Degradation Loss: A Self-supervised Loss Function for Low-image-quality Face Recognition

Wes Robbins* Terrance E. Boult*
Vision and Security Technology Lab, University of Colorado, Colorado Springs
wrobbins@uccs.edu, tboult@vast.uccs.edu

Abstract

Challenging imaging conditions such as atmospheric turbulence decrease face recognition performance. Prior work has developed image-restoration methods to mitigate the effects of atmospherics. In contrast, this work directly optimizes for recognition performance in atmospheric conditions. First, we propose simulated atmospheric turbulence as an augmentation for training face recognition models, which is found to be useful for multiple low-quality image conditions. Additionally, a loss term—called degradation loss—and a multi-stage training procedure is proposed for further increasing performance. Degradation loss improves performance by self-supervising features to be directionally invariant to atmospherics and express quality-aware magnitudes. The advantages of trajectory loss and our end-to-end approach are empirically demonstrated with experiments on simulated atmospheric and low-resolution test sets, including SoTA performance on the TinyFace dataset.

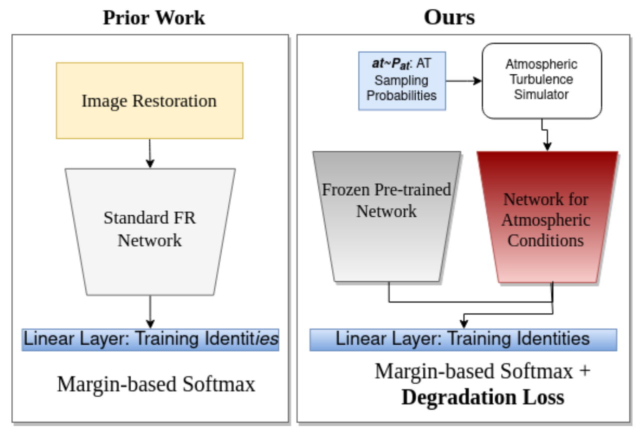

Figure 1. This work directly optimizes for the recognition problem in atmospherics. During training, an atmospheric turbulence simulator augments the training data with different levels of perturbation (based on sampling distribution $at \sim \mathcal{P}_{at}$). Our proposed degradation loss leverages a frozen pretrained network to self-supervise deep features to be "equivariant" to the effects of atmospheric turbulence. The training procedure and loss function are described in detail in Section 3.

1. Introduction

Face Recognition (FR) has been a well-studied problem in computer vision. Over the last several years, improvements surrounding deep learning have lead to high-performing FR models [5, 39]. Despite these improvements, handling low-quality imaging conditions, such as atmospheric turbulence, remains an open problem. Applications such as surveillance that use long-distance imaging are affected by atmospheric turbulence. Thus, it is important that models are robust to such conditions to mitigate inaccurate classifications.

Atmospheric turbulence is caused by fluctuating temperatures and turbulent airflow in the atmosphere. Light waves propagating through such conditions are refracted before before reaching an image sensor, which leads to a variable point spread function. Figure 2 shows the point spread function before and after atmospherics. Formally, the effect on an image of atmospheric turbulence can be described with the following equation [15, 36, 20]:

$$I_k = D_k(H_k(I)) + n_k \tag{1}$$

where k denotes time step, I the unobserved clean image, and I_k the observed image affected by atmospherics. D_k is a deformation operator, H_k is a blur operator, and n_k is signal noise. The terms in Equation 1 are non-linear and random, and in practice modeling the effect is highly ill-posed [27]. As I_k is of lower quality than I, com-

*This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via [2022-21102100003]. 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.

puter vision applications perform worse on perturbed image I_k. Furthermore, the deformations D_k are particularly challenging for face recognition because identifiable features such as eyes or mouth can be deformed or slightly re-positioned, which has an outsized effect on a neural network classifier [26].

To address the problem of face recognition in atmospheric conditions, this work looks at directly optimizing an end-to-end FR model for such conditions. This is in contrast to prior methods in the literature have approached mitigating atmospheric turbulence for face recognition with image-restoration methods [36, 35, 14, 15]. Figure 1 shows a high-level overview of our methodology. In particular, this this work proposes simulated atmospheric turbulence as a training augmentation, a novel loss term, and a multi-stage training procedure.

The methodology of this work is driven by simulating atmospheric turbulence on training data. A GPU-accelerated atmospheric turbulence simulator [19] is used to allow simulation at a large scale with a wide range of turbulence levels. With the use of simulated training data, we achieve state-of-the-art performance on atmospheric and low-resolution test data. This is the first work to use simulated atmospheric turbulence as a training augmentation for face recognition, which is experimentally demonstrated to be be useful for face recognition in low-quality regimes. Details on atmospheric augmentations can be found in Section 3.1. In addition to training with simulated atmospheric data, a novel loss term called degradation loss is proposed.

The motivation for degradation loss is to learn a feature space that is "equivariant" under atmospheric transformations. In particular, an atmospheric feature should have maximum cosine similarity with the original feature; and the magnitude of the feature should decrease monotonically as the strength of the atmospheric perturbation increases. The first property is desirable because maximum cosine similarity necessarily results in robust performance under atmospheric transformations, as cosine similarity is the distance metric used for open-set evaluation. The second property is desirable because magnitude can be used for quality assessment, where lower magnitude indicates lower quality [21, 13, 6]. Degradation loss simultaneously optimizes for the these two properties in a self-supervised manner by penalizing an atmospheric feature if it is not between the clean feature and the origin (within some margin). Degradation loss is depicted in Figure 3 and described formally in Section 3.2.

