# Contrastive Learning Enhanced Diffusion Model for Improving Tropical Cyclone Intensity Estimation with Test-Time Adaptation

Ziheng Zhou, Haojia Zuo, Ying Zhao(✉), and Wenguang Chen

Tsinghua University, Beijing, China
{zhouzihe18,zuohj19}@mails.tsinghua.edu.cn, {yingz,cwg}@tsinghua.edu.cn

**Abstract.** Tropical cyclone (TC) intensity estimation from satellite images is the very first and critical step of making TC forecasts, whose SOTA performance is achieved by methods built upon CNN based regression models. Unlike discriminative models trained for specific tasks, generative models on the other hand learns to comprehend data in a more sophisticated way through generation. In this paper, we explore the potential of using generative models to further improve the regression task of TC intensity estimation, distinguished from precedents that aim at classification tasks. Our proposed method ConDiff-RTTA optimizes a TC regression model during test time, by back-propagating the loss of a diffusion model conditioned on the regression outputs. More importantly, by enhancing the diffusion model's training process with our proposed contrastive loss, the diffusion model is more likely to align diffusion losses with prediction errors of the regression model. This enhancement leads to a better understanding of incorrect conditions which facilitates the adaptation of the regression model. We evaluate our proposed method on a benchmark dataset TCIR, where TCs of the latest two years are used as testing cases. Experimental results show that our proposed method ConDiff-RTTA improves the regression model in overall performance, especially on high intensity tropical cyclones. Our code is publicly available at https://github.com/maxmaxcu/ConDiff-RTTA/.

**Keywords:** Tropical cyclone · Intensity estimation · Diffusion Models · Test-time adaptation · Regression · Contrastive learning

## 1 Introduction

Tropical cyclones (TC) are among the most catastrophic weather events that can cause injuries and deaths as well as huge economic losses. Tropical cyclone monitoring and forecasting are among the most concerned missions for meteorologists and weather service centers worldwide. The very first and critical step of making TC forecasts is intensity estimation, which is defined as the maximum sustained


A. Bifet et al. (Eds.): ECML PKDD 2024, LNAI 14949, pp. 418–434, 2024.
https://doi.org/10.1007/978-3-031-70378-2_26

surface wind speed near the TC center (measured in knot, 1 kt $\approx$ 0.51 $ms^{-1}$). Since tropical cyclones usually occur on the open ocean, satellite images are mostly used for estimating the intensity. Traditional methods, such as Dvorak [8], DAV [22], and ADT [21] are based on cloud patterns recognized from satellite images. Recently, many efforts have been made in developing Neural Network (NN) based models [1–3,6,10,24,29,31] for the task of TC intensity estimation, which has become a promising direction to achieve more accurate estimations. All of these models are discriminative models that are inspired by the ability of automatically learning useful features from satellite images with various network architectures, data pre-processing methods or physics guided feature extractions. Backbone models of these works are often CNN based regression models, which predict numerical TC intensities directly.

Alternative to discriminative models, generative models are trained on a harder task, forcing them to learn a deeper and more sophisticated comprehension of the data so as to synthesize new samples, thereby improving their potential for discriminative tasks especially under limited data [11,20]. Inspired by the recent advancements of diffusion models that show promising ability in synthesizing high quality images following class or text conditions [7,12–14,26–28], there emerge a number of studies [4,5,17,18,23] that aim to unleash the potential of diffusion models on discriminative tasks. Among them, Diffusion-TTA [23] uses a pre-trained conditional diffusion model to tune an image classifier during test time and observes improvements on accuracy over the original classifier. The two models are attached in a way that, the classifier output serves as the condition to the diffusion model, such that the classifier can be adapted by back-propagating the diffusion loss.

It is natural to utilize diffusion models in a similar way on regression tasks to achieve improved TC intensity estimations. Unlike classification tasks, where the predicted attributes are categorical, regression tasks predict ordinal and numerical attributes, and face additional challenges. To successfully tune a regressor in a gradient descent manner, it should hold that given a biased prediction, the gradient of the diffusion loss on the condition points to the direction toward the ground truth, considering the ordinal nature of attributes to be regressed on. Existing studies like Diffusion-TTA focus on classification tasks and have not yet inspected into this problem. Furthermore, the level of diffusion loss should be connected to the degree in which the condition is biased, so as to encourage the expected gradient. However conditional diffusion models are typically trained with only correct conditions, lacking penalties for the incorrect ones let alone such "distance awareness", which could result in sub-optimal results.

In this paper, we propose a method driven by a contrastive learning enhanced diffusion model that meets the aforementioned challenges and can better resolve the tropical cyclone intensity estimation task. The main contributions of this paper are the following:

1. We propose a test-time adaptation method **Diff-RTTA** to improve performances of regression models utilizing diffusion models, and observe favorable loss characteristics that lead adaptations towards more accurate predictions.

2. We propose the **ConDiff-RTTA** method to enhance the diffusion model with contrastive learning such that it is aware of the distance between true and false conditions, which further optimizes the model to be more aligned with the regression task.
3. We conduct experiments on a benchmark dataset of TC intensity estimation and observe performance gains with our method, especially on high intensity tropical cyclones.

The rest of this paper is organized as follows. We first give a brief overview of related work in Sect. 2. Then, we introduce the preliminary knowledge on both diffusion models and test-time adaptation with diffusion models, and propose our constractive learning enhanced diffusion models in Sect. 3. Experimental results of our proposed method on TC benchmark dataset TCIR are shown in Sect. 4. Finally, we make concluding remarks in Sect. 5.

