

## 3D常见的损坏和数据增量

Fatih Kar

Teresa Yeo

Andrei Atanov

Amir Zamir

瑞士联邦理工学院 (EPFL)

<https://3dcommoncorruptions.epfl.ch/>

### 摘要

我们介绍了一组图像变换，它们可以被用作评估模型的鲁棒性以及训练神经网络的数据增强机制的破坏。所提议的变换的主要区别在于，与现有的方法如Common Corruptions[27]不同，场景的几何形状被纳入变换中--从而导致更有可能在真实世界发生的损坏。我们还引入了一组语义上的破坏（例如自然物体遮挡。见图1）。

我们表明这些转换是"高效的"（可以即时计算），"可扩展的"（可以应用于大多数图像数据集），暴露了现有模型的弱点，并且在作为"三维数据增强"机制使用时可以有效地使模型更加健壮。对一些任务和数据集的评估表明，将三维信息纳入基准测试和训练为鲁棒性研究开辟了一个很有前途的方向。

形状一致的偏移。这导致了更有可能在现实世界中发生的转变（见图1）。由此产生的集合包括20个腐败，每个腐败代表一个

### 1. 简介

部署在现实世界中的计算机视觉模型会遇到来自其训练数据的自然发生的分布变化。这些偏移的范围从较低层次的失真，如运动模糊和光照变化，到语义上的失真，如物体遮挡。它们中的每一个都代表了一个模型的可能失败模式，并且经常被证明会导致非常不可靠的预测[15,23,27,31,66]。因此，在现实世界中部署这些模型之前，对这些转变的脆弱性进行系统测试是至关重要的。

这项工作提出了一套分布转移，以测试模型的稳健性。与之前提出的在图像上进行统一的二维修改的偏移相比，如Common Corruptions (2DCC) [27]，我们的偏移结合了三维信息来产生与场景几何

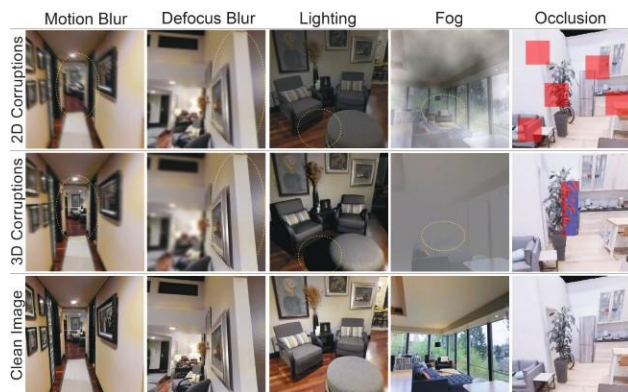


图1.使用三维信息来生成真实世界的损坏。最上面一行显示的是在图像上统一应用的二维破坏样本，例如在Common Corruptions[27]中，无视三维信息。这导致了在现实世界中不可能发生的破坏，例如在整个图像上有相同的运动模糊，而不考虑与相机的距离（左上）。中间一行显示的是来自3D Common Corruptions (3DCC) 的3D对应物。圈出的区域强调了纳入3D信息的效果。更具体地说，在3DCC中，**1.运动模糊**具有运动视差效应，离摄像机较远的物体似乎移动较少；**2.失焦模糊**具有景深效应，类似于真实摄像机中的大光圈效应，图像的某些区域可以被选择为焦点；**3.照明**在照亮场景时考虑到场景的几何形状，在物体上投下阴影。**4.**离摄像机越远，**雾气**越浓，**5.**通过改变摄像机的视角，让它的视线自然地被另一个物体，如植物（红色遮罩）所**遮挡**，从而产生一个焦油物体的**遮挡**，如冰箱（蓝色遮罩）。这与随机丢弃斑块的二维对应物形成对比[13]。图的视频版本见项目[页面](#)。

训练数据的分布变化，我们将其称为**3D 共同干扰**（3DCC）。3DCC解决了现实世界的几个问题，如相机运动、天气、遮挡、景深和照明。图2提供了一个所有损坏的概述。如图1所示，与单纯的2D方法相比，3DCC中的腐败现象更加多样化和现实化。

我们在第五节中表明，旨在提高鲁棒性的方法的性能，包括那些具有二维数据增强的方法，在3DCC下会急剧下降。此外，我们观察到，3DCC暴露的鲁棒性问题与通过现实合成产生的破坏有很大的关联。因此，3DCC可以作为一个挑战。