In order to train with degradation loss, GPU memory needs to simultaneously hold a clean feature and an atmospheric feature from the same image. The most straight forward way of implementing this is passing a clean image and the copied artificially degraded image through the network in the training loop. However, we find several implementations of this approach to hurt training efficiency and test performance. To overcome this, a two-stage training procedure is designed. The first stage pretrains a model which is then frozen to contribute to finetuning a final model with degradation loss in the second stage. This approach is more efficient and results in our best performing model on atmospherics. The implementation of our two-stage training procedure can be seen later in Figure 4 and is described in detail in Section 3.3.

In Section 4, we perform experiments on a wide range of simulated atmospherics and on low-resolution data. Comparisons are made with atmospheric image-restoration and prior face recognition methods. A principle result is that simulated atmospheric turbulence as a training augmentation leads to state-of-the-art performance on both simulated atmospheric turbulence and low-resolution test sets. It is also shown that using degradation loss results further improved performance on simulated atmospherics by creating features that are more robust to atmospherics. The paper is concluded in Section 5 with a discussion on the proposed methods and paths of future work.

In summary, this work makes the following contributions:

- Demonstrates simulated atmospheric turbulence as a useful training augmentation for face recognition.
- Proposes self-supervised degradation loss and a corresponding two-stage training procedure for improved performance in atmospheric conditions.
- Provide experimentation on wide range of simulated atmospheric conditions and low-resolution datasets and demonstrate State-of-the-art (SoTA) on TinyFace dataset.

Figure 2. **Left** The effect of atmospherics on a grid of point sources (inverted for visibility). The effect is spatially and temporally variable. **Right** A sample before and after simulated atmospheric turbulence. This work aims to create deep features that are robust to atmospheric transformations.

2. Related Work

The optical effects of atmospheric turbulence have a rich history of being studied in optics and physics [27]. One of the most effective approaches for mitigating atmospherics is adaptive optics [32, 28]. However, this method requires large and expensive equipment. Thus, there has been significant interest in mitigating atmospherics with image-processing approaches. This has been approached

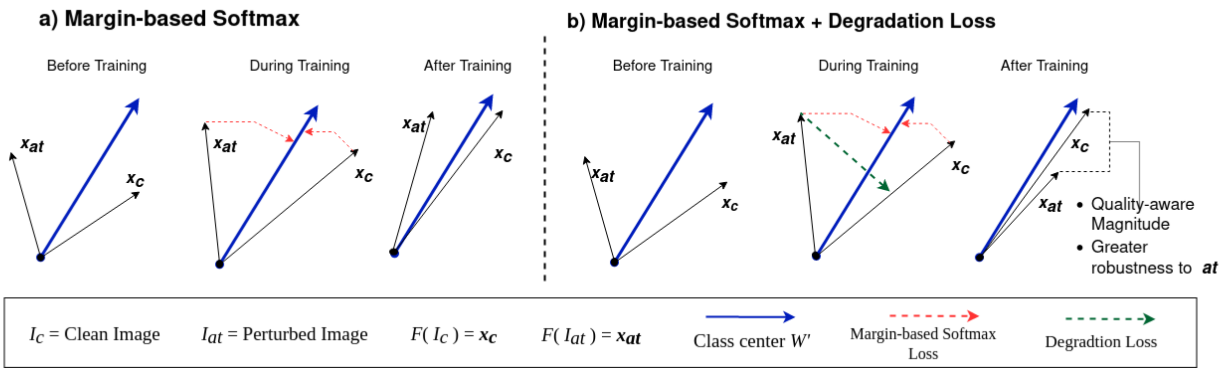

Figure 3. **Margin-based softmax + degradation loss in feature space.** Margin-based softmax losses such as ArcFace [5] and CosFace [34] supervise features to have minimum distance from a class center, where the class center is represented as the the ith entry of final fully connected layer W. We propose *degradation loss*, which offers a stronger inductive bias for robustness to degradations such as atmospheric turbulence. In deep feature space, during training, the feature vectors of artificially degraded samples are supervised to 1) have high cosine similarity with the clean feature vector and 2) be of lower magnitude.

with classical image-processing [38, 1, 16] as well as neural network-based approaches [20, 7, 17, 35].

Most similar to this work, several works have looked at image-restoration approaches as an upstream step to face recognition. Lau et al. proposed ATFaceGan, a method that uses two generators to restore deformation and blurring respectively[15]. In [35], they developed a learning-based approach to restore images by estimating uncertainty maps which are prior for a combination of both blur and geometric distortions in turbulence degraded images. Then, the estimated uncertainty maps are used to guide the network to obtain the restored image. Another proposed method uses facial landmarks to guide restoration [14]. In [36], a CNN-based network called TDRN is developed for image-restoration. Nair et al. [25] provide a study on different atmospheric simulators on face datasets. In [26], a baseline model is evaluated on simulated atmospheric turbulence and several challenges of face recognition under atmospheric turbulence are enumerated. Other works have provided studies mitigating atmospheric turbulence outside of face recognition [33, 8, 7, 20]. This work is different from the aforementioned works, as we propose a method for directly optimizing for face recognition rather than image-restoration.

Prior work on low-quality and low-resolution face recognition is also related to this paper. Cheng et al. propose an end-to-end network for super-resolution and identification called CSRI [4]. In [12], a distillation is used for resolution-invariant recognition. Other prior work has proposed probabilistic methods for estimating quality and uncertainty such as Probabilistic Face Embeddings [30], Spherical Face [18], and data uncertainty learning [3]. A multi-scale network for low-resolution recognition is proposed in [23]. [37] use a data synthesis method for training. A method for low-resolution recognition using multidimensional scaling is proposed in [2]. AdaFace [13] is a quality adaptive margin which is the current state-of-the-art on low-quality. Different from these works, this paper focuses heavily on atmospheric conditions.

3. Approach

The methodology proposed in this work consists of two parts. The first is atmospheric turbulence simulation as an augmentation, which is described in Section 3.1. The second the is degradation loss, which directly supervises feature vectors to be equivariant under atmospheric turbulence. Degradation loss is described in detail in Section 3.2. Given degradation loss, there are several implementation details that need to be considered for an efficient and effective implementation. To accomplish this, we implement a two-stage training procedure, which results in our highest-performing model. The two-stage training procedure is described in Section 3.3.

