On the Effect of Atmospheric Turbulence in the Feature Space of Deep Face Recognition

Wes Robbins* Terrance Boult*
University of Colorado, Colorado Springs
wrobbins@uccs.edu, tboult@vast.uccs.edu

Abstract

When captured over long distances, image quality is degraded by inconsistent refractive indexes in the atmosphere. This effect, known as Atmospheric Turbulence (AT), leads to lower performance for vision-based biometric systems such as face recognition. To account for AT, the literature has proposed methods to restore face-images from atmospheric turbulence, but has limited success. There is still a need to understand how atmospheric turbulence breaks recognition performance. We offer a first-look in this direction by providing a study on the effect of atmospheric turbulence in the feature space of deep-learning-based face recognition. We present results on recognition performance and feature space transformation caused by a wide range of AT levels. In deep feature space, we find interesting phenomena such as increasing feature magnitudes, which contradicts the expected result from the literature. From our results, we are able to identify an effect that makes face recognition under atmospheric turbulence uniquely difficult, which we call feature defection. In total, our findings suggest several areas of available improvement which can be used as a guideline for further progress in building models that are robust to AT.

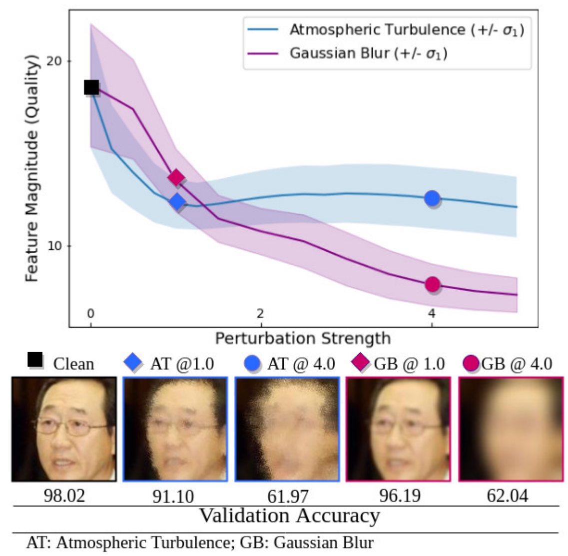

Figure 1. This paper addresses magnitude-based quality assessment on face-images perturbed with different levels of Gaussian blur (GB) and Atmospheric Turbulence (AT). Current belief is that magnitude is a quality measure that is positively correlated with recognition rate. As expected, under increasing GB, feature magnitude (i.e., quality) decreases monotonically. Surprisingly, under increasing AT, feature magnitude **increases** at medium perturbation levels, resulting in images with AT=1.0 and AT=4.0 to have equal assessed "magnitude-based quality" (shown with blue samples), despite having significantly different validation accuracy of 91.10% and 61.97%. (See Section 3 for details on AT and GB generation.)

1. Introduction

Face Recognition (FR) is used in critical applications and can improve safety and security in many settings. However, false classifications can be of high consequence and lead to public mistrust in identification technology. Thus, it is important that deployed recognition methods achieve both high performance and awareness of classification uncertainty. In good imaging conditions, current Face Recognition (FR) methods can achieve the necessary performance. Significant research in deep learning approaches has driven FR performance to over 99% accuracy on benchmark datasets [4, 15]. However, low-quality imaging conditions have been show shown to be more challenging [10]. Under low-quality conditions, both recognition performance and quality assessment accuracy is degraded.

Atmospheric Turbulence (AT) is one effect that creates low-quality imaging conditions and degrades face recogni-

*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.

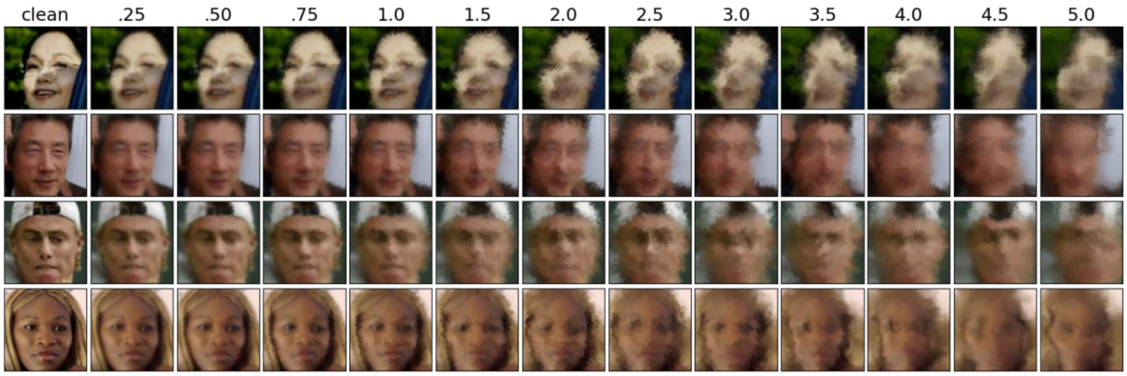

Figure 2. Images with simulated atmospheric turbulence—strengths ranging from 0.25 (weak) to 5.0 (severe). It can be seen that at low strengths identification features are retained. At high strengths, identities may be unrecognizable. Simulation details and settings can be found in Section 3.1.