## 2 Related Work

**Neural Network Models for TC Intensity Estimation.** Neural network based models for the TC intensity estimation problems fall into two categories in terms of their outputs, i.e., classification models and regression models. Classification models, e.g. [10,24], output TC categories or TC intensity ranges instead of the numerical intensity value, whose performance is inferior to that of regression models [1–3,6,29,31] in terms of estimation accuracy in RMSE or MAE. For regression models, recent works mostly focus on physics guided methods, through using extra data or features as inputs [2,31,32], or designing loss functions with TC knowledge [29,31]. Some works also focus on the network design [1,2], suggesting that the neural network should not be too deep and need to exclude dropout layers.

**Diffusion Generative Models for Discriminative Tasks.** There have been continuing attempts in aiming to unleash the potential of generative models on discriminative tasks, dated back to early studies [11,20,25]. With recent advancements in diffusion models, a number of works [4,5,17] face this challenge by sharing the idea that, a mildly noised image should be denoised by a diffusion model with the best effect when given the correct condition. In this light, they transform either class-conditional or text-to-image diffusion models to image classifiers by enumerating through classes and converting their corresponding diffusion losses to class probabilities. Diffusion models can also be seen as teacher models to optimize dedicated discriminative student models. DreamTeacher [18] distills knowledge from generative models pre-trained on large datasets onto a discriminative backbone, which is later trained on small downstream datasets. Diffusion-TTA [23] back-propagates the diffusion loss to a classifier, allowing test-time adaptation to improve the classification accuracy of the discriminative model. Our work is more similar to the latter than the former as we target at TC intensity estimation, an area where data are limited due to the fact that satellite images of TCs are only available since past few decades and have to be

fully used by both generative and discriminative models, not allowing for an up- and down-stream split.

**Contrastive Learning to Capture Data Divergence.** By contrasting positive samples with negative ones, contrastive learning serve as a powerful tool to capture various forms of divergence in the data. Such divergence could be data mismatching, label differences, or even the precise distances between values. For instance, SupCon [15] projects data to positions in the embedding space according to their class labels, and Rank-N-Contrast [30] further extends the idea to continuous label values, making embeddings repel each other in a degree of their label distances. CoDi [16] aims to generate tabular data entries which consist of both continuous and discrete parts by two co-evolving diffusion models, and penalizes mismatching between the two parts by utilizing contrastive learning. Inspired by these works, we use contrastive learning to make our diffusion model not only able to capture image-condition mismatching, but also be "distance aware" of correct and biased conditions.

## 3 Methodology

### 3.1 Preliminaries

**Diffusion Models**. For an image $x$ sampled from the real data distribution $x \sim q(x)$, a diffusion model learns to approximate the data distribution by gradually adding noise to $x$ in the diffusion process and predict the noise in the reverse process. Conditional diffusion models further learns the distribution $q(x|c)$, where $c$ is the condition input corresponding to the image $x$. The diffusion process, where a sequence of noise are added to the original input image $x$ (now denoted as $x_0$) generating a noised image sequence $x_1, x_2, ..., x_T$, is formally defined [12] as:

$$\begin{aligned} q(x_{1:T}|x_0) &:= \prod_{t=1}^{T} q(x_t|x_{t-1}), \\ q(x_t|x_{t-1}) &:= N(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t \mathbf{I}), \end{aligned} \tag{1}$$

where $\beta_1, ...\beta_T$, is a variance schedule that controls the level of the noise. We can further sample $x_t$ from $x_0$ using

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, \epsilon \sim \mathcal{N}(0, \mathbf{I}), \tag{2}$$

where $\alpha_t := 1 - \beta_t$ and $\bar{\alpha}_t := \prod_{s=1}^{t} \alpha_s$

A diffusion denoising network $\epsilon_\phi(x_t, t)$ learns to predict the noise with noisy image $x_t$ and the noise level $t$ as inputs. For conditional diffusion models that takes $c$ as an input condition during the reverse process, the diffusion loss for training is defined as:

$$\mathcal{L}_{\text{diff}}(\phi; \mathcal{D}) = \frac{1}{|\mathcal{D}|} \sum_{(x^i, c^i) \in \mathcal{D}} \| \epsilon_\phi(\sqrt{\bar{\alpha}_t}x^i + \sqrt{1-\bar{\alpha}_t}\epsilon, c^i, t) - \epsilon \|^2 \tag{3}$$

where $\mathcal{D} = \{(x^i, c^i)\}_{i=1}^N$ is a training batch of $N$ images with their corresponding conditions (labels).

Note that for the sake of simplicity, the above formulations are from DDPM [12], an origin of diffusion models. In our work we follow the framework of EDM [14] which includes altered design choices that boost the generative ability.

**Test-Time Adaptation with Diffusion Models**. Test-time adaptation refers to a procedure in which a pre-trained model is adapted on *unlabeled* test data [19]. Without labels, what is helpful to the adapted model can be another model that contains better knowledge about the test data. Diffusion-TTA [23] tackles this by using a pre-trained diffusion model, and tune an image classifier in an iterative manner. First, the classifier does inference on an image to provide an initial guess of class probabilities, from which a class condition is synthesized as weighted mixing of class embeddings. Then, a noise batch of different strengths is added onto the image, as inputs into the diffusion model along with the synthesized condition to compute the conditional diffusion loss. Last, loss gradients are back-propagated to the classifier, updating it to produce new class probabilities for the next iteration. After a specified number of iterations, the classifier is optimized on the image sample to produce a more accurate classification result with the help of the diffusion model, yielding better performance on the test set.