3.1. Simulated Atmospheric Turbulence as a Training Augmentation

Image augmentations allow for improved generalization by adding variation to training data. There is a vast space of augmentations that can be performed on an image, many of which have been remarkably successful for computer vision tasks. In this work, we adopt atmospheric turbulence as an augmentation, where an image is perturbed by atmospheric simulation software before being fed to a network during training. Simulated atmospheric turbulence is of practical interest because it is not tractable to collect real labeled data through atmospherics at a scale suitable for training a deep learning model. Experimentally, we find training with simulated atmospherics yields state-of-the-art performance on simulated atmospheric and low-resolution test sets. In the remainder of the work, training with At-

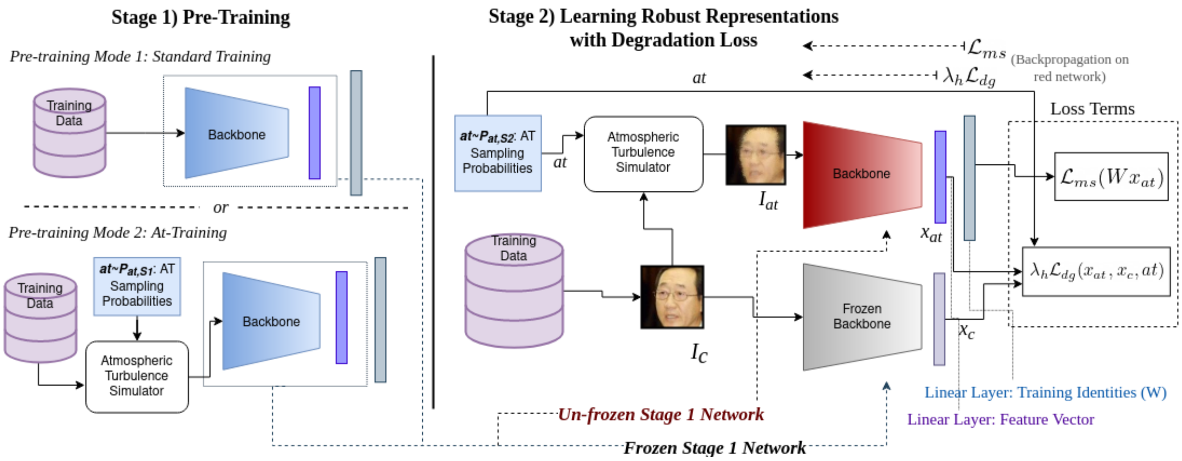

Figure 4. **A two-stage procedure for training with degradation loss.** In Stage 1, a model pretrained with standard training or AT-Training. For AT-Training, a batch of N images are extracted from the training set and passed to the atmospheric turbulence simulator. An atmospheric turbulence strength at is then sampled from probability distribution $at{\sim}P_{at}$ for each of N images. The simulator generates degraded copies of each image with strength at. In the Stage 2, two copies of the Stage 1 network are instantiated, one of which is frozen. The unfrozen network is trained with a margin-based softmax and degradation loss. The features from the frozen network are leveraged for calculating the degradation loss.

mospheric Turbulence (AT) as an augmentation is referred to as **AT-Training**. While adding simulated atmospheric turbulence to the training pipeline is simple to implement, there are several open questions for optimal training efficiency and test performance.

We propose the following question, which this work begins to answer: *At what frequency and strength should atmospheric perturbations be applied during training?* To formulate this, let at be the strength of the perturbation, which is sampled from a probability distribution $at{\sim}\mathcal{P}_{at}$ with support $[0, k]$. Note that the value 0 can be drawn from $\mathcal{P}_{at}$, which in practice is equivalent to a clean sample not passed to the simulator. Thus, the probability a sample is not passed to the simulator is represented as $\mathcal{P}(at{=}0)$. k is the maximum perturbation strength that can be drawn from $\mathcal{P}_{at}$.

In this work, we create a space of sampling distributions by selecting several values of k and several fixed values of $\mathcal{P}(at{=}0)$. For $\mathcal{P}(at \neq 0)$, atmospheric strengths are sampled uniformly from discrete levels within $(0, k]$. Discrete values of at are sampled because the atmospheric simulation software requires a pre-calculated tilt-matrix for each level of at. Therefore, it is intractable to sample continuously from $(0, k]$[1]. In Section 4.4, experiments are run to search over a space of a range of values of k and $\mathcal{P}(at{=}0)$. While we ascertain the best model from this search space, it is not exhaustive. A potential path of future work is further optimizing the sampling distribution with respect to a given test distribution. In the following section the degradation loss is discussed, which uses samples from $\mathcal{P}_{at}$.

3.2. Degradation Loss Term

Degradation loss is a self-supervised loss term to supplement margin-based softmax for face recognition training. Degradation loss optimizes for two properties in feature space. The first is for features vectors of a clean sample and an artificially degraded copy of that sample (e.g., with simulated atmospherics) to have maximum cosine similarity. High cosine similarity necessarily leads to robust performance under a given degradation, because cosine similarity is the distance metric used for evaluation. The second motivation for degradation loss is that features from clean samples should have greater magnitudes than the features of artificially degraded samples. This is based on the observations from prior work that lower quality samples have lower magnitudes [21, 13, 6]. This is a beneficial property because magnitude can be used as a quality-assessment measure [21, 13]. Compared to a standard margin-based softmax loss, degradation loss is a stronger level of supervision for robustness under degraded samples. For example, in the case of AT-Training, ArcFace [5] minimizes distance between 1) class-center to clean-sample and 2) class-center to atmospheric-sample. Comparatively, degradation loss directly minimizes the distance between the clean-sample and the atmospheric-sample. A comparison of margin-based softmax and degradation loss is depicted in Figure 3. Conceptually, the inductive bias implied by degradation loss is that deep features should be invariant to atmospherics.