tion performance. AT image perturbations, which can be severe in long-distance imaging systems, are caused by spatially and temporally variable refractive indexes in the atmosphere. The perturbations caused by AT can be modeled as a composition of geometric deformation and blur [19]. More formally, the effect of AT on unobserved clear image I at frame k can be described as the composition of a temporally varying deformation function H_k and point-spread-function D_k, plus sensor noise n_k:

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

The terms in Equation 1 are known to be random and non-linear [19], which has made handling AT a challenging problem in optics. While atmospheric turbulence can be partially mitigated with adaptive optics techniques, these techniques require large and expensive hardware. For this reason, there is a significant utility for face recognition models that can 1) maintain reasonable performance under AT and 2) classify uncertainty caused by AT.

Mitigating the effect of AT on deep-learning-based face recognition has only recently been studied. A few works have proposed methods for image-restoration to compensate for AT [11, 25]. As an alternative to image-restoration, we are interested in asking: can the effect of atmospheric turbulence on face recognition be mitigated directly in feature space? In order to answer this is question, it is helpful to understand *how* atmospheric turbulence breaks face recognition performance, which has not been studied in the literature. To this end, we provide a study on the effects of atmospheric turbulence on deep-learning-based face recognition. We perform analysis with a wide range of atmospheric turbulence levels, previously unstudied for face recognition, which is relevant in real-world applications where atmospheric turbulence levels are variable. Additionally, we compare the effects of atmospheric turbulence to Gaussian blur. Our focus is on deep feature space, allowing for more detailed observations than those attainable from only performance metrics. Specifically, we look at two properties in feature space: *feature magnitude* and *feature trajectory*.

Feature Magnitude In the literature, several works have shown magnitude to representative of sample quality, where lower magnitude implies lower quality [2, 5]. This includes a state-of-the-art Face-Image-Quality-Assessment (FIQA) method [15]. Surprisingly, we find magnitude does not decrease monotonically with increasing turbulence levels. That is, lower quality images are being assessed to be higher quality under AT. This result is highlighted in Figure 1, where we show that the samples at $AT = 1.0$ and $AT = 4.0$ have similar magnitudes, despite much lower validation scores at $AT = 4.0$, where AT is the strength of the AT perturbation (details on AT strengths in Section 3.1). Additionally, Figure 1 shows this effect does not occur under increasing levels of Gaussian blur.

Feature Trajectory The feature of a sample affected by atmospheric turbulence resides in a different section of feature space than the feature of a clean sample. For incremental increases in atmospheric turbulence strength, we examine feature trajectory vectors. We find that identity specific feature trajectories exist at low turbulence levels. However, at medium to high turbulence levels, samples follow trajectories that merge and rotate, leading to a single subspace within the feature space. In Section 5, we demonstrate this effect with t-SNE plots of feature trajectories and with inter-class and intra-class distance measurements.

From our results, we gain a better understanding of the effect of atmospheric turbulence. We conclude that at approximately $AT > 1.5$, AT artifacts are *misinterpreted by the recognition model as salient features for recognition*, which does not occur from similar Gaussian blur degradations. We call this effect *feature defection*. Specifically, we define *feature defection* as the result of image features that are blurred, deformed, or magnified and then subsequently misinterpreted as highly useful features for classification. In

Figure 2, many affected face features can be seen at high levels of turbulence.

We find *feature defection* to have two primary negative effects on model behavior. First, as previously mentioned, feature magnitude begins to increase at $AT = 1.5$, leading to inaccurate quality assessment. Second, samples degraded with AT have low inter-class distances, creating an AT super-class, which severely degrades performance. Based on these results, we identify the following problems for future work:

1. **Magnitude-Accuracy Alignment** While quality-aware magnitude is a promising technique for assessing face-image quality, more work is needed to increase the correlation between magnitude and recognition performance under atmospheric turbulence.
2. **Inter-Class Separation** At high levels of atmospheric turbulence, AT samples cluster together. Future work needs to improve inter-class separation.

Improvements in these directions can lead to more robust models under atmospheric turbulence and, thus, more reliable in surveillance applications. In future work, we plan to address these challenges with training procedures that optimize for both recognition performance and the desired feature space behavior.

In summary, the contributions of this work are as follows:

- Compare the effects of a wide range of atmospheric turbulence levels to the effects of Gaussian blur for both recognition performance and quality assessment.
- Identify *feature defection* as a cause of poor quality assessment and lower recognition performance under atmospheric turbulence.
- Demonstrate a state-of-the-art FIQA method performs poorly under atmospheric turbulence.
- Outline paths of future work to overcome challenges identified in this work.

The remainder of the paper is organized into six sections. In the following section, we review related work. In Section 3 we describe our experimental setup. In Section 4 we analyze feature magnitude under atmospheric turbulence, which is followed by an analysis of feature space trajectory in Section 5. In Section 6, we visualize the effects of atmospheric turbulence with input activation maps. Finally, we summarize our findings and discuss future work in Section 7.

