Single Domain Generalization via Normalised Cross-correlation Based Convolutions

WeiQin Chuah1* Ruwan Tennakoon1* Reza Hoseinnezhad1 David Suter2 Alireza Bab-Hadiashar1
RMIT University, Australia1 Edith Cowan University (ECU), Australia2
{wei.qin.chuah, ruwan.tennakoon, rezah, abh}@rmit.edu.au, d.suter@ecu.edu.au

Abstract

Deep learning techniques often perform poorly in the presence of domain shift, where the test data follows a different distribution than the training data. The most practically desirable approach to address this issue is Single Domain Generalization (S-DG), which aims to train robust models using data from a single source. Prior work on S-DG has primarily focused on using data augmentation techniques to generate diverse training data. In this paper, we explore an alternative approach by investigating the robustness of linear operators, such as convolution and dense layers commonly used in deep learning. We propose a novel operator called “XCNorm” that computes the normalized cross-correlation between weights and an input feature patch. This approach is invariant to both affine shifts and changes in energy within a local feature patch and eliminates the need for commonly used non-linear activation functions. We show that deep neural networks composed of this operator are robust to common semantic distribution shifts. Furthermore, our empirical results on single-domain generalization benchmarks demonstrate that our proposed technique performs comparably to the state-of-the-art methods.

1. Introduction

Deep learning techniques have achieved practical success in a variety of fields, including computer vision, natural language processing, and speech processing. However, this success is often limited to settings where the test data follows the same distribution as the training data. In many real-world situations, this assumption breaks down due to shifts in data distribution, known as domain-shift [2], which can significantly degrade performance [24].

Dealing with domain-shift is a challenging problem with important practical implications. There are two main approaches to address domain shift and enable the transfer of knowledge from previously seen environments (source domains) to a new environment (target domain) without using any labeled data of the target domain: (1) Domain Adaptation [31] (DA) where a model trained with source data is recalibrated using unlabeled data from the target domain, and (2) Domain generalisation [36] (DG) where a model is trained on multiple source domains but no target domain data is available for recalibration. The most data-efficient domain generalisation technique is the single domain generalisation (S-DG), which requires data from only a single source domain to train a model that is robust against unforeseen data shifts. Although practically desirable, S-DG has received little attention in the past.

S-DG presents a significant challenge due to two main factors. Firstly, the input data, derived from only one source domain, does not provide sufficient opportunity to observe the possible diversity in out-of-domain data. Secondly, the presence of spurious correlations or shortcuts can further complicate the issue by introducing biases and hindering generalization. Prior work on S-DG has primarily focused on increasing the diversity of input data using adaptive data augmentation techniques. In this work, we explore an alternative approach by investigating the robustness of linear operators, such as convolution and dense layers commonly used in deep learning. We propose a novel operator called “XCNorm” that computes the normalized cross-correlation between weights and an input feature patch. This approach is invariant to both affine shifts and changes in energy within a local feature patch and eliminates the need for commonly used non-linear activation functions. We show that deep neural networks composed of this operator are robust to common semantic distribution shifts. Furthermore, our empirical results on single-domain generalization benchmarks demonstrate that our proposed technique performs comparably to the state-of-the-art methods.

Figure 1. Comparison of accuracy for common normalization methods in image classification on CIFAR-10-C, considering five levels of domain discrepancy caused by corruptions. The evaluated methods include BN, IN, and our proposed approaches (XCNorm and R-XCNorm).
augmentation techniques. These include creating fictitious examples that mimic anticipated shifts in data distribution using random [34], adversarial [8, 21, 26, 35] or causality [5, 20] based data augmentation, as well as image style diversification [29].

The generalization of a model is largely influenced by its support, which refers to the diversity of its input data and its inductive biases [30]. While not explicitly stated as such, the success of the above-mentioned S-DG methods hinges on increasing the input diversity via data augmentation. An alternative and complementary approach that has received less attention is to incorporate inductive biases into the network components to make them more robust to domain shifts. In this paper, we explore the above approach and propose a robust alternative to linear operators, such as convolution and dense layers, which are fundamental components of most neural networks.