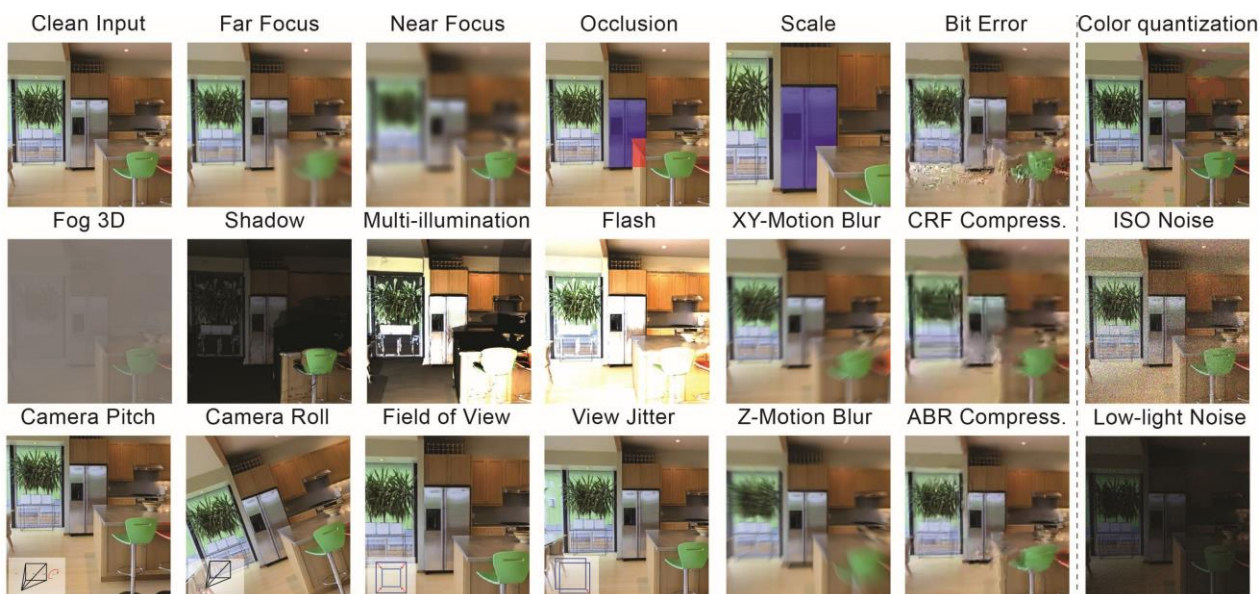


图2.新的破坏。我们提出了一系列不同的新的破坏操作，从散焦（近/远焦）到光照变化和三维语义的破坏，例如物体遮挡。这些破坏都是自动生成的，计算效率高，并且可以应用于大多数数据集（第3.3节）。我们表明它们暴露了模型中的漏洞（第5.2.1节），并且很好地接近了现实中的损坏（第5.2.3节）。

最后一列中标明的腐败的子集是新颖的，在现实世界中经常面临，但不是基于3D的。我们把它包括在我们的基准中。对于闭塞和比例腐败，蓝色和红色掩码表示物体的正交可见部分和闭塞部分，例如冰箱。

在现实世界中，特别是那些依赖于场景几何的损坏，是一个测试平台。

受此启发，我们的框架也引入了新的三维数据增强。与二维增强相比，它们考虑到了场景的几何形状，从而使模型能够针对更真实的破坏建立不变性。我们在第5.3节中表明，它们极大地提高了模型对这种破坏的鲁棒性，包括那些不能被二维增强所解决的破坏。

所提出的破坏是通过编程产生的，并带有暴露的参数，能够对鲁棒性进行细粒度分析，例如通过不断增加三维运动模糊度。它们的计算效率很高，并且可以在训练过程中作为数据扩展而即时生成，计算成本增加很少。它们也是可以扩展的，也就是说，它们可以应用于标准的视觉数据集。

例如，ImageNet[12]，它不带有三维标签。

## 2. 相关工作

这项工作提出了一个以数据为中心的稳健性方法[51,62]。在篇幅有限的情况下，我们对一些相关的主题进行了概述。

**基于腐败的稳健性基准。**有七项研究提出了鲁棒性基准，以了解模型对腐败的脆弱性。一个流行的基准，Common Corruptions (2DCC)[27]，在真实的图像上生成合成的损坏，显示出图像识别模型的敏感性。它导致了一系列的工作，要么创造新的损坏，要么在其他数据集上应用类似的损坏，用于不同的

任务[7,32,42,44,65,78]。与这些工作不同的是，3DCC利用三维信息修改真实的图像，以产生再现实的破坏。由此产生的图像与二维图像相比，在感知上是不同的，在模型预测上也暴露出不同的故障模式（见图1和8）。其他的工作在现实世界中创造和捕捉腐败，例如ObjectNet[3]。虽然是真实的，但它需要大量的手工劳动，而且不具有扩展能力。一个更可扩展的方法是使用基于计算机图形的三维模拟器来生成损坏的数据[37]，这可能会导致泛化的问题。3DCC旨在生成尽可能接近真实世界的破坏，同时保持可扩展性。

**稳健性分析**工作使用现有的基准来探测不同方法的稳健性，例如数据增加或自我监督训练，在几个分布变化下。最近的工作调查了合成和自然分布变化之间的关系[14,26,43,67]和架构进步的有效性[5,47,63]。我们选择了几种流行的方法来说明3DCC可以作为一个具有挑战性的基准（图6和7）。

**提高鲁棒性。**已经提出了许多方法来提高模型的鲁棒性，如用损坏的数据来增加数据[22,39,40,59]，纹理变化[24,26]，图像合成[80,82]和转换[29,79]。虽然这些方法可以泛化到一些未见过的例子，但性能的提高是不均匀的[22,60]。其他的方法包括自我训练[74]、预训练[28,49]、结构变化[5,63]和二维集合[33,50,76,77]。在这里，我们采用了一种以数据为中心的稳健性方法，即i。



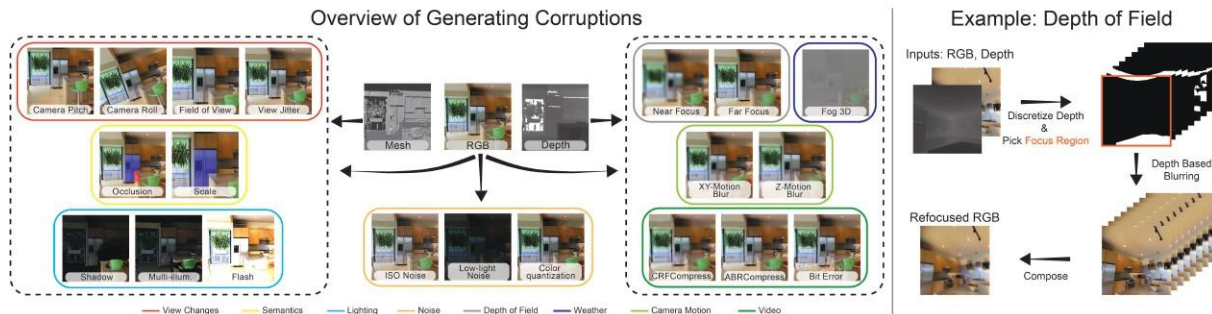


图3.左图：我们展示了创建每个腐败所需的输入，例如三维信息，如深度，和RGB图像。这些变形也根据它们的变形类型进行了分组（用实色线）。例如，要创建右边虚线框中的变形，只需要RGB图像和其相应的深度。对于左边虚线框中的那些，则需要3D网格。请注意，如果有全景图像，也可以在没有网格的情况下创建视图变化的破坏。右边：作为一个例子，我们展示了一个有效生成景深效果的概述。场景首先通过离散化的场景深度被分割成多个层次。接下来，选择一个区域来保持聚焦（这里是最靠近摄像机的区域）。然后，我们使用针孔相机模型，根据每层与焦点区域的距离计算出相应的模糊程度。最终重新聚焦的图像是通过合成模糊的图像层而得到的。