### 3.2 Conditional Diffusion Model for a Regression Task

Existing works that use diffusion models in discriminative tasks are limited to using categorical conditions such as one-hot class labels or text embedding during the training and inference of diffusion models. This raises a direct question that whether regression tasks can benefit from conditional diffusion models as well. In our TC intensity estimation task, the intensity is a numerical number with its range from 10 kt to 180 kt. Given the continuous nature of labels in regression tasks, it is infeasible to build a generative regressor by enumerating through labels as conditions and infer the target from corresponding conditional diffusion losses as in [4,17]. Therefore we build our method on top of Diffusion-TTA which is gradient based.

Towards this goal, we migrate Diffusion-TTA to TC intensity estimation in a simple yet effective fashion. We follow the process of Diffusion-TTA and make modifications to take the TC intensity value as the condition, instead of class text embedding as in Diffusion-TTA. First we train a conditional diffusion model on an open dataset of TCs (will be described in Sect. 4.1), where the intensity condition is passed through a linear layer, projected to an embedding vector and taken by the diffusion model. Then we take a CNN-based TC intensity regression model [31] and conduct TTA on it in an instance-wise manner. We denote this method as **Diff-RTTA**, whose overall architecture and pseudo code are shown in Fig. 1 and Algorithm 1, respectively.

We see improvements on Diff-RTTA over the regression model (reported in Sect. 4.4), but in this phase of study what we mainly want to inspect is the reason

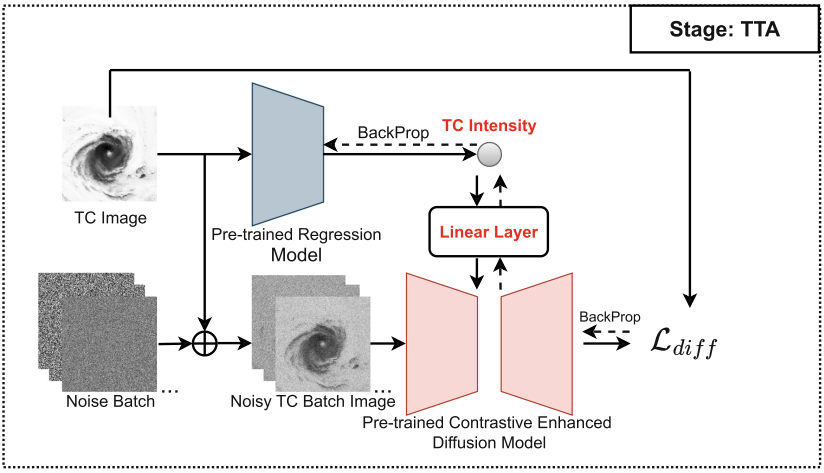


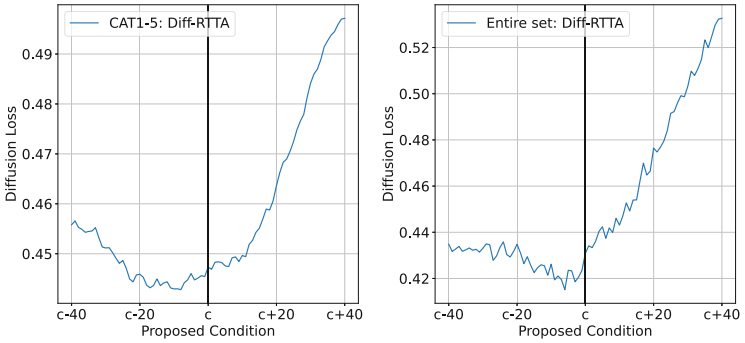

**Fig. 1.** Overall Architecture for Test-time Adaptation

**Algorithm 1** Test-time Adaptation

**Require:** Test image $x$, regression model weights $\theta$, diffusion model weights $\phi$, adaptation steps $N$

**for** $s \in [1, N]$ **do**
  Do inference on regression model to get prediction $\hat{c} \leftarrow f_\theta(x)$
  Project $\hat{c}$ to embedding $\boldsymbol{e}_{\hat{c}}$ by the linear layer
  Sample noise strength batch $\boldsymbol{t}$ following settings of Diffusion-TTA
  Sample noise batch $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$
  repeat $x$ to build batch $\boldsymbol{x}$
  Compute $\mathcal{L}_{\text{diff}} = \| \epsilon_\phi(\boldsymbol{x} + \boldsymbol{t} \odot \boldsymbol{\epsilon}, \boldsymbol{e}_{\hat{c}}, \boldsymbol{t}) - \boldsymbol{\epsilon} \|^2$
  Take gradient descent step on $\nabla_\theta \mathcal{L}_{\text{diff}}$ to update $\theta$
**end for**
**return** $f_\theta(x)$

why a diffusion model can indeed benefit regression tasks. To demonstrate it, for every TC image we enumerate the intensity condition as an integer from 10 kt to 180 kt, and collect diffusion losses over the enumeration. It is expected that the diffusion loss should be minimal at the correct intensity of the TC image. Figure 2 shows the loss enumerations on test set for TCs of CAT1-CAT5 categories (well be defined in Sect. 4.1) and the entire set. It can be observed that the loss curves tend to be U-shaped with the valley near the correct condition (denoted by $c$). With the U-shaped loss curves, it is made possible that a biased proposed value of intensity could be optimized towards the ground truth intensity by steps of gradient descent, whereas enumeration on possible conditions is far more costly for continuous values.