We draw on the classical idea of template matching and consider linear layers in a neural network as computing a matching between the template (represented by the weights) and the signal (represented by input feature maps) using cross-correlation, as detailed in Equation (4). Early works in template matching have shown that cross-correlation is not ideal for pattern matching as it fails when the local energy of the input (i.e., $\sum_{u,v} z(u,v)^2$) varies, and is not robust to affine shifts in the input signal [14]. More recently, Jin et al. [12] empirically showed that domain shift primarily causes variation in the local energy of feature representations. This suggests that the linear operator, which is sensitive to local energy, might degrade out-of-domain (OOD) generalization.

The above perspective enables us to use more robust template-matching techniques such as normalized cross-correlation [14] to replace convolutions or dense layers in neural networks and recover the underlying invariant features in the input. We call our method “XCNorm”, which performs cross-correlation between standardized (i.e., Z-score normalized) weights and patch-wise standardized input feature maps. This reduces the influence of input feature magnitude on the output and makes the operator invariant to affine transformations of the input. Moreover, we leverage robust statistics to improve the resilience of XCNorm to outliers and introduce a refined version of our method named R-XCNorm. As Figure 1 demonstrates, our methods achieve significantly better robustness to semantic distribution shifts on CIFAR-10-C, in contrast to other normalization techniques. Moreover, the advantage of our methods becomes more pronounced as the domain discrepancy increases. The contributions of this paper include:

- Leveraging robust statistics, we further enhance the robustness of XCNorm to outliers. Our experiments on several commonly used benchmarks in S-DG show that the proposed robust operator (“R-XCNorm”) is also complementary to augmentation-based methods and achieves state-of-the-art performance.

- We empirically show that a neural network based on “XCNorm” or “R-XCNorm”, is significantly more robust to semantic shifts compared to a network based on a typical linear operator.

2. Related Work

Domain Generalization: Domain generalization (DG) methods aim to learn robust models from several source domains that can generalize to unseen target domains, addressing the issue of domain shifts.

A particularly challenging yet practical variant of domain generalization (DG) is a single-domain generalization (S-DG), where only one source domain is available during training. S-DG is especially challenging because, unlike in DG, there is no access to multiple source domains that would allow for the observation of possible shifts in data and invariances between domains. To address this challenge, researchers have primarily focused on using data augmentation techniques to generate diverse training data and increase input diversity. A common technique is posing S-DG as a “distributionally robust optimization” problem and solving it using adversarial data augmentation (ADA) [26]. ADA lacks the ability to produce large semantic shifts that are common in real data. As a result, subsequent works have added additional constraints to adversarial augmentation [8, 15, 21, 35] or incorporated background knowledge about anticipated semantic shifts via random augmentations [34], causality based data augmentations [5, 20], or image style diversification [29].

An alternative that has received little attention is to incorporate inductive biases into the network components to make them more robust to domain shifts. The most closely related work in this direction is the Meta Convolution Neural Networks (Meta-CNN) proposed by Wan et al. [27], where the output feature-maps of each layer are reconstructed using templates learned from training data, resulting in universally coded images without biased information from unseen domains. Our proposed operator, XCNorm, offers a simpler implementation compared to [27]. While their method involves more complex operations, our approach simply replaces the convolution function with our XCNorm method. This simplicity makes our method more straightforward to implement and integrate into existing frameworks.

Normalization in Neural Networks: During the training process of a deep neural network, the input distribu-
3. Method

Single Domain Generalization (SDG) aims to learn a robust predictive function \( f_\theta : X \rightarrow Y \) that can achieve a minimum prediction error on an unseen target domain \( T \sim P_X^T \), where \( P_X^S \neq P_X^T \). However, \( X \) is constructed from a mix of domain-invariant factors \( X_0 \) and domain rendering factors \( R \), and only the former corresponds to the label \( Y \) as illustrated in Figure 2. Moreover, as \( R \) differs between source and target domains, causing domain shifts, a predictive function \( f_\theta \) that is invariant to \( R \) would greatly improve SDG.