引入新的三维数据增强，以提高对现实世界的破坏的鲁棒性（第5.3节）。

**逼真的图像合成**涉及生成逼真图像的技术。这些技术中的一些最近被用于创建腐败数据。这些技术一般都是针对单一的真实世界的腐败现象的。例如，恶劣的天气条件[19,30,61,68,69]，运动模糊[6,48]，景深[4,17,52]，70,71]，照明[25,75]，和噪音[21,73]。它们可用于纯粹的艺术目的或创造训练数据。我们的一些三维变换是这些方法的实例化，其下游目标是在一个统一的框架内测试和改进模型的鲁棒性，并具有广泛的腐败。

**图像修复**的目的是使用经典的信号处理技术[18,20,34,41]或基于学习的方法[1,8,45,46,56,83,84]撤销图像中的损坏。我们与这些作品不同的是，我们生成了损坏的数据，而不是去除它，以使用它们作为基准或数据的增加。因此，在后者中，我们用这些损坏的数据进行训练，以鼓励模型不受损坏的影响，而不是作为一个预处理步骤来训练模型以去除损坏。

**对抗性破坏**在输入中增加了难以察觉的最坏情况下的偏移来欺骗模型[11,35,40,66]。在现实世界中，大多数模型的失败案例都不是对抗性破坏的结果，而是自然发生的分布转移。因此，我们在本文中的重点是产生可能在现实世界中发生的腐败。

### 3. 生成3D常见的腐败现象

#### 3.1. 腐败类型

我们定义了不同的腐败类型，即景深、相机运动、灯光、视频、天气、视图变化、语义和噪音，导致20种腐败现象

在3DCC中。大多数损坏需要一个RGB图像和场景深度，而有些则需要3D网格（见图3）。我们使用一套利用三维合成技术或图像形成模型的方法来生成不同的腐败类型，下面会详细解释。进一步的细节在附录中提供。

**景深的破坏**会产生重新聚焦的图像。它们使图像的一部分保持在焦点上，而其余部分则变得模糊不清。我们考虑采用分层方法[4,17]，将场景分成多个层次。对于每一层，使用针孔相机模型计算出相应的模糊程度。然后用阿尔法混合法对模糊层进行合成。图3（右）显示了这个过程的概况。我们通过随机改变焦点区域到场景的近处或远处来产生近焦和远焦的破坏。

由于相机在曝光过程中的移动，相机运动会产生模糊的图像。为了产生这种效果，我们首先使用深度信息将输入图像转换为点云。然后，我们定义一个轨迹（相机移动）并沿着这个轨迹渲染新的视图。由于点云是由单一的RGB图像生成的，当相机移动时，它对场景的信息是不完整的。因此，渲染的视图会有不完整的假象。为了缓解这个问题，我们采用了[48]中的绘画方法。然后，生成的视图被组合起来以获得视差一致的运动模糊。当摄像机的主要运动是沿着图像的XY面或Z轴时，我们分别定义XY运动模糊和Z运动模糊。

**照明破坏**通过添加新的光源和修改原始照明来改变场景的照度。我们使用Blender[10]来放置这些新的光源，并为三维网格中的特定视点计算相应的照度。对于闪光破坏，光源被放置在摄像机的位置，而对于阴影破坏，它被放置在摄像机外壳以外的随机不同位置。同样，对于多光照度的破坏，我们从一组随机的光照度中计算出光照度。

不同位置和亮度的来源。

在视频的处理和流媒体过程中会出现视频损坏。利用场景三维，我们通过定义一个轨迹，使用单个图像的多个帧来创建一个视频，类似于运动模糊。受[78]的启发，我们生成平均比特率（ABR）和恒定速率因子（CRF）为

H.265编解码器的压缩伪影和比特误差，以捕获由不完善的视频传输通道引起的损坏。在对视频进行破坏后，我们挑选一帧作为最终破坏的图像。

由于介质中的干扰，天气破坏通过遮蔽场景的一部分来降低可见度。我们定义了一个单一的腐败，并将其表示为雾3D，以区别于2DCC中的雾腐败。我们使用雾的标准光学模型[19,61,69]。

$$\mathbf{I}(\mathbf{x}) = \mathbf{R}(\mathbf{x})\mathbf{t}(\mathbf{x}) + \mathbf{A}(1 - \mathbf{t}(\mathbf{x})). \quad (1)$$

其中 $\mathbf{I}(\mathbf{x})$ 是在像素 $x$ 处产生的雾状图像， $\mathbf{R}(\mathbf{x})$ 是干净的图像， $\mathbf{A}$ 是大气中的光线， $\mathbf{t}(\mathbf{x})$ 是描述光量的传输函数。

到达摄像机。当介质是均匀的，传输取决于与摄像机的距离， $\mathbf{t}(\mathbf{x}) = \exp(-\beta d(\mathbf{x}))$

其中 $d(\mathbf{x})$ 是场景深度， $\beta$ 是控制雾气厚度的衰减系数。视图的变化是由于摄像机外部的变化而引起的。

我们的框架能够使用Blender来渲染RGB图像，这些变化包括视场、相机的滚动和相机的俯仰。我们的框架能够使用Blender渲染以几种变化为条件的RGB图像，如视场、相机滚动和相机俯仰。这使我们能够以一种可控的方式分析模型对各种视图变化的敏感性。我们还生成了带有视图抖动的图像，可用于分析模型预测是否因视点的轻微变化而出现闪烁。

**语义学。**除了视图的变化，我们还通过选择场景中的一个物体并改变其遮挡程度和比例来渲染图像。在闭塞破坏中，我们生成一个被其他物体遮挡的物体的视图。这与像素的随机二维遮蔽形成不自然的闭塞效果是不同的，例如在[13,47]中（见图1）。遮挡率可以被控制，以探测模型对遮挡变化的鲁棒性。同样地，在比例腐败中，我们渲染一个物体与相机位置的不同距离的视图。请注意，这些破坏需要一个带有语义符号的网格，并且是自动生成的，类似于[2]。这与[3]相反，后者需要繁琐的手工操作。对象可以通过在场景中随机选取一个点或者使用语义注释来选择。