(a) CAT1-CAT5 (b) Entire Set

**Fig. 2.** Diffusion loss enumerations over conditions by Diff-RTTA: For each TC image, diffusion losses are calculated on each condition enumerated from range [$c$-40,$c$+40], where $c$ is the true condition of the corresponding TC image. The average of diffusion losses on each condition offset value of all TC images from CAT1 to CAT5 categories and from the entire test set are shown in (a) and (b), respectively.

### 3.3 Contrastive Learning Enhanced Diffusion Model

Observations on Diff-RTTA indicate that, by following the vanilla training procedure, a diffusion model conditioned on numerical values can exhibit our expected characteristic: the loss enumeration curve is U-shaped around the true condition. In other words, the diffusion model denoises noisy images the best around the true condition, and behaves worse when the proposed condition is farther away. Nevertheless, we suppose this favorable characteristic can be even strengthened, since the vanilla training way of the conditional diffusion model assumes the conditions are always correct, thus paying no attention on the relation between true and false conditions and their distances. It is reasonable because such knowledge can hardly be of use for pure generation, but it comes to importance in the context of our study. We expect that explicitly relating diffusion loss to condition distances can point the gradient more to the correct direction, and reduce the bias between the loss minimum point and correct condition.

Similar ideas can be seen in contrastive learning literature such as [30]. This motivates us to explore supervised contrastive learning for the enhancement. Contrastive learning works by contrasting similar samples (positive samples) with dissimilar ones (negative samples). In the TC estimation scenario, for a TC image $x$ with its true condition $c$, a positive-negative pair is defined as:

$$\begin{aligned} \text{Positive} &: [\text{aug}(x), c_{\text{pos}}], \\ \text{Negative} &: [\text{aug}(x), c_{\text{neg}}], \end{aligned} \tag{4}$$

where $\text{aug}(\cdot)$ is a data augmentation function, $c_{\text{pos}} := c$ and $c_{\text{neg}}$ is a false condition not equal to $c_{\text{pos}}$. Concerning the ordinal nature of our conditions and the local gradient we pursue, we sample the negative condition in a neighborhood of

the positive, which also serves as a harder negative compared to some arbitrarily positioned one. There are also common observations that regression models tend to exhibit larger estimation errors on TCs of high intensities [1,2,29,31,32], therefore we enlarge the sampling neighborhood for high intensities to cover the potential error bar with negative samples. The sampling strategy is defined as

$$c_{\text{neg}} = c_{\text{pos}} + \text{rand}(-(\log c_{\text{pos}})^2, (\log c_{\text{pos}})^2), \tag{5}$$

where $\text{rand}(a, b)$ draws a random number from a uniform distribution on the interval $(a, b)$.

With the defined positive-negative pair, we propose a contrastive loss term in the form of a triplet loss, which is formally defined as

$$\mathcal{L}_{\text{con}} = \max\left(\mathcal{L}_{\text{diffpos}} - \mathcal{L}_{\text{diffneg}} + \text{Margin}(c_{\text{pos}}, c_{\text{neg}}), 0\right), \tag{6}$$

where $\mathcal{L}_{\text{diffpos}}$ is the diffusion loss for the positive sample and $\mathcal{L}_{\text{diffneg}}$ is the diffusion loss for the negative sample. In the standard triplet loss, margin is defined as a constant to keep the positive away from the negative in a certain degree. Here, we propose margin as a distance aware function so that it adjusts the margin between positive and negative losses according to the distance between the corresponding conditions. With a larger distance, the negative loss should exceed the positive loss to a greater extent. The Margin function is defined as

$$\text{Margin}(c_{\text{pos}}, c_{\text{neg}}) = \log\left(1 + D(c_{\text{pos}}, c_{\text{neg}})\right) * \mathcal{L}_{\text{diffpos}}, \tag{7}$$

where $D(c_{\text{pos}}, c_{\text{neg}})$ is the distance between true and false conditions, and $\mathcal{L}_{\text{diffpos}}$ here only provides the value without contributing a gradient. We choose the current form to let the margin shrink when $c_{\text{neg}}$ gets close to $c_{\text{pos}}$. The margin is also proportional to $\mathcal{L}_{\text{diffpos}}$ because the loss scale differs through conditions and the margin should be adjusted in a relative manner. With this distance aware margin, the diffusion model learns to increase the diffusion loss under a false condition adaptive to the condition distance and the loss scale.

We propose the following contrastive learning enhanced diffusion loss for continuous training on the previously trained diffusion model,

$$\mathcal{L}_{\text{ConDiff}} = \mathcal{L}_{\text{diff}} + \lambda \mathcal{L}_{\text{con}} \tag{8}$$

where $\lambda$ is the weight for the contrastive loss, which is a hyper parameter. The training procedure is modified from the standard procedure of training a conditional diffusion model, where in each iteration the batch is doubled to construct the negative half whose conditions are sampled according to Eq. 5, and the doubled batch is fed into the model to update it via $\mathcal{L}_{\text{ConDiff}}$. The contrastive learning enhanced diffusion model is then used in the TTA stage. We denote this improved method as **ConDiff-RTTA**.

The pipeline for the contrastive enhanced diffusion model training phase is shown in Fig. 3. The overall pseudo code for training is shown in Algorithm 2.