To this extent, we draw on the classical idea of template matching and propose XCNorm that computes the normalized cross-correlation between weights and an input feature patch. Our method is invariant to both affine shifts and changes in energy within a local feature patch, recovering the underlying invariant features \( X_0 \) in the input. Additionally, XCNorm inherently integrates normalization and non-linear operations, negating the necessity for normalization layers and activation functions commonly employed in convolutional neural networks.

3.1. Normalized Cross-Correlation Layer

Typically, the linear units of a DNN layer compute the cross-correlation between the input feature maps \( z \in \mathbb{R}^{H \times W \times C_{in}} \) and the weights \( w \in \mathbb{R}^{K \times K \times C_{in} \times C_{out}} \). Here \( C_{in(out)} \) is the number of channels in the input (or output), \([H, W]\) are the feature map spatial dimensions and \( K \)

\[\text{Conv}(z; w) = \langle z, w \rangle = \|z\| \|w\| \cos(\varphi) = h(\|z\|, \|w\|) g(\cos(\varphi)) \]

where \( \cos(\varphi) \) is the angle between \( z \) and \( w \). This view allows for the separation of norm terms from the angle term (decoupling), and to change the form of \( h() \) and \( g() \) independently. Liu et al. [16] derived several decoupled variants of the functions \( h() \) and \( g() \). They demonstrated that the decoupled reparameterizations lead to significant performance gains, easier convergence, and stronger adversarial robustness. Later [3] introduced the “B-cos” operator and showed that it lead to better neural network interpretations.

Our proposed XCNorm operator can also be seen within this framework, where the dot product is taken between the centered and normalized versions of the input patch and the weights. While the general formula presented in Equation (2) encompasses a range of potential frameworks as demonstrated in [16], it’s important to note that our XCNorm and its robust variant have not been previously proposed or thoroughly examined. Moreover, unlike the methods mentioned above, we investigate the use of XCNorm for out-of-domain generalization in a single source domain setting.
are the mean and variance of the weights within each output channel is computed as:

\[ \Phi_{u,v,c}(z; w) = \langle z_{\bar{u},\bar{v}}, w_c \rangle \]

\[ = \sum_j \tilde{z}_j^{(j)} \cdot w_c^{(j)} \]  

(4)

Here \( z_{\bar{u},\bar{v}} \) is a patch of the input feature map centered at \( \bar{u}, \bar{v} \) with the same shape as the weight tensor \( w_c \), and \( j \) index the pixels within the patch. The map \( \nu : (u, v) \rightarrow (\bar{u}, \bar{v}) \) is determined by parameters of the convolution (or dense) layer (i.e., stride, kernel width).

The above operator is ineffective for pattern matching when the patch energy (i.e., \( \| z_{\bar{u},\bar{v}} \| = \sum_j (\tilde{z}_j^{(j)})^2 \)) is not uniform across a feature map. It also lacks robustness to affine transformations of the input, i.e., \( \langle z_{\bar{u},\bar{v}}, w_c \rangle \neq \langle Az_{\bar{u},\bar{v}} + b, w_c \rangle \) [14]. To overcome the above limitations, we propose using the normalized cross-correlation operator as a replacement for the linear operator. With this new operator, which we coin XNorm, the output feature at position \((u, v)\) of the \( c \)th feature channel is computed as:

\[ \Psi_{u,v,c}(z; w) = \sum_j \frac{\tilde{z}_j^{(j)} \cdot \mu_{z(\bar{u},\bar{v})}^j \cdot w_c^{(j)} \cdot \mu_{w_c}}{\| z_{\bar{u},\bar{v}} - \mu_{z(\bar{u},\bar{v})} \|} \]  

\[ = \frac{\sum_j \tilde{z}_j^{(j)} \cdot w_c^{(j)} \cdot \mu_{w_c}}{\alpha \sigma_{w,c} \sqrt{\frac{1}{\alpha} \sum_j (\tilde{z}_j^{(j)} \cdot z_{\bar{u},\bar{v}} - \mu_{z(\bar{u},\bar{v})}^2) + \epsilon}} \]  

(6)

Here, \( \mu_\cdot \) is the mean of \( \cdot \), \( \sigma_{w,c} = \sqrt{\frac{1}{\alpha} \sum_j (w_c^{(j)} - \mu_{w,c})^2} \), \( \epsilon \) is a small constant to ensure numerical stability, and \( \alpha = K \times K \times C_{in} \) is the number of pixels in the patch.