**噪声**破坏来自不完善的相机传感器。我们引入了以前的2DCC基准中不存在的新的噪声破坏。对于低光噪声，我们降低了像素强度并增加了泊松-高斯分布的噪声，以反映低光成像环境[21]。ISO噪声也遵循泊松-高斯分布，有固定的光子噪声（以泊松为模型）和变化的电子噪声（以高斯为模型）。我们还包括

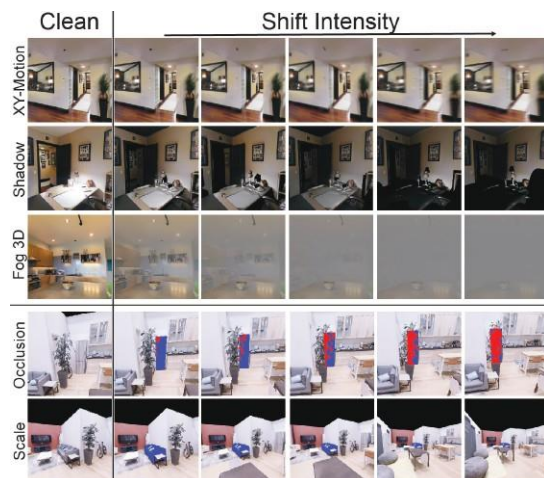


图4.随着移位强度的增加，3DCC的可视化。上图：增加移位强度会导致更大的模糊，更少的照明和更密集的雾。底部。使用计算出的视角变化，物体变得更加隐蔽或尺寸缩小。蓝色的遮罩表示冰箱/沙发的正负可见部分，红色的遮罩是被遮挡的部分。最左边一栏显示的是干净的图像。所有移位强度下的所有损坏的视觉效果都显示在补充文件中。

颜色量化是另一种减少RGB图像比特深度的破坏。只有这个子集的破坏不是基于三维信息。

### 3.2. 起步3D常见腐败数据集

我们发布了我们管道的全部开放源代码，这使得我们可以在任何数据集上使用所实现的腐败。作为初始数据集，我们在16k

Taskonomy[81]测试图像上应用了腐败。对于所有的损坏，除了视图变化和改变场景的语义外，我们遵循2DCC的协议，定义了5个移动强度，产生了大约100万张损坏的图像（16k×14×5）。直接应用这些方法来产生破坏，会导致与2DCC相比的未校准的移位强度。因此，为了能够与2DCC在更均匀的强度变化上进行统一比较，我们进行了一个校准步骤。对于在2DCC中具有直接对应关系的损坏，例如运动模糊，我们在3DCC中设置了损坏级别，使得对于2DCC中的每个位移强度，所有图像的平均SSIM[72]值在两个基准中是相同的。对于在2DCC中没有对应的损坏，我们调整变形参数以增加移位强度，同时保持在与其他类似的SSIM范围。对于视图变化和语义，我们使用Replica[64]数据集渲染了32k个参数平滑变化的图像，例如滚动角度。图4显示了具有不同移位强度的腐败实例。

### 3.3. 将3DCC应用于标准视觉数据集

虽然我们采用了具有完整场景几何信息的数据集，如Taskonomy[81]，但3DCC也可以应用于没有三维信息的标准数据集。我们在ImageNet[12]和COCO[38]的验证中对此进行了示范。



通过利用MiDaS[54]模型的深度预测，一个最先进的深度估计器。图5显示了具有近焦、远焦和雾状三维破坏的图像实例。生成的图像在物理上是合理的，这表明3DCC可以被社区用于其他数据集，以生成一组不同的图像损坏。在第5.2.4节中，我们定量地证明了使用预测深度来生成3DCC的有效性。

## 4. 三维数据扩增

虽然基准测试使用损坏的图像作为测试数据，但我们也可以使用它们作为训练数据的增强部分，以建立对这些损坏的不变性。对我们来说就是这样，因为与2DCC不同，3DCC的设计是为了捕捉更有可能出现在真实世界的损坏，因此它也有一个合理的增强值。因此，除了使用3DCC进行鲁棒性基准测试外，我们的框架也可以被看作是新的数据增强策略，将3D场景的几何形状纳入其中。在我们的实验中，我们用以下腐败类型进行增强：景深、相机运动和光线。在使用并行实现的训练过程中，这些增强可以有效地随即生成。比如说

充分，景深增强在单个V100

GPU上需要0.87秒（挂钟时间），批量大小为128张224×224分辨率的图像。相比之下，应用二维散焦模糊平均需要0.54秒。也可以预先计算某些选定的

扩增过程的一部分，例如照明扩增的照度，以提高效率。我们将这些机制纳入我们的实现中。我们在第5.3节中表明，这些增强可以显著提高对现实世界失真的鲁棒性。

## 5. 实验

我们进行了评估，证明3DCC可以暴露出2DCC无法捕捉到的模型（第5.2.1节）中的漏洞（第5.2.2节）。

破坏与昂贵的现实合成破坏相似（第5.2.3节），并适用于没有三维信息的数据集（第5.2.4节）。最后，所提出的三维数据增量从质量上和数量上提高了鲁棒性（第5.3节）。请参阅项目页面，以获得更多的定性结果。

### 5.1. 预备工作

**评价任务。**3DCC可以应用于任何数据集，而不考虑目标任务，例如密集回归或低维分类。在这里，我们主要以表面法线和深度估计作为社区广泛采用的目标任务进行体验。我们注意到，与分类任务相比，解决这类任务的模型的鲁棒性还没有得到充分的探索（见补充充分分类的再研究结果）。为了评估稳健性，我们计算了

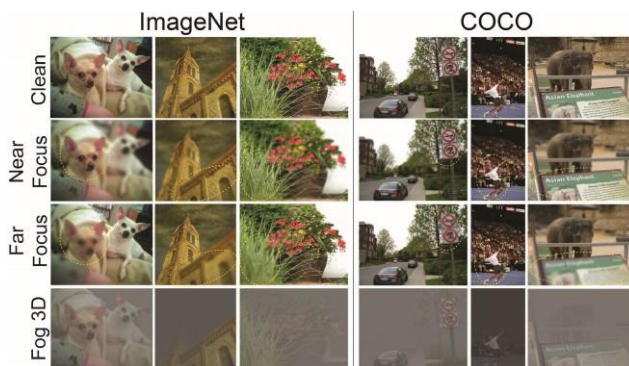


图5.3DCC可以应用于大多数数据集，甚至是那些不带有3D信息的数据集。来自ImageNet[12]和COCO[38]数据集的几张查询图像被应用了近焦、远焦和雾化的三维破坏。请注意，根据焦点区域和场景几何的不同，圈内的物体是如何从清晰到模糊的。为了获得创建这些破坏所需的深度信息，使用了MiDaS[54]模型的预测。这提供了一个足够好的近似值来产生真实的损坏（我们将在第5.2.4节进行量化）。