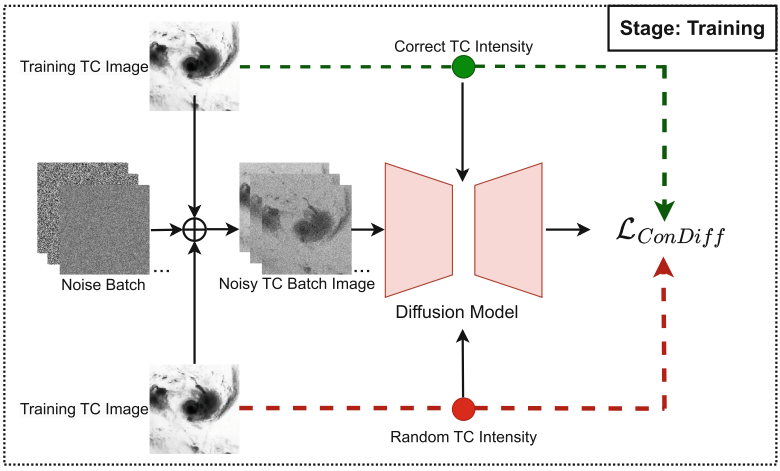



**Fig. 3.** Overall Architecture for Constructive Enhanced Diffusion Model Training

**Algorithm 2** Training of Contrastive Learning Enhanced Diffusion Model

**Require:** training set $\mathcal{D}_{\text{train}}$, diffusion model weights $\phi$
1: **while** not converged **do**
2: Sample training image-intensity batch $(\boldsymbol{x}_0, \boldsymbol{c}_{\text{pos}}) \sim \mathcal{D}_{\text{train}}$
3: Sample noise strength batch $\boldsymbol{t}$ following settings of EDM
4: Sample noise batch $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$
5: $\boldsymbol{x}_t \leftarrow \boldsymbol{x}_0 + \boldsymbol{t} \odot \boldsymbol{\epsilon}$
6: Synthesize false conditions $\boldsymbol{c}_{\text{neg}}$ according to Eq. 5
7: Compute $\mathcal{L}_{\text{ConDiff}}$ according to Eq. 8
8: Take gradient descent step on $\nabla_\phi \mathcal{L}_{\text{ConDiff}}$ to update $\phi$
9: **end while**

## 4 Experiments

### 4.1 Dataset

We use a publicly available benchmark dataset, the Tropical Cyclone Dataset for Image Intensity Regression (TCIR)[1] [1]. TCIR contains TCs in the North Eastern Pacific, the North Western Pacific, and the Atlantic Ocean. The satellite observations in TCIR are derived from two open datasets, GridSat and CMORPH. The best track intensities (IBTrACS) are derived from the Joint Typhoon Warning Center (JTWC) and the Atlantic Hurricane Database (HURDAT2).

As shown in Table 1, we classify TCs according to the Saffir-Simpson Hurricane Wind Scale, which consists of 7 classes, with higher classes representing higher maximum sustained winds. We use a total of 36566 image frames from TCs in 2003-2013 as training data, 3245 from TCs in 2014 as validation data,

[1] Available at https://www.csie.ntu.edu.tw/~htlin/program/TCIR/.

and 7570 frames from TCs in 2015-2016 as testing data. Each frame has $201 \times 201$ pixels and a total of 4 channels per pixel, i.e., infrared (IR), water vapor (WV), visible channel (VIS), and passive microwave rain-rate (PMW). In our experiments, we use the IR channel, and normalize it to have zero mean and unit standard deviation, and resize it to $65 \times 65$ pixels as the input.

**Table 1.** Number of Frames in TCIR from [1]

| Category | Training | Validation | Testing |
|---|---|---|---|
| TD (33≤kt) | 13766 | 1154 | 2353 |
| TS (34∼63 kt) | 13850 | 1194 | 3048 |
| CAT1 (64∼82 kt) | 3793 | 388 | 787 |
| CAT2 (83∼95 kt) | 1909 | 178 | 490 |
| CAT3 (96∼112 kt) | 1381 | 129 | 418 |
| CAT4 (113∼136 kt) | 1558 | 147 | 394 |
| CAT5 (≥137 kt) | 309 | 55 | 80 |
| Total | 36566 | 3245 | 7570 |

### 4.2 Models and Metrics

**Regression Model**. To achieve SOTA performance on TC intensity estimation, it is needed to include physics-guided features in the regression network, and conduct special post-processing such as sliding windows and rotation ensembles [1,2,29,31,32]. These techniques are along different dimensions compared to our method, and will require a lot of extra efforts and computational resources. Therefore we set our goal to explore the ability of diffusion models on improving the intensity estimation performance of CNN based backbone models. We use ResNet-18 [9] as the backbone for feature extraction and train it on the TCIR training set with L2 loss. This regression model achieves comparable performance to backbone models in [1,31] on the TCIR validation set and test set. We refer to this model as Regression or Reg Model in the experiments.

**Diffusion Models**. For diffusion models, we use the implementation framework of EDM [14] and the U-Net backbone model from [28]. **Diff-RTTA**: This is a diffusion model trained with our modification for the regression task as discussed in Sect. 3.2. The trained diffusion model is then used to adapt the Reg Model during test time. **ConDiff-RTTA**: We fine-tune the above diffusion model with our proposed $\mathcal{L}_{\text{ConDiff}}$ loss. Same as in Diff-RTTA, $\mathcal{L}_{\text{diff}}$ is used during TTA stage.