2. Related Work

Previous work has studied the effect of blur and other noise on face recognition. Lie et. al. [12] provided a nice initial theory for hallucinating a high-resolution face image from low-resolution data. Heflin et al. proposed an improved technique for single-image deblurring [7] with fewer assumptions and evaluated its impact on recognition performance. Punnappurath et al. propose an algorithm for jointly handling non-uniform blur and poor illumination [18]. In contrast to these works, we focus on deep-learning-based face recognition.

Another line of work has studied the general effects of low quality images on deep learning models. Karahan et al. studied the effect of Gaussian noise and occlusions on early CNN architectures [9]. The UG2 challenge has been proposed to advance image understanding in poor imaging conditions [24]. Specifically addressing face-images, Banerjee et al. proposed a GAN-based method for image reconstruction [1]. Other deep-learning based approaches have also been explored [3,21]. Their work differs from ours as we focus on atmospheric turbulence rather than general image blur and distortions.

Recently there has been work on restoring face-images degraded by atmospheric turbulence. Lau et al. proposed ATFaceGan, a method that uses two generators to handle deformation and deblurring [11] respectively. In [25], 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. The estimated uncertainty maps are then used to guide the network to obtain the restored image. Differing from these works, rather than image quality restoration or just learning maps, we focus on understanding the impacts of turbulence on face recognition feature spaces and the subsequent impact on recognition. The insights from this paper could probably be combined with some of the above restoration approaches, but that is left for future work.

3. Experimental Setup

3.1. Atmospheric Turbulence Simulation

In order to obtain sufficient data for experiments, we use a simulator to generate atmospheric turbulence. *Accelerating Atmospheric Turbulence Simulation via Learned Phase-to-Space Transform*, published in ICCV 2021, is used [14]. Atmospheric turbulence strength 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 (caused by air motion) and distance. Readers can find detailed information at [19].

The AT software accepts a ratio of $\frac{D}{r_0}$. We vary $\frac{D}{r_0}$ from 0.25 to 5.00 at increments of 0.25 for a total of 20 different AT strengths. $\frac{D}{r_0} = 0.25$ is minor turbulence and $\frac{D}{r_0} = 5.0$ is severe. Figure 2 shows sample images perturbed with twelve different levels of AT. In Figure 2, it can be seen that low perturbation levels appear as only slight blur, and important identity features are retained. At high turbulence, significant deformation and blur causes identifiable features to be affected (e.g., distance between face landmarks) or lost completely.

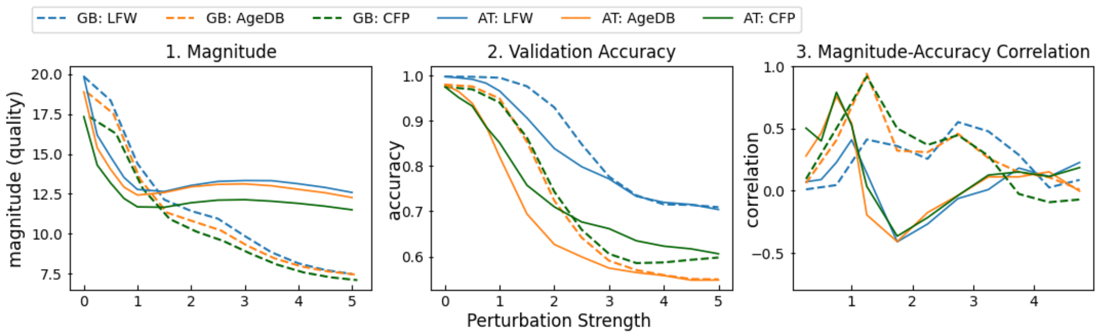

Figure 3. The effect of two different perturbations on feature magnitude and model performance. Dashed lines represent Gaussian blur (GB) perturbations and solid lines represent Atmospheric Turbulence (AT) perturbations, and each line color represents a different dataset. The x-axis shows increasing perturbation strength (details on perturbation generation can be found in Section 3). Plot 1 shows feature magnitude, which is a means for quality assessment; Plot 2 shows one-to-one face recognition performance; and Plot 3 shows correlation between magnitude and recognition performance. In Plot 3, in can be seen that magnitude and accuracy have significantly lower—and even negative—correlation under AT than GB.

3.2. Gaussian Blur Image Degradation

To better understand the effects of atmospheric turbulence on face recognition, we compare them to the effects of Gaussian blur (GB). Since we use a wide range of strengths of AT (Section 3.1), we also generate images with a wide range of strengths of GB—where greater strength implies greater degradation.

To make the GB to AT comparison useful, we generate GB such that validation accuracy is equally low for both the greatest strengths of GB and AT. To achieve this, GB is generated with Gaussian kernel size 31x31 and standard deviation σ ranging from 0 to 10 (larger σ is greater perturbation). At $\sigma = 10$, validation accuracy matches $AT = 5$ on all three validation datasets.