Since the patch-wise mean, \( \mu_\cdot \), can be computed using linear operations (i.e., convolution of a constant weight tensor with all elements being equal to \( 1/\alpha \)), the XNorm can be realized using linear operators:

\[ \Psi(z; w) = \frac{\phi(z; w) - \alpha \Phi(z; w\alpha) \mu_w}{\alpha \sqrt{\Phi(z; w\alpha) - \Phi(z; w\alpha)^2} \sigma_w + \epsilon} \]  

\[ = \phi(z; w) - \alpha \mu_z \mu_w \]  

\[ \frac{1}{\alpha} \sqrt{\mu_z^2 - \mu_w^2} \sigma_w + \epsilon \]  

(8)

Here, \( w_{\alpha} \in \mathbb{R}^{K \times K \times C_{in} \times 1} \) is a constant matrix with all elements equal to \( 1/\alpha \), and \( \mu_w \in \mathbb{R}^{1 \times C_{out}} \) and \( \sigma_w \in \mathbb{R}^{1 \times C_{out}} \) are the mean and variance of the weights within each output channel.

### 3.2. Robust XNorm

The sensitivity of XNorm to outliers in the input patch [7] can lead to issues when the input distribution changes unpredictably or includes outliers. For instance, salt and pepper noise can cause large variations in the input energy (first term in the denominator of Equation (5)) and affect the output of XNorm. To overcome this, we propose a robust version of the operator called R-XNorm, which modifies Equation (5) as follows:

\[ \Gamma_{u,v,c}(z; w) = \sum_j \left[ \phi(z_j^{(j)} - \mu_{z(\bar{u},\bar{v})}) \cdot w_c^{(j)} - \mu_{w_c} \right] \]  

\[ + \epsilon \| w_c - \mu_{w_c} \| \]  

(9)

Here, \( \phi(\cdot) \) can be any robust function such as the Huber, Cauchy (aka Lorentzian), Tukey, or Welsch function. In this work, we use the Welsch function due to its simplicity, which is defined as follows [6]:

\[ \phi(z) = c \left[ 1 - \exp \left( -\frac{z^2}{2c^2} \right) \right] \]  

(10)

Here, \( c \) is a learnable parameter that controls the amount of penalty for the outliers. The influence of \( c \) is depicted in Figure 3a. During the training process, we adopt an initialization strategy where we set the value of the parameter \( c \) to a large value. Subsequently, we update the value of \( c \) for each layer by computing the mean of the patchwise standard deviation of the input over the training dataset. Similar to batch normalization, we incorporate a moving average component to enhance the stability and effectiveness of the normalization process.

### 3.3. Improving Convergence

To enhance the convergence of XNorm and R-XNorm, we incorporate some modifications to their base formulations. We use the notation \( \Upsilon(z; w) \) to denote either \( \Psi(z; w) \) or \( \Gamma(z; w) \) in the following text.

**Sharpening:** The peaks and valleys of the output \( \Upsilon(z; w) \) can be emphasized (or de-emphasized) by raising it to power \( r \):

\[ \Upsilon_r(z; w) = \{ \max(0, \Upsilon(z; w)) \}^r \]  

(11)
Here, we solely consider the positive outputs, as our empirical observations indicate that this choice stabilizes the training process and prevents convergence to undesirable optima. Figure 3b shows the relationship between the input and output of the above operation. Similar, approaches have also been used in cosine similarity-based techniques [32]. However, our approach differs from cosine similarity-based methods in that we do not pre-determine the value of \( \tau \). Instead, we treat it as a learnable parameter and optimize it alongside the weights.