**Evaluation Metrics**. We report the TC intensity estimation accuracy of various models in terms of Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

### 4.3 Implementation Details

Models are trained on 8 RTX 4090 GPUs. We train a total of 21M TC images randomly sampled from the training dataset with batch size 256 for the Diff-RTTA and continue to train 3M TC images with batch size 128 for the ConDiff-RTTA, with the rest of training settings following the default of EDM. For test-time adaptation, the noise batch size is 200 with 10 adaptation steps and Adam optimizer is used with a learning rate of $5 \times 10^{-5}$.

### 4.4 Overall Performance

**Diff-RTTA as Regression Model**. To get a better understanding on using diffusion model alone as a regression model, we test the performance of Diff-RTTA model without using the pre-trained Reg Model but instead with 50 kt as the initial conditional inputs for all the test TC images. 50 kt is the mean value of the TC intensities from training set and has an overall RMSE of 30.39 on the entire test set. The performance of using 50 kt as initial conditions with Diff-RTTA is shown in Table 2 labelled as Diff-RTTA (50). Even with the initial condition of 50 kt, the overall RMSE of Diff-RTTA improves to 14.83, showing its ability as a regression model. In Fig. 4, Diff-RTTA (50) results for each TC image are ordered by the true conditions (from IBTrACS) from left to right. We can see that the predicted intensities are spread along the true conditions, which indicates the important fact that the correct adaptation directions are likely to be found using the diffusion loss as the feedback.

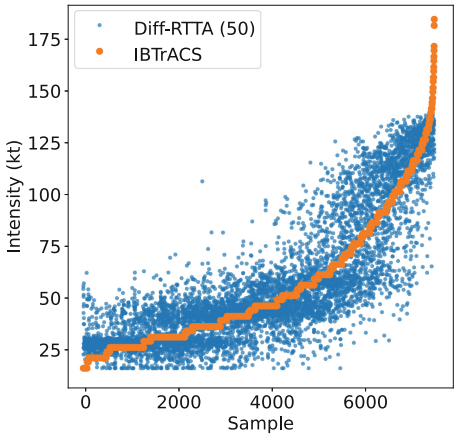



**Fig. 4.** Diff-RTTA (50) results for each TC image

**Comparisons to Baselines**. The performances of Reg Model, Diff-RTTA and ConDiff-RTTA are shown in Table 2. Diff-RTTA shows an improvement of 0.33

on RMSE over Reg Model, from 11.22 to 10.89. ConDiff-RTTA further improves the overall performance to 10.76. Although ConDiff-RTTA achieves mildly better results over Diff-RTTA in overall performance, a detailed inspection reveals that improvements on each TC category are made differently, as shown in Fig. 5. ConDiff-RTTA shows more significant improvements over Diff-RTTA as the TC intensity becomes higher, roughly between 0.6 to 1.0 compared to Reg Model on CAT1-5, in which the most destructive TCs reside. We attribute this observation to the stronger contrastive effect on high intensities due to larger contrastive margins and wider negative sampling windows, which we design deliberately to enhance the regression model's performance on strong TCs.

**Table 2.** RMSE and MAE results on TCIR test set

| | TD | | TS | | CAT1-5 | | Overall | |
|---|---|---|---|---|---|---|---|---|
| Method | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE |
| Regression | 6.76 | 4.95 | 9.89 | 7.70 | 15.76 | 12.16 | 11.22 | 8.17 |
| Diff-RTTA (50) | 9.25 | 6.91 | 12.49 | 9.28 | 21.21 | 17.12 | 14.83 | 10.84 |
| Diff-RTTA | **6.49** | **4.69** | 9.62 | 7.49 | 15.31 | 11.89 | 10.89 | 7.93 |
| ConDiff-RTTA | 6.66 | 4.80 | **9.58** | **7.43** | **14.97** | **11.57** | **10.76** | **7.85** |

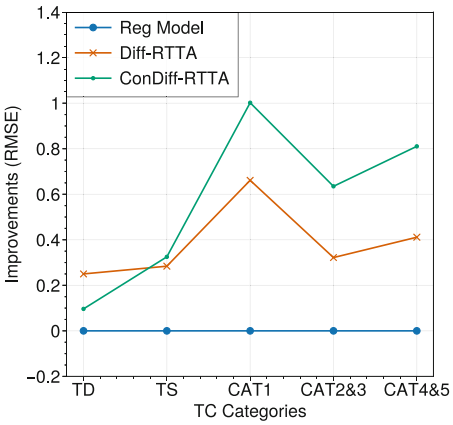
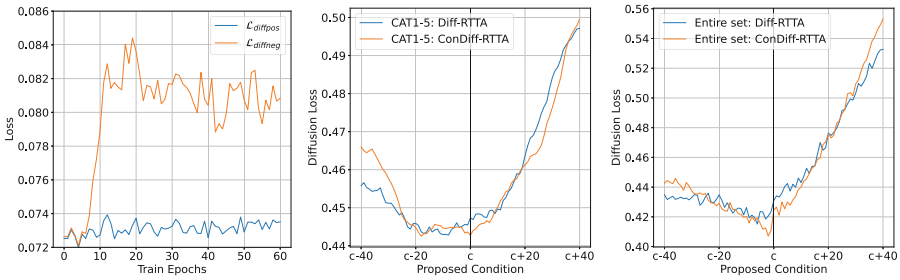



**Fig. 5.** Improvements on different categories over baseline Reg Model

### 4.5 Diffusion Loss Analysis

We show the training curves of $\mathcal{L}_{\text{diffpos}}$ and $\mathcal{L}_{\text{diffneg}}$ in Fig. 6 (a). $\mathcal{L}_{\text{diffpos}}$, the diffusion loss given true conditions, remains at a low level while $\mathcal{L}_{\text{diffneg}}$, the

diffusion loss given false conditions, increases significantly during training. Diffusion loss enumerations of the diffusion model trained in Diff-RTTA and that trained in ConDiff-RTTA are shown in Fig. 6 (b) and (c), on CAT1 to CAT5 TCs and the entire test set respectively. We can see that on both figures, the enumeration curves (yellow) with ConDiff-RTTA are sharper than the curves (blue) with Diff-RTTA. The valley of the curve with ConDiff-RTTA also shifts more towards the center (true condition $c$) compared to Diff-RTTA. These figures comply with our intention to still learn $p(x|c)$ as well as impose stronger constraints on false condition scenarios.