1 预测图像和地面真实图像之间的误差。**训练细节。**我们在Taskonomy[81]上训练UNet[58]和DPT[53]模型，使用学习率 $5 \times 10^{-4}$ ，权重衰减 $2 \times 10^{-6}$ 。

。我们使用AMSGrad[55]优化具有拉普拉斯先验的似然损失，遵循[77]。

除非特别说明，所有的模型都使用相同的UNet backbone（如图6）。我们还试验了在Omnidata[17]上训练的DPT模型，该模型混合了各种训练数据集。按照[17]，我们用学习率 $1 \times 10^{-5}$ ，权重衰减 $2 \times 10^{-6}$ ，角度和英镑

损失进行训练。**评估了稳健性机制。**我们评估了几种流行的数据增强策略。DeepAugment [26], style augmentation [24], and adversarial training [35].我们还包括跨域合集（X-

DE）[77]，该策略最近被证明可以通过输入转换创建不同的合集组件来提高对腐败的鲁棒性。训练细节请参考补充文件。最后，我们训练一个带有2DCC[27]腐败的增强模型（2DCC增强），以及另一个带有3D数据增强的模型（2DCC + 3D增强）。

### 5.2. 3D常见损坏基准

#### 5.2.1 3DCC会暴露出漏洞

我们对现有的模型与3DCC进行了基准测试，以了解其脆弱性。然而，我们注意到，我们的主要贡献不是所进行的分析，而是基准本身。最先进的模型可能会随着时间的推移而改变，而3DCC的目的是确定稳健性的趋势，与其他基准类似。

**稳健性机制的影响。**图6显示了不同鲁棒性机制在3DC的表面法线和深度估计任务中的平均表现。这些机制的性能比基线有所提高，但与清洁数据的性能相比仍有很大差距。

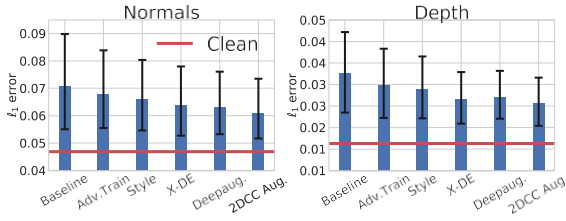


图6.现有的鲁棒性机制被认为不足以解决由3DCC近似的真实世界的损坏。图中显示了不同鲁棒性机制的模型在3DCC下对表面法线（左）和深度（右）估计任务的表现。这里所有的模型都是UNets，用Taskonomy数据进行训练。每个条形图显示的是在所有3DCC损坏中平均的 $\ell_1$ 误差（越低越好）。黑色的误差条显示了最低和最高移位强度下的误差。红线表示基线模型在干净（未损坏）数据上的表现。这表明现有的鲁棒性机制，包括那些具有不同增强功能的机制，在3DCC下表现不佳。

这表明3DCC暴露了鲁棒性问题，可以作为模型的一个挑战性测试平台。2DCC扩增模型返回的误差略低， $\ell_1$ ，表明多样化的2D数据扩增只能部分地帮助防止3D腐败。

**数据集和架构的影响。**我们在图7中提供了对3DCC性能的分类。我们首先观察到，在Taskonomy上训练的基线UNet和DPT模型具有相似的性能，特别是在视图变化的破坏上。通过用OmniData的更大、更多样化的数据进行训练，DPT的性能有所提高。类似的观察也发生在用于分类的视觉转换器上[5,16]。这种改进在视图改变的情况下是不存在的，而对于其他的改变，误差从0.069下降到0.061。这表明，将结构上的进步与多样化和大量的训练数据相结合，可以在对抗3DCC的鲁棒性方面发挥重要作用。此外，当与三维增强技术相结合时，它们可以提高对现实世界中腐败的鲁棒性（第5.3节）。

### 5.2.2 3DCC和2DCC中腐败的冗余度

在图1中，对3DCC和2DCC进行了定性的比较。前者产生了更真实的图像，而后者没有考虑到场景的三维，而是对图像进行了统一的修改。在图8中，我们旨在量化3DCC和2DCC之间的相似性。在图8的左边，我们计算了3DCC和2DCC的相关度。

1，基线模型对一个腐败子集的清洁预测和腐败预测之间的误差（全套在补充文件中）。3DCC在基准内部以及与2DCC相比都产生了较少的相关性（平均相关性是2DCC-2DCC为0.32，3DCC-3DCC为0.28，以及0.30。2DCC-3DCC）。在深度方面也得到了类似的结论估计（在补充文件中）。在右边，我们通过计算干净图像和损坏图像之间的英磅误差，对RGB域进行了同样的分析，再次表明3DCC产生了较低的相关性。

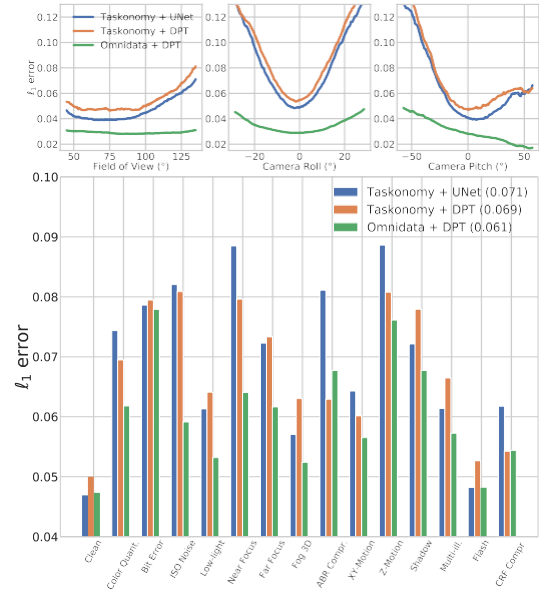


图7.3DCC性能的分类。该基准可以暴露出趋势和模型对各种破坏的敏感性。我们通过在Taskonomy[81]或OmniData[17]上训练模型，并采用UNet[58]或DPT[53]架构来证明这一点。每种腐败的所有移位强度的平均 $\ell_1$ 误差显示（越低越好）。顶部：我们观察到，与OmniData公司的训练模型相比，Taskonomy模型更容易受到视野、相机滚动和俯仰的影响，这与他们的方法一致。底部：图例中的数字是在所有腐败中的平均表现。我们可以看到，所有的模型都对三维破坏很敏感，例如，Z-运动模糊和阴影。总的来说，我们观察到用大型多样化数据（如OmniData）进行训练并使用DPT可以明显提高性能。