Let I_c be the clean image and I_{at} be a degraded copy of I_c that has been perturbed by the atmospheric simulator,

[1] The tilt matrices have a small memory footprint of <1MB. Multiple tilt matrices can be pre-calculated.

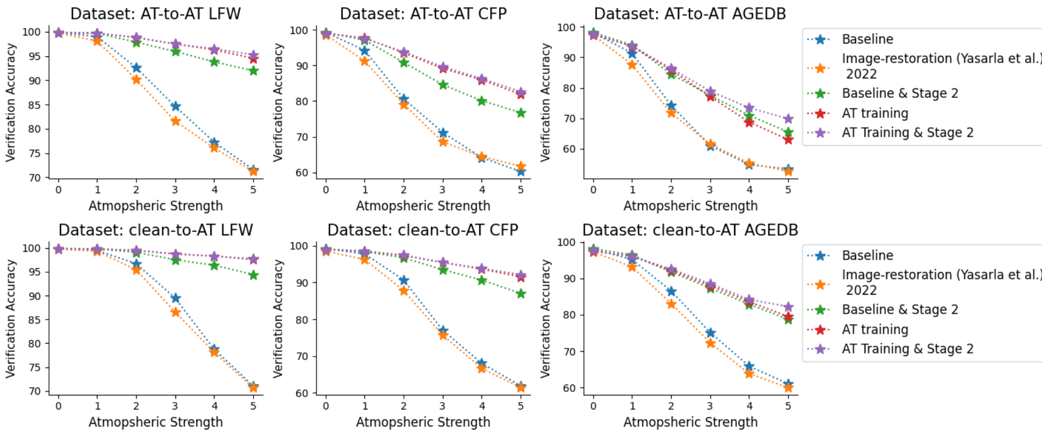

Figure 5. 1:1 Verification results on LFW [9], AgeDB [24], and CFP [29] for simulated atmospheric turbulence levels 0.0–5.0 [26, 19]. The first row shows atmospheric-to-atmospheric pairs and the second row show clean-to-atmospheric pairs. It is found that the pretrained atmospheric image-restoration model does not improve upon the baseline. Adding AT-Training (shown in red) significantly improves upon baseline and the model trained with AT-Training. and AT-Training followed by degradation loss is Stage 2 (shown in purple) is the best performing model.

where at is strength of the atmospheric turbulence. Higher at implies higher levels of atmospheric turbulence and thus a more degraded image. I_c and I_{at} are passed through the backbone network $\mathcal{F}$:

$$x_c = \mathcal{F}(I_c), x_{at} = \mathcal{F}(I_{at}) \tag{2}$$

which yields feature vectors x_c and x_{at}. Then, x_c is multiplied by a scalar $\sigma(at)$ where $l \leq \sigma(at) \leq 1$ and l is a hyperparameter. The result is a down-scaled feature vector $\hat{x}_c$:

$$\hat{x}_c = \sigma(at)x_c. \tag{3}$$

The function $\sigma(at)$ is chosen to be a sigmoid function fit to the accuracy curve plotted along increasing levels of at, as shown in Figure 3.2. Once $\hat{x}_c$ is obtained, the squared l_2 distance between $\hat{x}_c$ and x_{at} is calculated:

$$\mathcal{L}_d = ||x_{at} - \hat{x}_c||_2^2. \tag{4}$$

Given that $\hat{x}_c$ is colinear in the the same direction as x_c, $\mathcal{L}_d$ is optimizing x_{at} for maximum cosine similarity with x_c *and* for $||x_{at}|| < ||x_c||$, because $||\hat{x}_c|| < ||x_c||$ by construction of $\hat{x}_c$. $\mathcal{L}_d$ is obtained in a self-supervised manner because x_{at} is obtained following an augmentation of I_c, and at is randomly sampled from $\mathcal{P}_{at}$. Additionally, as a measure to prevent overfitting to highly noisy samples, a regularization margin r is subtracted from $\mathcal{L}_d$. Regularization margin r is calculated as

$$r(x_c, at) = (||x_c||_2 - ||\hat{x}_c||_2)\sin\theta_r \tag{5}$$

where θ_r is a hyperparameter. Margin r is monotonically increasing with at. The effect margin r is that high-atmospheric samples are penalized less severely for not falling exactly on the path between the clean feature and the origin. The margin is only a function of the clean feature and at—not the atmospheric feature. Finally, for each sample, the margin is subtracted from distance $\mathcal{L}_d$ and the batch of N samples is sum-reduced:

$$\mathcal{L}_{dg} = -\frac{1}{N}\sum_{i=1}^{N}(\textsc{max}(\mathcal{L}_{d,n} - r(x_{c,n}, at_n), 0)^2 \tag{6}$$

where $\mathcal{L}_{dg}$ is the final value of the degradation loss.

Combining $\mathcal{L}_{dg}$ with Margin-based Softamx Over recent years, many loss functions have been proposed that optimize a cosine feature space with the help of an inter-class margin [5, 21, 34, 13], which have been widely adopted for training face recognition models. We use the term *margin-based softmax* to refer generally of this class of methods. In this work, we train models using a combination of both the margin-based softmax and the degradation loss. In particular, experiments are performed using ArcFace [5], which can be described as follows. Let θ_{y_i} be intra-class similarity, θ_j inter-class distance, s a feature scaling factor, m the inter-class margin, and N the mini-batch size:

$$\mathcal{L}_{ms} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s\cos(\theta_{y_i}+m)}}{e^{s\cos(\theta_{y_i}+m)} + \sum_{j\neq y_i} e^{s\cos\theta_j}}. \tag{7}$$

Finally, the the total loss $\mathcal{L}_{total}$ is a combination of $\mathcal{L}_{tr}$and $\mathcal{L}_{ms}$:

$$\mathcal{L}_{total} = \mathcal{L}_{ms} + \lambda_h\mathcal{L}_{tr} \tag{8}$$

where λ_h is a scalar hyperparameter.