**Gradient Scaling:** The weight normalization in Equation (9) tends to reduce the gradient magnitude, which in turn leads to slower convergence [16]. To mitigate this issue, we propose gradient scaling using a learnable scaling factor \( A \in \mathbb{R}^{C_l} \), where \( C_l \) is the feature channel dimension of layer \( l \). More specifically, we apply different \( A \) to the output of \( \bar{\Upsilon}_{[1]}(z; w) \) at every layer, as shown below:

\[
\bar{\Upsilon}_{[2]}(z; w) = A \odot \bar{\Upsilon}_{[1]}(z; w).
\]

Given that scaling the output of a function by a constant is equivalent to augmenting the gradient of the function by that constant, the proposed method effectively addresses the reduced gradient issue and accelerates the convergence process.

**Norm-Based Attention Mask (NBAM):** The input norm \( \|\bar{z}\| \) signifies the importance of the local patch within the input. Here, \( \bar{z} = z_{o,d} - \mu_{z(a,d)} \) for \( XCNorm \) and \( \bar{z} = \phi(z_{o,d} - \mu_{z(a,d)}) \) for \( R-XCNorm \). Normalizing with \( \|\bar{z}\| \) assigns equal importance to all patches, but this may be problematic when \( \|\bar{z}\| \) is very small. Such small values indicate low-variation (low-information) input areas that should not be equally weighted as high-information areas, as this may cause spurious matches with templates.

To address this issue, we propose the *Norm-based Attention Mask* technique, which leverages a lightweight single convolution layer with sigmoid activation, denoted as \( \psi \). This function learns to dynamically assign importance weights to image patches. Specifically, by taking \( \|\bar{z}\| \) as input, \( \psi \) learns to generate patch-wise importance weights \( m \in [0, 1] \). Subsequently, the output is redefined using the following computation:

\[
\bar{\Upsilon}_{[3]}(z; w) = \psi(\|\bar{z}\|) \odot \bar{\Upsilon}_{[2]}(z; w) \\
+ \left( 1 - \psi(\|\bar{z}\|) \right) \odot \left\{ \bar{\Upsilon}_{[2]}(z; w) \odot \|\bar{z}\| \right\}
\]

where \( \Gamma(z; w) \odot \|\bar{z}\| \) represents the output without normalization.

To further ensure that each feature is weighted equally, we normalize each channel in the output tensor:

\[
\bar{\Upsilon}_{[4]}(z; w) = \frac{\bar{\Upsilon}_{[3]}(z; w) - \mu_{\bar{\Upsilon}}}{\sigma_{\bar{\Upsilon}}}
\]

where \( \mu_{\bar{\Upsilon}} \) and \( \sigma_{\bar{\Upsilon}} \) are the mean and the variance of \( \bar{\Upsilon}_{[3]}(z; w) \) along each channel.

### 4. Experiments

#### 4.1. Datasets and Settings

**Digits-DG:** consists of five distinct subsets: MNIST, MNIST-M, SVHN, SYN and USPS. Each subset represents a different domain with variations in writing styles and quality, scales, backgrounds, and strokes. We mainly utilise the Digits-DG benchmark for single-source domain evaluation and ablation studies. Following [21, 26, 27, 29], we chose the first 10,000 images from both the MNIST training and validation sets as the source dataset.

**CIFAR-10-C** [10]: is typically used for corruption robustness benchmarking. It contains 15 different corruption types that mimic real-world scenarios, such as noise, blur, weather, and digital artifacts. Each corruption type has five levels of severity. We follow the setup of [27, 29] and use the CIFAR-10 training set as the source dataset while images in CIFAR-10-C are used for evaluation.

**Camelyon-17-Wilds** [13]: comprises 455k histopathology image patches extracted from 1000 whole-slide images (WSIs) of sentinel lymph nodes [4]. These WSIs were obtained from five distinct medical centres, with each centre representing a unique domain within the dataset. The primary objective of this dataset is to classify input patches and determine whether the central region contains any tumour tissue. In our experimental setup, we selected each of the domains as the source domain in turn and used the rest of the domains as target domains.