In order to improve visual comparisons, we plot both AT and GB on a *perturbation strength* axis ranging from 0 to 5 (where 0 is no perturbation). AT strength maps to this axis unchanged, and GB maps to this axis with $\frac{1}{2}\sigma$. The *perturbation strength* axis is found in Figure 1 and Figure 3.

3.3. Recognition Model Parameters

A Resnet-50 is used as the backbone of our model, followed by a fully connected layer to the embedding layer of size 512. The model is trained for 25 epochs on MS1M-arcface dataset [4]. For a loss function, we use arcface [4] + Magface quality-aware margin and regularization term [15] (discussed further in Section 4.1). Magface scaling factor λ_g is set at 35.

SGD is used as an optimizer with momentum of 0.9 and weight decay of $1e^{-5}$. Training batch size is 512, initial learning rate is 0.1, and we use learning rate drop by a factor of 0.1 at epochs 10, 18, and 22. In total, our training set-up closely follows Meng et al., which achieves state-of-the-art on recognition benchmarks [15].

3.4. Validation Datasets

Using MagFace model [15] pretrained on MS1M-arcface [4], we perform our experiments on the following validation datasets: Labeled-Faces-in-the-Wild (LFW) [8], AgeDB [16], and Celebrities in Fronal-Profile (CFP) [20]. We do one-to-one recognition with 6,000 pairs from each dataset. For the CFP dataset, we use the frontal-profile procedure, where validation pairs contain one camera-facing image and one profile (side face) image. A copy of each validation dataset is made for each level of atmospheric turbulence and Gaussian blur.

4. Feature Magnitude Analysis

Deep feature magnitude is known to be strongly related to misclassification rates [5]. Several papers have built on feature magnitude results to develop losses to manage magnitude to improve recognition and robustness [2, 5]. Feature magnitude, which impacts recognition rates, has also been suggested to be representative of face-image quality, where lower magnitude implies lower quality. Meng et al. propose Magface, a loss function that enforces quality-aware feature magnitudes [15]. In a recent survey of Face-Image Quality Assessment (FIQA), Magface is found to be a state-of-the-art method [6].

These prior works motivate us to examine feature magnitude as an indication of the effect of atmospheric turbulence and as a quality assessment method. First, in Section 4.1, we review the common geometric interpretation of face recognition feature space as a hypersphere (where hypersphere radius is equivalent to feature magnitude), and the use of feature magnitude as a quality assessment. In Section 4.2, we present feature magnitude results under atmospheric turbulence.

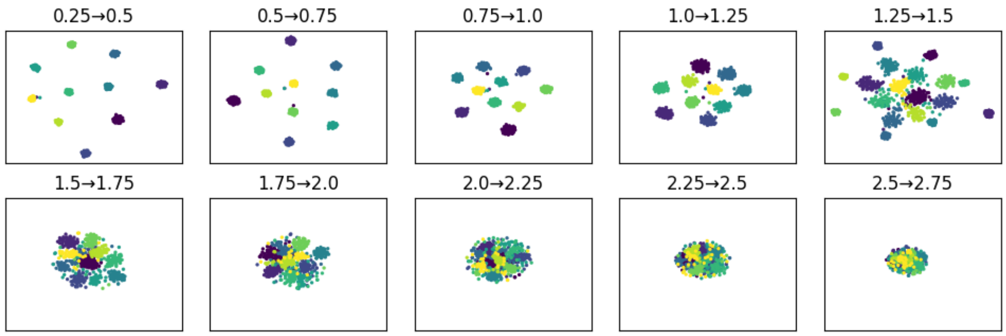

Figure 4. t-SNE plot of the trajectories of ten different identities. Each color represents a separate identity, and each point represents a single sample. There are 50 samples per identity. At low levels the identifies maintain unique trajectory clusters. At approximately $AT = 1.5$, identities splint into multiple trajectory clusters and begin rotating. Then, at high AT, all identities follow similar trajectories. In the bottom right corner, $2.50 \rightarrow 2.75$ is the last trajectory plot shown as levels greater than $AT = 2.75$ produce plots that are visually the same as $2.50 \rightarrow 2.75$ (i.e., a single central cluster).

4.1. Review of Hypersphere Feature Space

In open-set face recognition, recognition performance depends on intra-class compactness and inter-class separation. While early works used euclidean distance to measure feature space distances, recent approaches benefit from using geodesic distances. Many approaches have proposed loss modifications to optimize for geodesic distances during training [4, 13, 22]. Prominently, ArcFace loss $\mathcal{L}_{af}$ penalizes both intra-class separation θ_{y_i} and inter-class similarity θ_j:

$$\mathcal{L}_{af} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{s\cos(\theta_{y_i}+m)}}{e^{s\cos(\theta_{y_i}+m)} + \sum_{j \neq y_i} e^{s\cos\theta_j}}. \quad (2)$$

where a scaling factor s is used to scale normalized features to a fixed-radius hypersphere.