(a) Training loss curves (b) CAT1-CAT5 (c) Entire Set

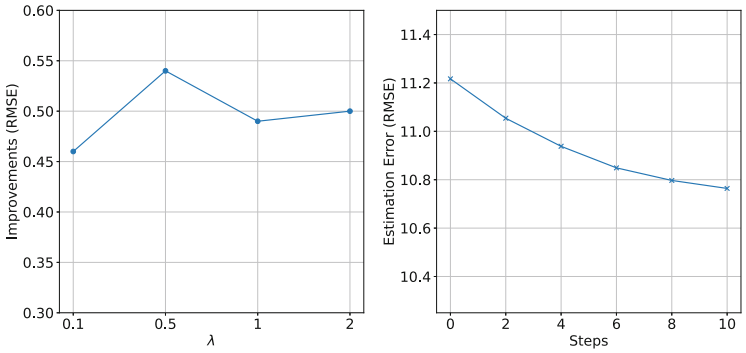
**Fig. 6.** (a) Training loss curves and diffusion loss enumerations over conditions on (b) CAT1-CAT5 and (c) the entire test set

### 4.6 Parameter Study

A parameter study is conducted using validation set on the hyper parameter $\lambda$, which is the weight of our proposed contrastive loss. Figure 7 (a) shows the overall improvements over baseline Reg Model with different $\lambda$ values of 0.1, 0.5, 1.0, 2.0. It shows that the overall performance improves even with a small $\lambda$ value. $\lambda = 0.5$ is selected according to our parameter study for reporting ConDiff-RTTA results. We also perform another parameter study w.r.t. the number of adaptation steps during test-time adaptation and the results are shown in Fig. 7 (b). It shows that by extending the adaptation steps, the overall RMSE keeps decreasing. As more adaptation steps lead to more running time, we stop the adaptation step at 10.

### 4.7 Case Study

We select from our test set the Super Typhoon Meranti, one of the most disastrous typhoons of this century for our case study. Meranti impacted South Eastern Asia and Southern China areas in September 2016, causing numerous deaths and injuries along with massive economic loss. It was recognized as a CAT5 typhoon during its peak times.




(a) Hyper Parameter $\lambda$ (b) Adaptation Steps

**Fig. 7.** Parameter study on (a) Hyper Parameter $\lambda$ and (b) Adaptation Steps

Figure 8 shows the best track intensities (from IBTrACS) and model intensity estimations of Meranti throughout its lifetime. The regression model underestimates the peak intensities, which is likely due to the rareness of violent typhoons in the nature and therefore in the TCIR dataset. As a comparison, Our proposed method ConDiff-RTTA revises the estimations upward such that they are closer to IBTrACS values. This case demonstrates that with the assist of our contrastive learning enhanced diffusion model, over-fitting in the regression model can be mitigated, resulting in a more accurate discriminative estimation on rare data.

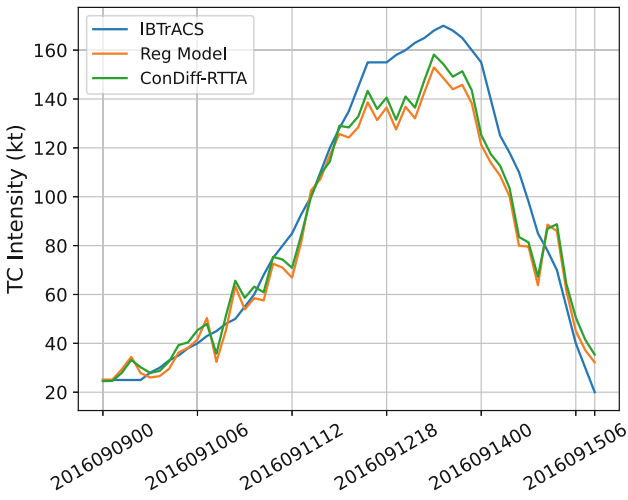



**Fig. 8.** The intensities of Super Typhoon Meranti over its lifetime

## 5 Conclusion

In this paper, we propose a new method ConDiff-RTTA to improve TC intensity estimation performance. We find that TC regression network can be optimized during test time by a diffusion model conditioned on ordinal intensity numbers instead of categorical labels as in previous works. Furthermore, we enhance the diffusion model by training in a contrastive learning approach in order to improve the alignment between diffusion losses and prediction errors of the regression model. Experimental results show that the diffusion model pre-trained from TC satellite images improves TC estimation performance, and ConDiff-RTTA achieves further overall performance gains, especially significant on high intensity TCs.

**Acknowledgments.** This work was supported by the National Key Research and Development Program of China (2022YFC3004102, 2017YFA0604502), Science and Technology Project of Qinghai Province (No.2023-QY-208), and High performance computing center of Qinghai University.