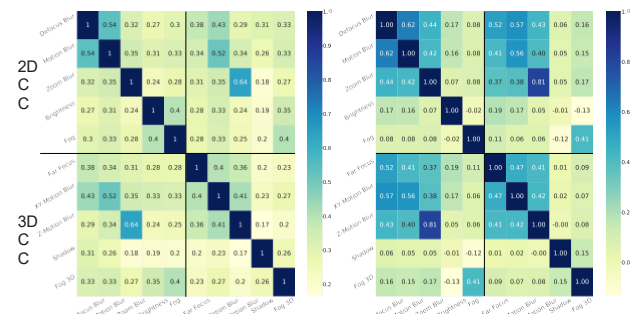


图8.腐败之间的冗余度。我们通过比较2DCC和3DCC的表面法线预测（左）和RGB图像（右）的 $\ell_1$ 误差中的相关性，量化了2DCC和3DCC的腐败子集的成对相似性。3DCC在基准内部以及与2DCC相比都产生了较少的相关性。因此，3DCC有一个多样化的损坏集合，这些损坏与2DCC没有明显的重叠。使用深度作为目标任务会产生类似的结论（完整的亲和矩阵在补充文件中提供）。

### 5.2.3 健全性。3DCC VS 昂贵的合成

3DCC的目的是暴露一个模型在某些真实世界的破坏中的脆弱性。这就要求3DCC生成的破坏与真实的破坏数据相似。由于生成这种标记的数据是昂贵的，而且几乎没有可用的



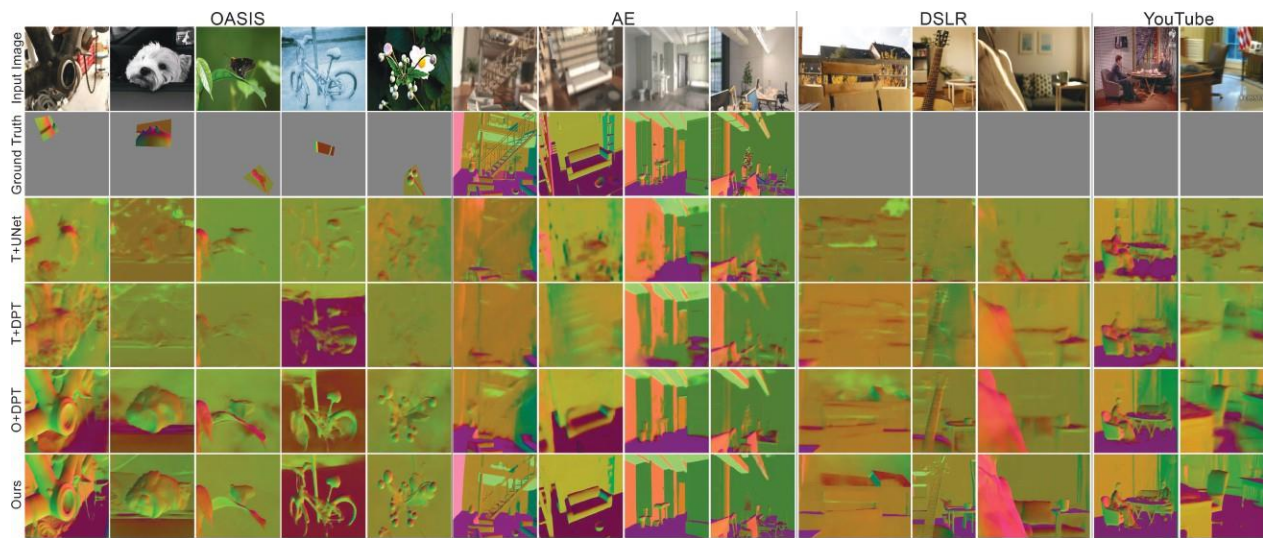


图9.在OASIS[9]、AE（第5.2.3节）、手动收集的单反相机数据和野生的YouTube视频的表面法线的随机查询中，使用3D数据增强学习的定性结果。地面真相不可用时为灰色，例如YouTube。最后一行的预测（我们的）是来自O+DPT+2DCC+3D模型。它们明显更清晰、更准确。更多结果请见项目[页面](#)和[补充材料](#)。

作为替代评估，我们将3DCC的真实性与Adobe After Effects (AE) 的合成进行比较，后者是一种商业产品，用于生成高质量的逼真数据，通常依赖于昂贵的人工处理。为了达到这个目的，我们使用了Hypersim[57]数据集，该数据集带有高分辨率的Z-深度标签。然后我们使用3DCC和AE生成了200幅近景和远景的图像。图10显示了两种方法生成的图像样本，它们在感知上是相似的。接下来，我们计算了当输入来自3DCC或AE时，基线Normal模型的预测误差。<sup>1</sup>误差的散点图在图11中给出，显示了两种方法之间的强烈相关性，即0.80。为了校准和控制，我们还提供了来自2DCC的一些腐败的散点图，以显示相关的意义。它们与AE的相关性明显较低，表明通过3DCC产生的景深效应与AE产生的数据相当匹配。

### 5.2.4 将3DCC应用于其他数据集的效果

我们在图5中定性地展示了3DCC可以应用于标准的视觉数据集，如ImageNet[12]和COCO[38]，利用来自MiDaS[54]的最先进模型的预测深度。在这里，我们定量地展示了使用预测深度而不是地面真相的影响。为此，我们使用了带有地面真相深度标签的Replica[64]数据集。然后，我们使用地面真实深度和来自MiDaS[54]的预测深度生成了1280幅corrupted图像，而没有对Replica进行微调。图12显示了使用地面真相和预测深度生成的3DCC的三种腐败的趋势。趋势是相似的，误差的相关性很强(0.79)。这