Recently, the MagFace loss is proposed to create feature space structure to relate feature magnitude and sample quality [15]. Prior work had demonstrated managing magnitude to improve image classification [5], and MagFace applies this notion to face recognition with loss modifications to enforce magnitudes representative of face-image quality. Precisely, MagFace loss $\mathcal{L}_{mf}$ modifies the fixed angular margin m from Equation 3 to dynamic margin $m(a_i)$, where $a_i = ||F(x_i)||$ is feature magnitude for learned function F and input x_i. Additionally, a regularizer $g(a_i)$ with scaling factor λ_g is added:

$$\mathcal{L}_{mf} = \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}_i, \quad \text{where}$$

$$\mathcal{L}_i = -\log \frac{e^{s\cos(\theta_{y_i}+m(a_i))}}{e^{s\cos(\theta_{y_i}+m(a_i))} + \sum_{j \neq y_i} e^{s\cos\theta_j}} + \lambda_g g(a_i). \quad (3)$$

MagFace demonstrates improved recognition performance, and creates a feature space-based quality metric, where features of lower quality samples lie on hyperspheres with smaller radiuses. A recent survey on Face-Image Quality Assessment (FIQA) finds MagFace to be a state-of-the-art FIQA method [6].

The properties of Magface, and strong prior results, make it advantageous for face recognition under atmospheric turbulence, where feature space structure can be used for both improved recognition performance and identifying samples highly affected by turbulence. For this reasons, we adopt $\mathcal{L}_{mf}$ when training the model used throughout this work.

4.2. Feature Magnitude Results

In this section, we present feature magnitude results under a wide range of strengths of Gaussian blur and atmospheric turbulence. The procedure for generating perturbations is described in Section 3. We calculate magnitudes on images perturbed with Gaussian blur for a baseline comparison with atmospheric turbulence. Figure 3 Plot 1 shows average feature magnitudes for each of three datasets perturbed with Gaussian blur. It can be seen that feature magnitude decreases monotonically with increasing Gaussian blur. This is expected behavior as images under greater GB perturbations are being assessed to be lower quality.

Figure 3 Plot 1 also shows feature magnitudes under increasing atmospheric turbulence. It can be seen that under atmospheric turbulence, feature magnitude initially drops, and then *increases* from $AT = 1.5$ to $AT = 3$. This is an unexpected result. *Samples degraded by greater levels of AT are being assessed to be of higher quality than samples with weaker AT degradations.* By referencing Figure 2, it is clear that samples of higher AT have fewer identifiable face features than samples perturbed by lower AT. Thus, higher AT samples should have lower magnitude (i.e., quality score).

In Figure 3 Plot 2, validation accuracy for each dataset is shown. In Plot 2, it can be seen that under both atmo-

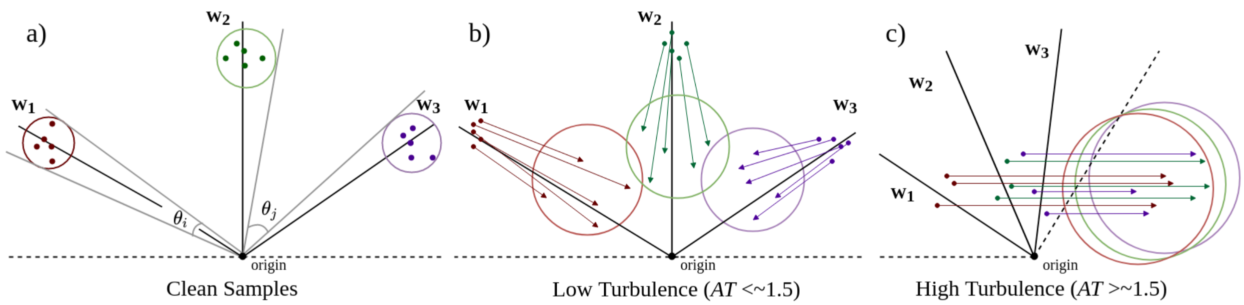

Figure 5. Visualization of feature space-shift under increasing atmospheric turbulence. Each color represents a separate class, and each circle represents the distribution of that class. A single point represents a single sample, and arrows represent feature trajectories. Vectors from the origin $W_{1..3}$ represent three class centers before AT transformations. In **a**, clean samples of the same identity cluster tightly—low intra-class distance θ_i and large inter-class distance θ_j are observed. In diagram **b**, low AT causes a decrease in feature magnitudes and overlapping class distributions, leading to a drop in performance. In **c**, high AT causes equal or increased magnitudes, and the formation of a high-AT superclass. The observations represented in this diagram are based on results shown in Figure 3 (magnitude), Figure 4 (feature trajectory), and Figure 7 (inter- and intra-class distance).

spheric turbulence and Gaussian blur, validation accuracy decreases with increasing perturbation strength. The validation accuracy under AT perturbation and GB perturbations are the same at perturbation strength equal to 5, despite having vastly different magnitudes.

In Figure 3 Plot 3, we show the correlation between magnitude and validation accuracy (the two previous plots). Plot 3 shows significantly lower correlation under AT than GB. At some AT levels, there is **negative correlation** between accuracy and quality assessment (the correlation coefficient gets as low as -0.44). This is due to the magnitude increase from $AT = 1.5$ to $AT = 3$, while accuracy is decreasing. Comparatively, GB has high correlation at similar perturbation levels.