**Implementation Details:** As mentioned in Section 3, XC-
Table 2. Experiments of single domain generalization on CIFAR-10 classification. Models are trained on the CIFAR-10 train set and evaluated on the CIFAR-10-C benchmark. "*" represents our implementation of existing work. Our method is also complementary to existing data augmentation methods such as Random Convolution (RC). The best results are denoted as bold and underlined, while the second best results are indicated by being underlined.

<table>
<thead>
<tr>
<th>Method</th>
<th>Weather</th>
<th>Blur</th>
<th>Noise</th>
<th>Digits</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERM</td>
<td>67.28</td>
<td>56.73</td>
<td>30.02</td>
<td>62.30</td>
<td>54.08</td>
</tr>
<tr>
<td>ADA</td>
<td>72.67</td>
<td>67.04</td>
<td>39.97</td>
<td>66.62</td>
<td>61.58</td>
</tr>
<tr>
<td>M-ADA</td>
<td>75.54</td>
<td>63.76</td>
<td>54.21</td>
<td>65.10</td>
<td>64.65</td>
</tr>
<tr>
<td>ME-ADA</td>
<td>74.44</td>
<td>71.37</td>
<td>66.47</td>
<td>70.83</td>
<td>70.78</td>
</tr>
<tr>
<td>RandConv*</td>
<td>74.90</td>
<td>74.95</td>
<td>55.71</td>
<td>76.80</td>
<td>70.59</td>
</tr>
<tr>
<td>L2D</td>
<td>75.98</td>
<td>69.16</td>
<td>73.29</td>
<td>72.02</td>
<td>72.61</td>
</tr>
<tr>
<td>MetaCNN</td>
<td>77.44</td>
<td>76.80</td>
<td>78.23</td>
<td>81.26</td>
<td>78.43</td>
</tr>
<tr>
<td>XCNorm</td>
<td>72.98</td>
<td>73.76</td>
<td>48.15</td>
<td>77.52</td>
<td>68.10</td>
</tr>
<tr>
<td>R-XCNorm</td>
<td>75.57</td>
<td>75.72</td>
<td>60.81</td>
<td>77.61</td>
<td>72.43</td>
</tr>
<tr>
<td>R-XCNorm^{+RC}</td>
<td>76.99</td>
<td>79.61</td>
<td>63.95</td>
<td>82.47</td>
<td>75.70</td>
</tr>
</tbody>
</table>

Norm incorporates normalization and non-linear operations by default. Thus, in our implementation, we replace all Conv-Norm-ReLU blocks as well as the final classifier layer within our chosen convolutional neural network architecture with our proposed XCNorm or R-XCNorm.

Complete details regarding the experimental setup, including the network architecture and model selection, can be found in the supplementary document, providing a comprehensive understanding of our methodology.

4.2. Comparisons on Digits-DG

Results: Table 1 provides a comprehensive comparison of the out-of-domain generalization performance between our proposed method and state-of-the-art approaches. The results highlight the significant improvements achieved by our proposed method. Both our base method (XCNorm) and robust variant (R-XCNorm) showcase substantial enhancements over the ERM baseline (49.3% → 69.8% and 49.3% → 74.2%, respectively), without the need for data augmentation techniques or extensive network modifications. Notably, our methods outperform adversarial augmentation-based domain generalization approaches, including ADA, M-ADA, and ME-ADA, by a considerable margin. Also, Table 1 highlights the complementary nature of our method with data augmentation techniques, such as Random Convolution (RC) [34]. The combined approach achieves even higher performance, positioning it competitively alongside the complex state-of-the-art method, MetaCNN.

4.3. Comparisons on CIFAR-10-C

Results: The average accuracy for four categories of level-5 severity corruptions is presented in Table 2. Our proposed method showcases substantial improvements over the baseline model (ERM), achieving an impressive accuracy boost from 54.08% to 68.10% without relying on data augmentation techniques. Notably, our approach demonstrates a remarkable 17.03% improvement for blur corruptions and a noteworthy 15.22% improvement for digits corruptions. Furthermore, our R-XCNorm method effectively enhances model robustness against noise corruption, elevating the accuracy from 48.15% to 60.81%. It is worth highlighting that our R-XCNorm outperforms most of the data augmentation-based approaches by a respectable margin. Additionally, when combined with RC [34], our method exhibits exceptional performance across all four categories and demonstrates competitive results compared to the current state-of-the-art method.