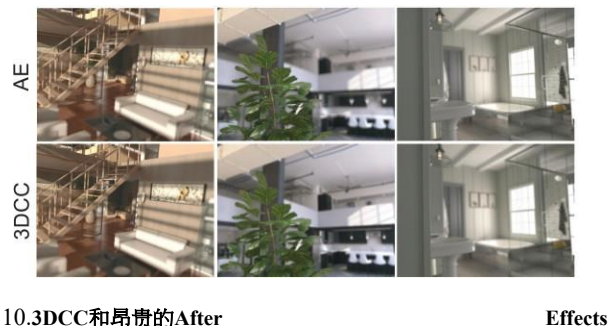


图10.3DCC和昂贵的After Effects (AE)对Hypersim查询图像产生的景深效果的视觉比较。3DCC产生的损坏在视觉上与AE的损坏相似。

这表明，预测的深度可以有效地用于其他数据集的3DCC，而且性能可以随着更好的深度预测而提高。更多关于ImageNet的分析和定量评估请见[补充说明](#)，3DCC可以在模型开发过程中通过暴露非线性趋势和2DCC没有捕捉到的弱点而起形成作用。

### 5.3. 三维数据扩增以提高稳健性

我们从质量和数量上证明了所提议的增强措施的有效性。我们评估了在Taskonomy上训练的UNet和DPT模型（T+UNet, T+DPT）以及在Omnidata上训练的DPT（O+DPT），以了解训练数据集和模型结构的影响。训练过程如5.1节所述。对于其他模型，我们从O+DPT模型初始化，并在此基础上训练2DCC增强（O+DPT+2DCC）和3D增强（O+DPT+2DCC+3D），即我们提出的模型。定性评价。我们考虑 i. OASIS [9], ii.



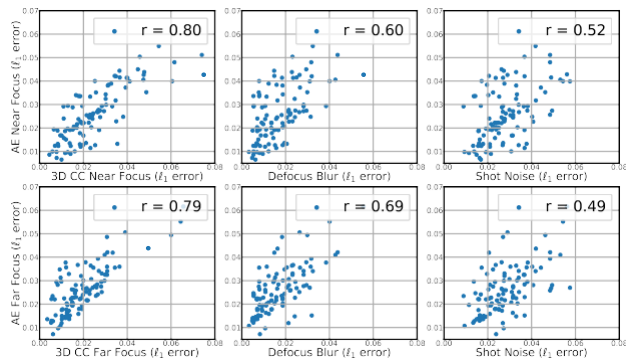


图11.3DCC的破坏与昂贵的现实合成的破坏相似，而产生的成本更低。英镑 errors 的散点图，来自3DCC的基线模型预测与Adobe After Effects (AE) 创建的预测。3DCC的近（远）焦和AE的近（远）焦之间的相关性是最强的（数字在左列的腿部）。我们还添加了来自2DCC的最相似的腐败（失焦模糊）进行比较，产生了较弱的相关性（中间）。镜头噪音（右）是一个控制基线，即一个随机选择的腐败，以校准相关性测量的意义。

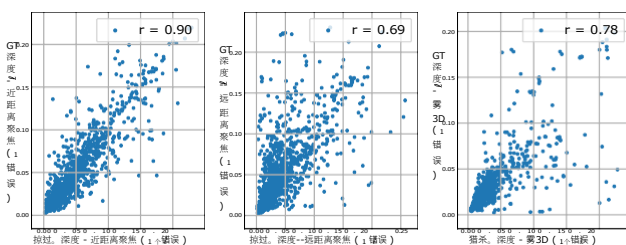


图12.应用3DCC的效果，没有地面真实深度。使用MiDaS[54]模型对未见过的Replica数据进行深度预测，产生了3DCC的三种破坏。散点图显示了当使用预测深度（X轴）或地面实况（Y轴）生成腐败时，基线模型的 $\epsilon_1$ 误差。两种破坏的数据结果之间的趋势相似，表明预测深度是生成3DCC的有效近似值。更多的测试，包括控制基线，请见补充说明。

### 5.2.3节中的AE损坏数据，iii.

手工收集的DSLR数据，以及iv.

野外的YouTube视频。图9显示，与基线相比，建议的模型所做的预测明显更稳健。我们还建议观看项目页面上的片段。

**定量评价。**在表1中，我们计算了模型在2DCC、3DCC、AE和OASIS[9]数据（无微调）上的误差。同样，拟议的模型在不同的数据集上产生了较低的误差，显示了增强的有效性。请注意，在不牺牲野外清洁数据（即OASIS）性能的情况下，对损坏数据的鲁棒性得到了改善。

## 6. 结论和局限性

我们引入了一个框架来测试和提高模型对现实世界分布变化的鲁棒性，特别是那些以三维为中心的分布变化。实验表明，所提出的三维共同破坏是一个具有挑战性的

基准	模型	T+UNet	T+DPT	O+DPT	O+DPT+2DCC	O+DPT+2DCC+3D (我们的)
2DCC[25] (1英镑错误)		8.15	7.47	6.43	5.78	5.32
3DCC (1英镑错误)		7.08	6.89	6.13	5.94	5.42
AE (5.23节) (1英镑错误)		12.86	12.39	7.84	6.50	4.94
OASIS[9] (角度误差)		30.49	32.13	24.42	23.67	24.65

表1.使用不同基准量化的三维增强的效果。1英镑的误差乘以100的阅读能力。我们的模型在所有的基准中产生了较低的误差。2DCC和3DCC被应用在相同的Taskonomy测试图像上。更多的结果是不完整的。在OASIS上的评估有时显示出很大的差异，因为它的地面真相很稀疏。

该基准暴露了模型在现实世界中可能出现的腐败现象下的脆弱性。此外，拟议的数据增强导致了与基线相比更加稳健的预测。我们相信这项工作通过展示三维腐败在基准测试和训练中的作用，为鲁棒性研究开辟了一个有前途的方向。在此，我们简要地讨论一些局限性。