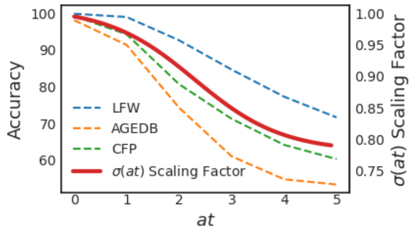

Figure 6. **Feature scalar $\sigma(at)$.** The x-axis is increasing levels of atmospherics at. The dashed lines show accuracy (left y-axis) of a baseline ArcFace model on atmospherics. The solid red line shows a sigmoid function $\sigma(at)$ fit to the the accuracy plots and then re-scaled to l–1.0 (right y-axis). The motivation for $\sigma(at)$ is that if magnitude is a proxy for quality-assessment, it should be correlated with accuracy [26, 21]. Lower bound l is 0.8. Degradation loss supervises feature magnitude using $\sigma(at)$.

3.3. Multi-stage Training

For degradation loss to be calculated, both a clean feature x_c and the atmospheric feature x_{at} need to be in GPU memory simultaneously. A simple way to achieve this is to pass a batch of clean images and a batch of clean atmospheric images within the training loop. For backpropagation, this can be implemented as two separate gradient updates (where the degradation loss is only applied to the second) or as a single gradient update. For a single gradient update, the batch size must be halfed, so gradients from both forward passes can be stored in memory. We find that both of these implementations have lower performance than a vanilla AT-Training model and take longer to train. To overcome this challenge we design a two-stage training procedure. In the first stage, a model is pretrained. Stage 1 can be performed with wight AT-Training or standard training. In the second stage, a copy of the stage 1 model is frozen. Clean images are passed through the frozen network while atmospheric samples are passed through the unfrozen network. As the frozen pretrained network does not change, the clean embeddings only need calculated one time (for all Stage 2 training runs). The result of this set-up is that Stage 2 training is as efficient as stage 1, while also employing the degradation loss. Note that degradation loss is used if and only if Stage 2 is used. In our experimentation, a model trained with Stage 2 indicates that the degradation loss is used. By design, degradation loss is never used in Stage 1. Figure 4 shows an overview of the two-stage training set-up.

4. Experimental Results

The experimental section of this work is organized as follows. Details for model training parameters, atmospheric simulation, and datasets are in Section 4.1. Results on simulated atmospherics are in Section 4.2 and results on low-resolution data is in Section 4.3. In Section 4.4, comparison are shown between various sampling probabilities for AT-Training. A discussion of results is provided in Section 4.5.

4.1. Experimental Set-up

4.1.1 Training Settings

ResNet-100 is used as a the backbone model with an embedding size of 512. For training, we use a batch size of 256 per GPU on each of 4 GPUs for a total batch size of 1028. Training is completed with machines with either 3xRTX3090,1xA6000 or 4xTitanRTX GPUs. For more efficient training, mixed-precision floating point [22] is used. Training is completed over 20 epochs with Stochastic Gradient Descent (SGD) optimizer, polynomial weight decay of $5e^{-4}$, and momentum of 0.9. A base learning rate 0.1 is used with polynomial learning rate scheduler. Horizontal-flip is used as augmentation. ArcFace loss [5] is used for training all models with margin of 0.5 and scaling factor s set to 64. For degradation loss $\mathcal{L}_{dg}$, lower feature vector scaling limit l is set to 0.8. The regularization margin angle θ_r is set to 10 and term scalar λ_h is set to 60.

4.1.2 Atmospheric Turbulence Simulation

For atmospheric turbulence simulation software, the *Phase-to-Space* model from ICCV 2021 is used [19]. This simulator is chosen as it is it can be accelerated on a GPU. For the simulator, the atmospheric turbulence 'strength' (i.e., at) is calculated by aperture diameter D divided by fried parameter r_0. Fried parameter r_0 is the net optical effect of atmospheric turbulence, which is a function of varying refractive indexes and distance. The AT simulator accepts a ratio of $\frac{D}{r_0}$. We fix D at 0.1 meters, and vary r_0 such that $at = \frac{D}{r_0}$ falls between 0.25 to 5.00. $\frac{D}{r_0} = 0.25$ is minor turbulence and $\frac{D}{r_0} = 5.0$ is severe. For testing, five levels of at are used: $1.0, 2.0, 3.0, 4.0, 5.0$. Figure 7 shows samples images from this range. It can be seen that at high turbulence levels, significant deformation and blur causes identifiable features to be affected, such as distances between face landmarks.

4.1.3 Datasets

Training is done on the WebFace4M dataset [39]. For evaluation, three dataset types are used: simulated atmospherics, mixed-quality, and low-resolution. For simulated atmospherics, we follow [26] and use a range of atmospherics on three high-image-quality datasets: LFW [9], CFP [24], and AgeDB [29]. CFP is avaulated with the CFP-FP protocol and AgeDB is evaluate with the AgeDB-30 protocol. For mixed-quality the IJB-C [11] dataset is used. TAR@FAR=0.0001 is reported for IJB-C. For low-resolution data, the TinyFace dataset [4] is used, which has

Model		Test Sets			
Stage 1	Stage 2	Simulated AT	TinyFace Rank 1	TinyFace Rank 5	L-p2p
Baseline	–	75.40	70.76	74.11	84.02
AT-Training	–	88.23	**73.18**	**75.99**	86.17
Baseline	$\mathcal{L}_{dg}$	86.76	70.84	74.20	85.76
AT-Training	$\mathcal{L}_{dg}$	**89.34**	72.85	75.86	**86.34**

L-p2p=low paq2piq blind image-quality set

Table 1. A comparison of models before and after the two stage training procedure. Th best result in each column is highlighted in bold. The model trained with degradation loss $\mathcal{L}_{dg}$ in Stage 2 is the best performing model on simulated atmospherics and L-p2p. The vanilla AT-Training model is the best performing model in TinyFace rank-1 and rank-5. Refer to Figure 4 for an overview on our two-stage training.