4.4. Comparisons on Camelyon-17

Results: While the Camelyon17 dataset is commonly employed for conventional domain generalization tasks (generalizing from multiple source domains to a single target domain), it has not been extensively explored for single domain generalization. In this study, we investigate the single domain generalization performance on the Camelyon17 dataset using the AUROC metric, as shown in Figure 4. Notably, the ERM model exhibits poor generalization when trained on a single source domain, particularly in domains 2 and 3. We attribute this observation to significant variations in the staining agent’s color across different hospitals (refer to the supplementary document for image samples collected from different domains). In contrast, the Random Convolution (RC) approach demonstrates impressive domain generalization capabilities on the Camelyon17 dataset.

Remarkably, our proposed XCNorm method and its robust variant, R-XCNorm, consistently outperform the ERM baseline across all domains, without relying on data augmentation techniques. In fact, our methods even surpass the performance of the RC method. Moreover, when combined with RC, our approach achieves a robust model with exceptional domain generalization performance. These re-
Figure 5. Model Robustness Score (MRS) comparison of different models on the CIFAR-10-C dataset. The MRS values were computed for various categories, including weather (W), blur (B), digital compression (D), noise (N), and chromatic (C) augmentations. Lower MRS values indicate better model robustness against corruption. The baseline ERM model exhibits poor performance, while our proposed method demonstrates significantly better robustness compared to ERM.

Table 3. Ablation results on Digits-DG benchmark, and the Top-1 accuracy is reported. Our method can substantially improve the generalization performance of the baseline model.

results highlight the effectiveness of our method in enhancing domain generalization and its potential to improve model robustness in practical applications, such as medical imaging classification.

5. Discussion

5.1. Ablation Study

In this section, we present the results of our ablation study conducted on the Digits-DG benchmark to evaluate the efficacy of each component of our proposed method. Specifically, we examine the effectiveness of our XC-Norm method, the robust variant (R-XCNorm), and the norm-based attention mask (NBAM) proposed to improve model convergence, as discussed in Section 3.3.

Table 3 reports the classification results of the four variants of our original framework, including the baseline ERM model for comparison. Our XCNorm method, without any extension, achieves a significant 19.7% performance improvement over the baseline, demonstrating the effectiveness of our approach. Also, the integration of NBAM, which relaxes the normalization of input features, leads to additional performance gains. This highlights the importance of selectively normalizing important input regions, as a global normalization approach may mistakenly assign equal importance to insignificant regions and hinder training.

Furthermore, we evaluate the effectiveness of the robust variant, R-XCNorm, specifically designed for outlier rejection. As illustrated in Table 3, the integration of the Welsch robust function enhances performance across all domains except for the USPS dataset. Moreover, combining the robust variant with NBAM yields even greater improvements across all domains without sacrificing performance. These results highlight the benefits of incorporating the robust variant and NBAM in our framework, demonstrating their potential for enhancing model performance and domain generalization.

5.2. Gradient Scaling

In this section, we investigate the effect of gradient scaling, as proposed in Section 3.3, on the performance of XC-Norm. As mentioned, The weight normalization employed in Equation (9) tends to reduce the gradient magnitude during backpropagation, as shown in Equation (15):

$$\frac{\partial}{\partial \hat{w}} \Gamma(\hat{z}; \hat{w}) = \frac{\hat{z} - \hat{w}^{\top} \cdot \hat{z} \cdot \hat{w}}{\|\hat{w}\|^4}$$

where $\hat{z} = \hat{z}/\|\hat{z}\|$ and $\hat{w} = \hat{w}/\|\hat{w}\|$. From Equation (15), we observe that the large norm $|\hat{w}|$ can diminish the gradients, significantly slowing the convergence of the optimization process.

