years as an Assistant Professor and two years as an Associate Professor with the CS Department, Columbia University. He moved from Columbia to Lehigh, Bethlehem, PA, USA, working there from 1994 to 2003. At Lehigh, he was an Endowed Professor and eventually founded Lehigh's CS Department. In 2003, he joined UCCS as an El Pomar Professor. He is a member of the IEEE Golden Core and has been an IEEE Distinguished Lecturer. He has won multiple teaching awards, research/innovation awards, best paper awards, best reviewer awards, and IEEE service awards. He was the Co-Founder of the Computer Vision Foundation and was very active in organizing/managing computer vision conferences. On the education side, he is the Founder, a Primary Architect, and the Co-Director of the world's first and only Bachelor of Innovation Family of Degrees at UCCS. This awarding family of degrees combines a core of innovation and entrepreneurship with a significant multiyear "team emphasis" and all the rigor of bachelor degrees in their fields, serving more than 600 students per year across 22 different majors spanning four colleges.