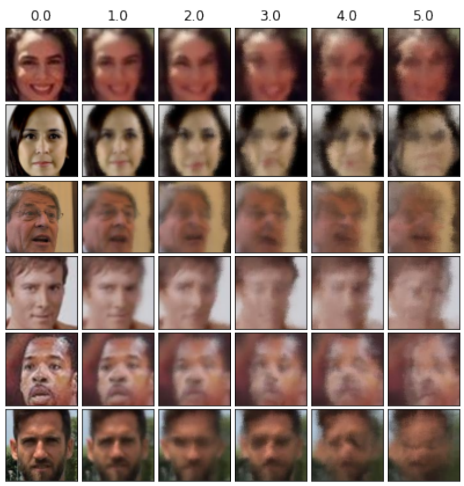

Figure 7. Sample images from the Mao et al. simulator [19] for atmospheric turbulence strengths (at) 0.0–5.0.

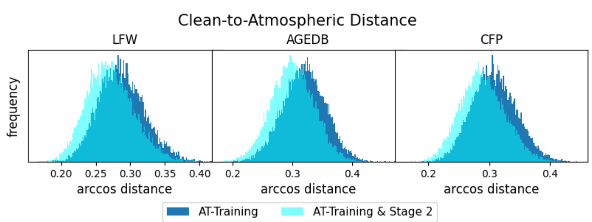

Figure 8. Distribution plots of the distance between the feature of a clean image and a feature from a copy of the image with atmospherics at $at = 3.0$. Lower arccos distance indicates a feature space more robust to atmospherics.

2,569 probe identities and 157,871 gallery images. Following previous work [13], 1:N Rank 1 and Rank 5 is presented for TinyFace. As an additional low-quality test set, 10,000 1:1 verification pairs of images with lowest 10 percentile paq2piq [37] blind-image quality are sampled from the WebFace42M dataset [39] (excluding WebFace4M identities).

4.2. Simulated Atmospheric Results

Prior work has reported face recognition results on simulated atmospheric turbulence with image-restoration methods [36, 35, 14, 15]. Yasarla et al. 2022 [36] is the state-of-the-art and used for comparison. In Figure 5, plots are shown for LFW, CFP, and AgeDB under increasing levels of atmospheric turbulence. It can be seen that methods using AT-Training significantly outperform the baseline model and the image-restoration based approach on all levels of atmospherics. The method utilizing Stage 2 and the degradation loss further improves over vanilla AT-Training. Across all models, it can also be seen that performance is lowest on AgeDB under atmospherics. This may be do to the fact that AgeDB has 1:1 verification pairs have 30+ year separation, which often results in a black-and-white to colored-image comparison. It can also be seen that clean-to-atmospheric comparisons score much higher than atmospheric-to-atmospheric for models trained with atmospherics. However, performance between clean-to-atmospheric and atmospheric-to-atmospheric is nearly equally low for the baseline model.

In Table 1, performance over all atmospheric sets is averaged. AT-Training outperforms the baseline ArcFace model by 12.83%. Using Stage 2 and degradation loss after AT-Training further improves performance by 1.11%. Using Stage 2 and degradation loss *after standard training* improves upon the baseline, but scores lower than the aforementioned models. To understand improvements from degradation loss, the distance between clean-to-atmospheric samples of the same image is calculated in Figure 8. In Figure 8, clean-to-atmospheric distance is less great for the model trained with degradation loss. That is, features supervised with degradation loss have less variance from the original feature under atmospherics. This coincides with our motivation that degradation loss supervises for robustness under atmospherics.

4.3. Low-image-quality Results

The publicly available TinyFace dataset [4] is used as a low-resolution dataset. In Table 1, it can be seen than all of our models outperform the baseline ArcFace model. The

Method	Venue	Dataset	Rank 1	Rank 5
CF [10]	CVPR20	MS1Mv2	63.68	67.65
URL [31]	CVPR20	MS1Mv2	63.89	68.67
AdaFace [13]	CVPR22	MS1Mv2	68.21	71.54
ArcFace [5]	CVPR19	WF4M	70.76	74.11
AdaFace [13]	CVPR22	WF4M	72.02	74.52
AT-Train (Ours)	*manuscript*	WF4M	**73.18**	**75.99**

Table 2. A comparison of AT-Training to prior works on TinyFace [4]. Our AT-Training method acheives state-of-the-art on TinyFace rank-1 and TinyFace rank-5. *Due to dataset redaction, our model is not trained with MS1Mv2.*

$\mathcal{P}(at{=}0)$	k	Sim-AT	TF-R1	TF-R5	IJB-C
0.50	3	87.07	72.29	75.59	97.08
0.66	3	86.32	72.31	75.00	97.16
0.50	5	**88.23**	**73.18**	**75.99**	97.03
0.66	5	87.28	72.91	75.43	**97.18**
1.00	–	75.40	70.76	74.11	97.16

TF-R1=TinyFace rank1; TF-R5=TinyFace rank-5

Table 3. Performance under different at sampling conditions for AT-Training. The last row is the baseline model. IJB-C reported with TAR@FAR=0.001. Best result per column highlighted in bold.

model trained with vanilla AT-Training is the best performing model. Degradation loss does not outperform vanilla AT-Training, which may have to do with the different characteristics of low-resolution and atmospheric (further discussed in Section 4.5). In Table 2, comparison is shown with prior methods on the TinyFace dataset. The previous State-of-The-Art (SoTA) on TinyFace is AdaFace [13], which scores 72.02 Rank-1 and 74.52 Rank-5. Our AT-Training method improves Rank-1 by 1.16% and Rank-5 by 1.47% to scores of 73.18% and 75.99%, for SoTA performance.