To address this issue, we propose a gradient scaling method. As shown in Figure 6, we compare the conver-
The causal factor $X$ severity level of 5 indicates the most severe corruption.  

A severity level of 0 represents no corruption, while a negative of S-DG is to develop a model that is invariant to the problem of diminishing gradients caused by weight normalizations, ultimately speeding up the training convergence of our proposed gradient scaling method in overcoming the problem of diminishing gradients caused by weight normalization, ultimately speeding up the training convergence.

5.3. Coherence of XCNorm and Random Convolution

Integrating Random Convolution (RC) [34] with XCNorm further boosts single domain generalization performance, especially with blur, noise, and digital corruptions (see Table 2). We postulate that this is due to the diversification of texture information introduced by RandConv which prevents learning highly “texture-sensitive templates”. RC generates new images with comparable global shapes to the source data while introducing random local textures strengthening the model’s resistance against local texture variations inherent in the mentioned corruptions.

5.4. Model Sensitivity to Input Perturbations

In this section, we investigate the sensitivity of our model to input perturbations using the CIFAR-10-C dataset. Our analysis focuses on the variance in predicted class probabilities, $P(y|x)$, when using a set of corrupted images with different severity levels (denoted as $s$) ranging from 0 to 5. A severity level of 0 represents no corruption, while a severity level of 5 indicates the most severe corruption.

As depicted in Figure 2, the input $X$ is influenced by the causal factor $X_0$ (e.g., semantics) and the rendering factor $R$, which contributes to domain shift. Since the objective of S-DG is to develop a model that is invariant to $R$ as much as possible, it is crucial to examine the effect of $R$ on the predictions of the developed model. We measure this by computing the Model Robustness Score (MRS), which quantifies the discrepancy between the predictions obtained using clean and perturbed inputs:

$$
MRS(x, \xi) = \frac{1}{5} \sum_{s=1}^{5} KL(f_\theta(\xi(x; 0)) ; f_\theta(\xi(x; s)))
$$

where, $\xi(x; s)$ represents the perturbation (e.g., blur, noise, compression, weather) included in the CIFAR-10-Corruption dataset. A model that relies solely on $X_0$ for prediction would exhibit a lower MRS compared to a model that incorporates $R$ in its predictions.

Figure 5 shows that the ERM model is highly susceptible to $R$, leading to large MRS values across all categories. In contrast, the proposed XCNorm effectively improves the model’s robustness, resulting in models that are less reliant on $R$ for prediction. Specifically, our method significantly reduces MRS for the weather, blur, and compression categories. Moreover, our robust variant R-XCNorm further diminishes the impact of $R$ on the model’s predictions, particularly in the noise category, thereby enhancing both the model’s robustness and domain generalization performance. These findings highlight the effectiveness of our approach in mitigating the influence of domain-specific factors and improving the model’s sensitivity to the underlying semantics $X_0$ rather than the rendering factor $R$.

6. Conclusion

In this paper, we introduce XCNorm, a novel normalization technique based on normalized cross-correlation. XCNorm exhibits invariance to affine shifts and changes in energy within local feature patches, while also eliminating the need for non-linear activation functions. The robust variant, R-XCNorm, focuses on outlier rejection, resulting in improved performance in challenging domains while maintaining competitiveness in others. Our proposed masking method selectively normalizes important input regions, enhancing model stability and out-of-domain performance. The integration of both methods showcases their complementary nature, leading to further improvements across all domains. We demonstrate the practical applicability of XCNorm in medical imaging classification, where it enhances model robustness and sensitivity to underlying semantics. Overall, our work provides effective methods for enhancing model performance and robustness, making notable contributions to the field of single-source domain generalization. These contributions have potential implications for various real-world applications.

Acknowledgements

This research was partially funded by the Australian Government and Ford Motor Company through an Australian Research Council Linkage Project (LP190100165).
References


[24] Rohan Taori, Achal Dave, Vaishaal Shankar, Nicholas Carlini, Benjamin Recht, and Ludwig Schmidt. Measuring ro-