## References

1. Chen, B., Chen, B.F., Lin, H.T.: Rotation-blended cnns on a new open dataset for tropical cyclone image-to-intensity regression. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 90–99 (2018)
2. Chen, B.F., Chen, B., Lin, H.T., Elsberry, R.L.: Estimating tropical cyclone intensity by satellite imagery utilizing convolutional neural networks. Weather Forecast. **34**(2), 447–465 (2019)
3. Chen, Z., Yu, X.: A novel tensor network for tropical cyclone intensity estimation. IEEE Trans. Geosci. Remote Sens. **59**(4), 3226–3243 (2020)
4. Chen, H., Dong, Y., Wang, Z., Yang, X., Duan, C., Su, H., Zhu, J.: Robust classification via a single diffusion model. arXiv preprint arXiv:2305.15241 (2023)
5. Clark, K., Jaini, P.: Text-to-image diffusion models are zero shot classifiers. Adv. Neural Inform. Process. Syst. **36** (2024)
6. Combinido, J.S., Mendoza, J.R., Aborot, J.: A convolutional neural network approach for estimating tropical cyclone intensity using satellite-based infrared images. In: 2018 24th International Conference on Pattern Recognition (ICPR), pp. 1474–1480. IEEE (2018)
7. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. Adv. Neural. Inf. Process. Syst. **34**, 8780–8794 (2021)
8. Dvorak, V.F.: Tropical cyclone intensity analysis and forecasting from satellite imagery. Mon. Weather Rev. **103**(5), 420–430 (1975)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
10. Higa, M., et al.: Domain knowledge integration into deep learning for typhoon intensity classification. Sci. Rep. **11**(1), 1–10 (2021)
11. Hinton, G.E.: To recognize shapes, first learn to generate images. Prog. Brain Res. **165**, 535–547 (2007)

12. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Adv. Neural. Inf. Process. Syst. **33**, 6840–6851 (2020)
13. Ho, J., Salimans, T.: Classifier-free diffusion guidance. arXiv preprint arXiv:2207.12598 (2022)
14. Karras, T., Aittala, M., Aila, T., Laine, S.: Elucidating the design space of diffusion-based generative models. Adv. Neural. Inf. Process. Syst. **35**, 26565–26577 (2022)
15. Khosla, P., et al.: Supervised contrastive learning. Adv. Neural. Inf. Process. Syst. **33**, 18661–18673 (2020)
16. Lee, C., Kim, J., Park, N.: Codi: co-evolving contrastive diffusion models for mixed-type tabular synthesis. In: International Conference on Machine Learning, pp. 18940–18956. PMLR (2023)
17. Li, A.C., Prabhudesai, M., Duggal, S., Brown, E., Pathak, D.: Your diffusion model is secretly a zero-shot classifier. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2206–2217 (2023)
18. Li, D., et al.: Dreamteacher: pretraining image backbones with deep generative models. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 16698–16708 (2023)
19. Liang, J., He, R., Tan, T.: A comprehensive survey on test-time adaptation under distribution shifts. arXiv preprint arXiv:2303.15361 (2023)
20. Ng, A., Jordan, M.: On discriminative vs. generative classifiers: a comparison of logistic regression and naive bayes. Adv. Neural Inform. Process. Syst. **14** (2001)
21. Olander, T.L., Velden, C.S.: The advanced dvorak technique: continued development of an objective scheme to estimate tropical cyclone intensity using geostationary infrared satellite imagery. Weather Forecast. **22**(2), 287–298 (2007)
22. Piñeros, M.F., Ritchie, E.A., Tyo, J.S.: Estimating tropical cyclone intensity from infrared image data. Weather Forecast. **26**(5), 690–698 (2011)
23. Prabhudesai, M., Ke, T.W., Li, A., Pathak, D., Fragkiadaki, K.: Test-time adaptation of discriminative models via diffusion generative feedback. Adv. Neural Inform. Process. Syst. **36** (2024)
24. Pradhan, R., Aygun, R.S., Maskey, M., Ramachandran, R., Cecil, D.J.: Tropical cyclone intensity estimation using a deep convolutional neural network. IEEE Trans. Image Process. **27**(2), 692–702 (2017)
25. Raina, R., Shen, Y., Mccallum, A., Ng, A.: Classification with hybrid generative/discriminative models. Adv. Neural Inform. Process. Syst. **16** (2003)
26. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10684–10695 (2022)
27. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)
28. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. In: International Conference on Learning Representations (2021). https://openreview.net/forum?id=PxTIG12RRHS
29. Tian, W., Zhou, X., Huang, W., Zhang, Y., Zhang, P., Hao, S.: Tropical cyclone intensity estimation using multi-dimensional convolutional neural network from multi-channel satellite imagery. IEEE Geosci. Remote Sensing Lett. (2021)
30. Zha, K., Cao, P., Son, J., Yang, Y., Katabi, D.: Rank-n-contrast: Learning continuous representations for regression. In: Thirty-seventh Conference on Neural Information Processing Systems (2023)

31. Zhou, Z., Zhao, Y., Qing, Y., Jiang, W., Wu, Y., Chen, W.: A physics-guided nn-based approach for tropical cyclone intensity estimation. In: Proceedings of the 2023 SIAM International Conference on Data Mining (SDM), pp. 388–396. SIAM (2023)
32. Zhuo, J.Y., Tan, Z.M.: Physics-augmented deep learning to improve tropical cyclone intensity and size estimation from satellite imagery. Mon. Weather Rev. **149**(7), 2097–2113 (2021)