As mentioned in Section 4.1.3, a test set of low paq2piq [37] blind-image-quality assessment (L-p2p test set) is evaluated on. Table 1 shows all of our models outperform the baseline on L-p2p. The model trained with degradation loss is best performing model on L-p2p.

4.4. Atmospheric Sampling Probabilities

As introduced in Section 3.1, an open question is determining an optimal sampling distribution for atmospheric levels. In this experiment, at sampling is modified with two parameters for Stage 1 AT-Training. The first is k, which is the highest level of at, such that $at \in [0, k]$. The second is $\mathcal{P}(at{=}0)$. Four full training runs are completed with each the four combinations of $k = 3, 5$ and $\mathcal{P}(at{=}0) = 0.50, 0.66$. $k = 5$ is the wider range at and $k = 3$ excludes the most noisy atmospheric samples (see Figure 7 for reference). For $k = 3$, at is sampled uniformly from the set $\{0.25, 0.5, 1.0, 2.0, 3.0\}$. For $k = 5$, at sampled uniformly from the set $\{0.25, 0.5, 1.0, 2.0, 3.0, 4.0, 5.0\}$.

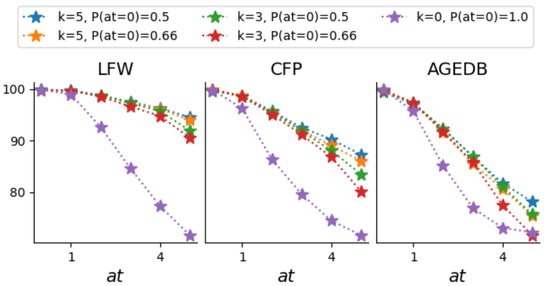

Figure 9. Accuracy on simulated atmospherics for AT-Training models trained with different at sampling distributions.

Figure 10. Both low-resolution and simulated atmospheric are challenging conditions for recognition. Low-resolution are pixelated and lack detail, but are not deformed. Atmospheric sample are blurry and deformed. Methods from this work improve performance on both.

Table 3 shows results for each run on as well as the baseline model (i.e., $\mathcal{P}(at{=}0) = 1.0$). It can be seen that the best performing model on both the simulated atmospherics and low-resolution is with parameters $k = 5; \mathcal{P}(at{=}0) = 1.0$. Additionally the model trained with less frequent atmospheric data (i.e., $k = 5; \mathcal{P}(at{=}0) = 1.0$) scores 97.18 on IJB-C TAR@FAR=0.0001, slightly outperforming the baseline. The results on simulated atmospherics can bee in more detail in the plots in Figure 9, where five different levels of atmospherics are show shown.

4.5. Discussion of Results

In our experimentation we find AT-Training significantly improves over the baseline on simulated atmospheric turbulence (Figure 5). The degradation loss further improves performance (Figure 5, Table 1) by creating feature that are most robust (Figure 8). Degradation loss also improves performance on the low-image-quality L-p2p set. For low-resolution, degradation loss and AT-Training improve over prior SoTA; and AT-Training is the best performing model on low-resolution. We hypothesize that degradation loss improves performance on by learning a stronger level of robustness to atmospheric deformations specifically. As can be seen in Figure 10, low-resolution images do not suffer from the same face-feature deformations as atmospherics. A path of future work is collecting and evaluating on real atmospheric data.

5. Conclusion

In this work, we have proposed simulated atmospheric turbulence as a training augmentation, the degradation loss, and a corresponding training procedure for face recognition in low-image-quality conditions. Our methodology outperforms a baseline model on simulated atmospherics by a considerable margin. Degradation loss improves upon vanilla AT-Training by learning a deep feature space that is more robust to atmospherics. SoTA is also achieved on the low-resolution TinyFace dataset with our AT-Training model.

References

[1] Nantheera Anantrasirichai, Alin Achim, Nick G Kingsbury, and David R Bull. Atmospheric turbulence mitigation using complex wavelet-based fusion. *IEEE Transactions on Image Processing*, 22(6):2398–2408, 2013.

[2] Soma Biswas, Kevin W Bowyer, and Patrick J Flynn. Multidimensional scaling for matching low-resolution facial images. In *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–6. IEEE, 2010.

[3] Jie Chang, Zhonghao Lan, Changmao Cheng, and Yichen Wei. Data uncertainty learning in face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5710–5719, 2020.

[4] Zhiyi Cheng, Xiatian Zhu, and Shaogang Gong. Low-resolution face recognition. In *Asian Conference on Computer Vision*, pages 605–621. Springer, 2018.

[5] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019.

[6] Akshay Raj Dhamija, Manuel Günther, and Terrance Boult. Reducing network agnostophobia. *Advances in Neural Information Processing Systems*, 31, 2018.

[7] Hamidreza Fazlali, Shahram Shirani, Michael BradforSd, and Thia Kirubarajan. Atmospheric turbulence removal in long-range imaging using a data-driven-based approach. *International Journal of Computer Vision*, 130(4):1031–1049, 2022.

[8] Md Hasan Furhad, Murat Tahtali, and Andrew Lambert. Restoring atmospheric-turbulence-degraded images. *Applied optics*, 55(19):5082–5090, 2016.

[9] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*, 2008.

[10] Yuge Huang, Yuhan Wang, Ying Tai, Xiaoming Liu, Pengcheng Shen, Shaoxin Li, Jilin Li, and Feiyue Huang. Curricularface: adaptive curriculum learning loss for deep face recognition. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5901–5910, 2020.