The negative correlation between accuracy and quality assessment is a surprising result with several implications. Work on "biometric completeness" [17], shows that solving face image quality is equivalent to solving the face recognition problem. Thus, if magnitudes are a direct measure of quality, they should be positively correlated with recognition rates, at least for the algorithm that uses that feature space. Given that magnitude and accuracy are not strictly positively correlated, two conclusions follow:

1. MagFace quality assessment performs poorly under certain levels of atmospheric turbulence.
2. Artifacts from atmospheric turbulence have a unique effect on face recognition.

In the following section we further explore the implications of conclusion 2.

5. Feature Trajectory Analysis

In Section 4, it is found that at low turbulence levels, features move closer to the origin. That is, they have trajectories generally towards the origin. At increments between higher levels of atmospheric turbulence, it is found that feature trajectories do not decrease distance from the origin. In this section, we obtain a more fine-grained understanding of feature trajectories under increasing AT.

5.1. Identity-Specific Trajectories

The relationship between trajectories of different samples is of interest for understanding how feature space reacts under increasing turbulence. If trajectories of all samples are unrelated, it suggests the feature space responds randomly to AT. If trajectories from all samples are similar, it suggests all samples are affected similarly—and moving to a single subspace of the feature space. Lastly, if intra-class trajectories are similar, but inter-class trajectories are dissimilar, it suggests identities are uniquely affected by the increasing perturbations.

To visualize the relationship between trajectories, we plot a 2D projection of the trajectories of 50 samples for each of 10 identities using a t-Distributed Stochastic Neighbour Embedding (t-SNE) plot. In the t-SNE plot, trajectory vector $\vec{\mathbf{x_{i,j}}}$ is calculated as follows:

$$\vec{\mathbf{x_{i,j}}} = (\mathbf{x_j} - \mathbf{x_i}) - \mathbf{x_i} \tag{4}$$

where $\mathbf{x_i}$, $\mathbf{x_j}$ are feature vectors and $i < j$ are turbulence levels (shown above each t-SNE plot). Subtracting $\mathbf{x_i}$ the second time has the effect of moving the trajectory vector to the origin. We note that without moving the trajectory vector to the origin, the vectors do not cluster as well. The t-SNE plot of trajectory vectors is shown in Figure 4.

In Figure 4, we show that at low turbulence levels, trajectories have identity-specific trajectories. However, as turbulence levels increase, feature trajectories become more similar, eventually to the point that feature trajectories from all identities form a single central cluster. This result implies that at low turbulence levels, identity-specific face features

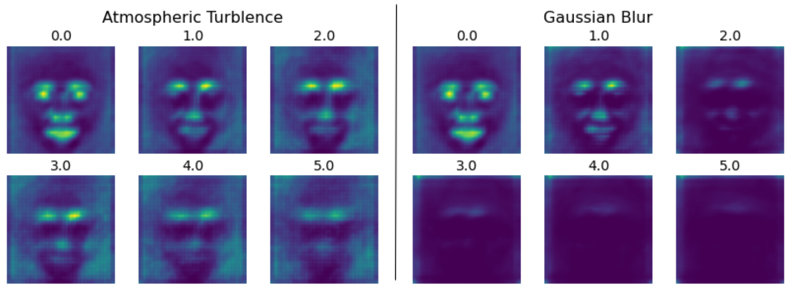

Figure 6. Input activation maps from the recognition model—generated with Score-CAM [23] and averaged over 1,000 images. Five perturbation strengths are shown for Atmospheric Turbulence and Gaussian blur. Activations are from Block 2 out of 4 in our ResNet, which allow us to visualize lower level features extracted by the model. A surprisingly stark difference can be seen between AT and GB, at higher perturbations. We attribute this to *feature defection*. While blurring removes feature activation, AT can perturb feature locations resulting in moderate activations in the map, which impact identity matching.

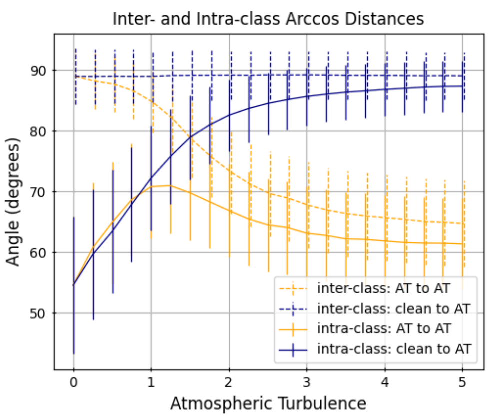

Figure 7. Inter- and intra-class angular distances. Blue lines represent distances between AT samples and clean samples, and orange lines represent distances between samples of matching AT levels. Distances shown are averages over the three datasets listed in Section 3.4.

(e.g., pupils, eyebrows) are lost, causing each identity to follow a unique path from the original class position to the origin (decreasing low-AT magnitudes shown in Section 4.2).

Comparatively, at higher turbulence levels, Figure 4 t-SNE plots show that all feature vectors, regardless of class, begin to follow the same feature space trajectory. It can be observed that identity-specific trajectories end at the same AT levels that feature magnitude begins to increase (Figure 3 Plot 1). This point is approximately $AT = 1.5$.

5.2. Inter- and Intra-Class Distance

In Section 5.1, we show that at low turbulence levels there are identity specific trajectories, and at high-turbulence levels all samples have similar trajectories. To better understand the global effect of trajectories on feature space structure, we empirically show inter- and intra-class distances in Figure 7. In Figure 7, we show the distance between samples with matching AT levels (AT-to-AT), and the distance between clean samples and AT samples (clean-to-AT). First we discuss AT-to-AT then clean-to-AT.

For samples unaffected by turbulence ($AT = 0$), Figure 7 shows relatively small intra-class distances and large inter-class distances, which corresponds with high recognition accuracy. Between $AT = 0$ and $AT = 1$, Figure 7 shows increasing geodesic distance for AT-to-AT distances. However, at $AT > 1.5$, AT-to-AT intra-class distance begins to *decrease*. At the same time, AT-to-AT inter-class distance also decreases. This leads to similar intra- and inter-class distances at high AT, which corresponds with low recognition performance. Additionally, low inter- *and* intra-class distances suggest that all samples exist in a single sub-space and create a high-AT superclass. This result corresponds with Figure 4, where we show that at higher levels of AT, all samples are following similar trajectories.

In comparison to AT-to-AT distances, clean-to-AT inter-class distances remain high, while clean-to-AT intra-class distances increase monotonically. This shows that clean samples have large geodesic distances from AT samples. This is true even for intra-class distances, thus AT samples are far from their original class-centers.

5.3. Interpretation of Results

In this section, we bring together previously presented results. In our results, a point of interest is at $AT = 1.5$, which is an inflection point for all of the following measurements: magnitude, trajectory similarity, and AT-to-AT class distance. Additionally, at this point, there is the lowest correlation between accuracy and magnitude—meaning validation accuracy is dropping quickly while magnitude is beginning to increase (Figure 3 Plot 3).

At approximately $AT = 1.5$, the increase in magnitude suggests the model is finding features representative of increased quality. At the same time, feature trajectories become similar, leading to lower AT-to-AT inter-class distances. Lower AT-to-AT inter-class distances suggest new, similar features are appearing in all images. It follows that: *the face recognition model misinterprets atmospheric turbulence artifacts as salient features for identification*. As discussed in Section 1, we call this effect *feature defection*, where defected (deformed, blurred, magnified) image features remain salient for model behavior. *Feature defection* is based on observations of increasing quality assessment, decreasing recognition performance, and the existence of the high-AT superclass.

In Figure 5, we overview the transformation caused by atmospheric turbulence. In Figure 5 diagram **a**, a clean feature space is shown with high inter-class separation and intra-class compactness. In diagram **b**, the effect of low turbulence is shown—decreased feature magnitudes and class distributions begin to overlap, causing performance to being to drop. In diagram **c**, we show the effect of *feature defection*, which occurs at approximately $AT > 1.5$. Under *feature defection*, most feature space structure is lost, as samples follow similar trajectories into a shared sub-space of feature space, which is far the original class centers. In the next section, we visualize *feature defection* with input activation maps.

6. Input Activation Maps

In order to visualize the the effects of *feature defection*, we show input activation maps. To calculate input activations, we use the Score-CAM method [23]. In Figure 6, we show average activations at Block 2 out of 4 from the Resnet backbone of our face recognition model. By looking at Block 2 activations, we are able to understand what lower level features are being extracted by the model. Figure 6 shows a significant difference between features extracted under Gaussian blur and atmospheric turbulence. Many noisy features are extracted under high-AT, and greater average activation can bee seen (i.e., more yellow). Comparatively, few features are extracted under high Gaussian blur. Under AT, there is significant activations around the eyebrow region, which indicate eyebrows are a feature commonly affected by *feature defection*, which results in misclassifications.

In Figure 8, we show sample images with Block 4 activations. Block 4 activations are an approximation of the regions of the input the model used most for classification. In the left columns, we show samples degraded by $AT = 0.75$. Activations at $AT = 0.75$ mostly surround identifiable face features. However, these sample have low quality assessment due to the $AT = 0.75$ degradations.

In the right columns of Figure 8, we show samples degraded by $AT = 3.0$. Bright activation regions can be seen around deformed edges in the face image (e.g. eyebrows, hairline, mount). Despite the greater deformations and misguided input activations, the samples at $AT = 3.0$ have greater magnitudes than at $AT = 0.75$. This demonstrates the essence of *feature defection*: defected features interpreted as salient regions for identity classification.