[11] Nathan D Kalka, Brianna Maze, James A Duncan, Kevin O'Connor, Stephen Elliott, Kaleb Hebert, Julia Bryan, and Anil K Jain. Ijb–s: Iarpa janus surveillance video benchmark. In *2018 IEEE 9th international conference on biometrics theory, applications and systems (BTAS)*, pages 1–9. IEEE, 2018.

[12] Syed Safwan Khalid, Muhammad Awais, Zhen-Hua Feng, Chi-Ho Chan, Ammarah Farooq, Ali Akbari, and Josef Kittler. Resolution invariant face recognition using a distillation approach. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2(4):410–420, 2020.

[13] Minchul Kim, Anil K. Jain, and Xiaoming Liu. Adaface: Quality adaptive margin for face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18750–18759, June 2022.

[14] Chun Pong Lau, Amit Kumar, and Rama Chellappa. Semi-supervised landmark-guided restoration of atmospheric turbulent images. *IEEE Journal of Selected Topics in Signal Processing*, 15(2):204–215, 2021.

[15] Chun Pong Lau, Hossein Souri, and Rama Chellappa. Atfacegan: Single face image restoration and recognition from atmospheric turbulence. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, pages 32–39. IEEE, 2020.

[16] Dalong Li, Russell M Mersereau, and Steven Simske. Atmospheric turbulence-degraded image restoration using principal components analysis. *IEEE Geoscience and Remote Sensing Letters*, 4(3):340–344, 2007.

[17] Nianyi Li, Simron Thapa, Cameron Whyte, Albert W. Reed, Suren Jayasuriya, and Jinwei Ye. Unsupervised non-rigid image distortion removal via grid deformation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2522–2532, October 2021.

[18] Shen Li, Jianqing Xu, Xiaqing Xu, Pengcheng Shen, Shaoxin Li, and Bryan Hooi. Spherical confidence learning for face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15629–15637, 2021.

[19] Zhiyuan Mao, Nicholas Chimitt, and Stanley H Chan. Accelerating atmospheric turbulence simulation via learned phase-to-space transform. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14759–14768, 2021.

[20] Kangfu Mei and Vishal M Patel. Ltt-gan: Looking through turbulence by inverting gans. *arXiv preprint arXiv:2112.02379*, 2021.

[21] Qiang Meng, Shichao Zhao, Zhida Huang, and Feng Zhou. Magface: A universal representation for face recognition and quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14225–14234, 2021.

[22] Paulius Micikevicius, Sharan Narang, Jonah Alben, Gregory Diamos, Erich Elsen, David Garcia, Boris Ginsburg, Michael Houston, Oleksii Kuchaiev, Ganesh Venkatesh, and Hao Wu. Mixed precision training. In *International Conference on Learning Representations*, 2018.

[23] Nayaneesh Kumar Mishra, Mainak Dutta, and Satish Kumar Singh. Multiscale parallel deep cnn (mpdcnn) architecture for the real low-resolution face recognition for surveillance. *Image and Vision Computing*, 115:104290, 2021.

[24] Stylianos Moschoglou, Athanasios Papaioannou, Christos Sagonas, Jiankang Deng, Irene Kotsia, and Stefanos Zafeiriou. Agedb: the first manually collected, in-the-wild age database. In *proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 51–59, 2017.

[25] Nithin Gopalakrishnan Nair, Kangfu Mei, and Vishal M Patel. A comparison of different atmospheric turbulence simulation methods for image restoration. *arXiv preprint arXiv:2204.08974*, 2022.

[26] Wes Robbins and Terrance E. Boult. On the effect of atmospheric turbulence in the feature space of deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 1618–1626, June 2022.

[27] Michael C Roggemann, Byron M Welsh, and Bobby R Hunt. *Imaging through turbulence*. CRC press, 1996.

[28] AL Rukosuev, AV Kudryashov, AN Lylova, VV Samarkin, and Yu V Sheldakova. Adaptive optics system for real-time wavefront correction. *Atmospheric and oceanic optics*, 28(4):381–386, 2015.

[29] Soumyadip Sengupta, Jun-Cheng Chen, Carlos Castillo, Vishal M Patel, Rama Chellappa, and David W Jacobs. Frontal to profile face verification in the wild. In *2016 IEEE winter conference on applications of computer vision (WACV)*, pages 1–9. IEEE, 2016.

[30] Yichun Shi and Anil K Jain. Probabilistic face embeddings. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6902–6911, 2019.

[31] Yichun Shi, Xiang Yu, Kihyuk Sohn, Manmohan Chandraker, and Anil K Jain. Towards universal representation learning for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6817–6826, 2020.

[32] Robert K Tyson and Benjamin West Frazier. *Principles of adaptive optics*. CRC press, 2022.

[33] Engin Uzun, Ahmet Anıl Dursun, and Erdem Akagündüz. Augmentation of atmospheric turbulence effects on thermal adapted object detection models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 241–248, June 2022.

[34] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5265–5274, 2018.

[35] Rajeev Yasarla and Vishal M. Patel. Learning to restore images degraded by atmospheric turbulence using uncertainty. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 1694–1698, 2021.

[36] Rajeev Yasarla and Vishal M. Patel. Cnn-based restoration of a single face image degraded by atmospheric turbulence. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(2):222–233, 2022.

[37] Zhenqiang Ying, Haoran Niu, Praful Gupta, Dhruv Mahajan, Deepti Ghadiyaram, and Alan Bovik. From patches to pictures (paq-2-piq): Mapping the perceptual space of picture quality. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[38] Xiang Zhu and Peyman Milanfar. Removing atmospheric turbulence via space-invariant deconvolution. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):157–170, 2012.

[39] Zheng Zhu, Guan Huang, Jiankang Deng, Yun Ye, Junjie Huang, Xinze Chen, Jiagang Zhu, Tian Yang, Jiwen Lu, Dalong Du, et al. Webface260m: A benchmark unveiling the power of million-scale deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10492–10502, 2021.