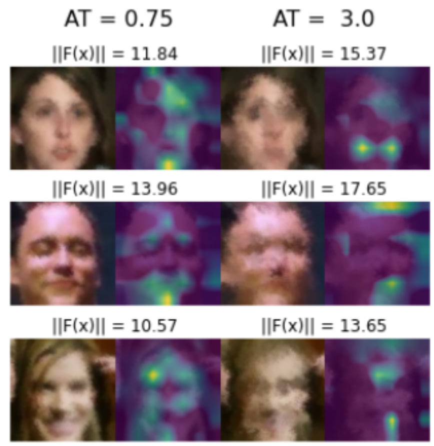

Figure 8. Sample images under atmospheric turbulence levels $AT = 0.75$ and $AT = 3.0$, shown with Block 4 activation maps. Samples shown have an increase in feature magnitude ($||F(x)||$) from $AT = 0.75$ to $AT = 3.0$, which is shown above each image. Bright yellow regions indicate greater activation. It should be noted that the amount of input activation (i.e., quantity of bright yellow) does not indicate magnitude.

7. Conclusion & Future Work

In this work, we have studied the effect of atmospheric turbulence on the feature space of deep face recognition. We present a surprising result where feature magnitudes increase under certain levels of atmospheric turbulence. Based on our results, we create a more complete view of the transformation caused by atmospheric turbulence in deep feature space. We identify *feature defection*—where the recognition model misinterprets AT artifacts as salient features for identification—as a cause of lower recognition performance and unexpected feature magnitudes.

Based on our results, we outline two challenges for future work. The first is magnitude-accuracy alignment, which, if improved, has the potential to be a meaningful quality assessment for FR under AT. The second is inter-class separation, as AT-samples collapse into a shared subspace. In future work, we plan to address these issues by designing a training procedure that optimizes for both desired feature space behavior and recognition performance under AT. We also note that a significant limitation of current work on atmospheric turbulence is that both training and testing are on simulated data with no reference to ground truth AT, and research validating simulated data is a part of future work.

References

[1] Samik Banerjee and Sukhendu Das. Lr-gan for degraded face recognition. *Pattern Recognition Letters*, 116:246–253, 2018. 3

[2] G Chen, P Peng, X Wang, and Y Tian. Adversarial reciprocal points learning for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 2, 4

[3] Grigorios G Chrysos and Stefanos Zafeiriou. Deep face deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 69–78, 2017. 3

[4] 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. 1, 4, 5

[5] Akshay Raj Dhamija, Manuel Günther, and Terrance Boult. Reducing network agnostophobia. *Advances in Neural Information Processing Systems*, 31, 2018. 2, 4, 5

[6] Biying Fu, Cong Chen, Olaf Henniger, and Naser Damer. A deep insight into measuring face image utility with general and face-specific image quality metrics. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 905–914, January 2022. 4, 5

[7] Brian Heflin, Brian Parks, Walter Scheirer, and Terrance Boult. Single image deblurring for a real-time face recognition system. In *IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society*, pages 1185–1192, 2010. 3

[8] 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. 4

[9] Samil Karahan, Merve Kilinc Yildirum, Kadir Kirtac, Ferhat Sukru Rende, Gultekin Butun, and Hazim Kemal Ekenel. How image degradations affect deep cnn-based face recognition? In *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–5, 2016. 3

[10] Martin Knoche, Stefan Hormann, and Gerhard Rigoll. Cross-quality lfw: A database for analyzing cross-resolution image face recognition in unconstrained environments. In *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, pages 1–5. IEEE, 2021. 1

[11] 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. 2, 3

[12] Ce Liu, Heung-Yeung Shum, and William T Freeman. Face hallucination: Theory and practice. *International Journal of Computer Vision*, 75(1):115–134, 2007. 3

[13] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 212–220, 2017. 5

[14] 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. 3

[15] 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. 1, 2, 4, 5

[16] 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. 4

[17] P Jonathon Phillips and J Ross Beveridge. An introduction to biometric-completeness: The equivalence of matching and quality. In *2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–5. IEEE, 2009. 6

[18] Abhijith Punnappurath, Ambasamudram Narayanan Rajagopalan, Sima Taheri, Rama Chellappa, and Guna Seetharaman. Face recognition across non-uniform motion blur, illumination, and pose. *IEEE Transactions on image processing*, 24(7):2067–2082, 2015. 3

[19] Michael C Roggemann, Byron M Welsh, and Bobby R Hunt. *Imaging through turbulence*. CRC press, 1996. 2, 3

[20] 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. 4

[21] Ziyi Shen, Wei-Sheng Lai, Tingfa Xu, Jan Kautz, and Ming-Hsuan Yang. Deep semantic face deblurring. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8260–8269, 2018. 3

[22] 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. 5

[23] Haofan Wang, Zifan Wang, Mengnan Du, Fan Yang, Zijian Zhang, Sirui Ding, Piotr Mardziel, and Xia Hu. Score-cam: Score-weighted visual explanations for convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 24–25, 2020. 7, 8

[24] Wenhan Yang, Ye Yuan, Wenqi Ren, Jiaying Liu, Walter J Scheirer, Zhangyang Wang, Taiheng Zhang, Qiaoyong Zhong, Di Xie, Shiliang Pu, et al. Advancing image understanding in poor visibility environments: A collective benchmark study. *IEEE Transactions on Image Processing*, 29:5737–5752, 2020. 3

[25] 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. 2, 3