**3D质量：**3DCC是由3D数据的质量所决定的。目前的3DCC是一个不完美但有用的近似值。

正如我们所展示的那样，对现实世界中的3D损坏的模仿。随着更高分辨率的感光数据和更好的深度预测模型，保真度有望提高。

**非穷尽性集合。**我们的一套三维腐败和增强技术并不是详尽无遗的。相反，它们是供研究人员实验用的启动集。该框架可以用于生成更多的特定领域的分布转移，只需最小的人工努力。

**大规模的评估。**虽然我们在分析中评估了一些最近的obustness方法，但我们的主要目标是表明3DCC成功地暴露了漏洞。因此，进行全面的稳健性分析超出了这项工作的范围。我们鼓励研究人员针对我们的腐败测试他们的模型。

**平衡基准。**我们没有明确地平衡我们的基准中的腐败类型，例如有相同数量的噪声和模糊的扭曲。我们的工作可以进一步受益于试图校准腐败基准的平均性能的加权策略，如[36]。

**扩增的使用案例。**虽然我们关注的是稳健性，但在其他应用中调查它们的用处。例如，自我监督学习，可能是值得的。

**评估任务。**我们用密集的回归任务进行实验。然而，3DCC可以应用于不同的任务，包括分类和其他语义任务。使用我们的框架对语义模型的失败案例进行调查，例如对平滑变化的闭塞率，可以提供有用的见解。

**鸣谢。**我们感谢Zeynep Kar和Abhijeet Jagdev。这项工作得到了ETH4D和EPFL EssentialTech中心人道主义行动挑战资助的部分支持。

## 参考文献

- [1] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. 智能手机摄像头的高质量去噪数据集。在 *IEEE 计算机视觉和模式识别会议论文集* 中, 第1692-1700页, 2018.<sup>3</sup>
- [2] Iro Armeni, Zhi-Yang He, JunYoung Gwak, Amir R Zamir, Martin Fischer, Jitendra Malik, and Silvio Savarese. 3D场景图。统一语义、三维空间和凸轮时代的结构。In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5664-5673, 2019.<sup>4</sup>
- [3] Andrei Barbu, David Mayo, Julian Alverio, William Luo, Christopher Wang, Danny Gutfreund, Joshua Tenenbaum, and Boris Katz. Objectnet: 一个大规模的偏见控制的数据集, 用于推动物体识别模型的极限。2019.<sup>2,4</sup>
- [4] Brian A Barsky和Todd J Kosloff. Algorithms for rendering depth of field effects in computer graphics. In *Proceedings of the 12th WSEAS international conference on Computers*, volume 2008. 世界科学与工程院和学会 (WSEAS), 2008.<sup>3</sup>
- [5] Srinadh Bhojanapalli, Ayan Chakrabarti, Daniel Glasner, Daliang Li, Thomas Unterthiner, and Andreas Veit. *arXiv preprint arXiv:2103.14586*, 2021.<sup>2,6</sup>
- [6] Tim Brooks和Jonathan T Barron. 学习合成运动模糊。在 *IEEE/CVF 计算机视觉和模式识别会议* 上, 第6840-6848页, 2019.<sup>3</sup>
- [7] Prithvijit Chattopadhyay, Judy Hoffman, Roozbeh Mottaghi, and Aniruddha Kembhavi. Robustnav: 争取在具身导航中实现基准的稳健性。 *arXiv 预印本 arXiv:2106.04531*, 2021.<sup>2</sup>
- [8] 陈晨, 陈奇峰, 徐佳, 和Vladlen Koltun. 学习在黑暗中观察。在 *IEEE 计算机视觉和模式识别大会* 上, 第3291-3300页, 2018.<sup>3</sup>
- [9] 陈伟峰, 钱晟怡, 范大伟, 小岛纪之, Max Hamilton和邓佳. Oasis: 一个用于野外单幅图像的大规模数据集。在 *IEEE/CVF 计算机视觉和模式识别会议* 上, 第679-688页, 2020.<sup>7,8</sup>
- [10] Blender在线社区. *Blender* - 一个三维建模和渲染包。Blender基金会, Stichting Blender基金会, 阿姆斯特丹, 2018.<sup>3</sup>
- [11] Francesco Croce, Maksym Andriushchenko, Vikash Sehwag, Edoardo De Benedetti, Nicolas Flammarion, Mung Chiang, Prateek Mittal, and Matthias Hein. Robustbench: 一个标准化的对抗性鲁棒性基准。 *arXiv 预印本 arXiv:2010.09670*, 2020.<sup>3</sup>
- [12] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: 一个大规模的分层图像数据库。In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248-255. Ieee, 2009.<sup>2,4,5,7</sup>
- [13] Terrance DeVries and Graham W Taylor. *ArXiv preprint arXiv:1708.04552*, 2017.<sup>1,4</sup>
- [14] Josip Djolonga, Jessica Yung, Michael Tschannen, Rob Romijnders, Lucas Beyer, Alexander Kolesnikov, Joan Puigcerver, Matthias Minderer, Alexander D'Amour, Dan Moldovan, et al. On robustness and transferability of convolutional neural networks. 在 *IEEE/CVF 计算机视觉和模式识别会议* 上, 第16458-16468页, 2021.<sup>2</sup>
- [15] Samuel Dodge和Lina Karam. 视觉失真下人类和深度学习识别性能的研究和比较。在 *2017年第26届国际计算机通信与网络会议 (ICCN)* 上, 第1-7. IEEE, 2017.<sup>1</sup>
- [16] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: *arXiv preprint arXiv:2010.11929*, 2020.<sup>6</sup>
- [17] Ainaz Eftekhari, Alexander Sax, Jitendra Malik, and Amir Zamir. Omnidata: 一个可扩展的管道, 用于从三维扫描中制作多任务中级视觉数据集。 *IEEE/CVF 计算机视觉国际会议论文集*, 第10786-10796页, 2021.<sup>3, 5, 6</sup>
- [18] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing*, 15(12):3736-3745, 2006.<sup>3</sup>
- [19] Raanan Fattal. Single image dehazing. *ACM图形交易 (TOG)*, 27(3):1-9, 2008.<sup>3,4</sup>
- [20] Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T Roweis, and William T Freeman. 从一张照片中消除相机的抖动。在 *ACM SIGGRAPH 2006* 论文中, 第787-794页。2006.<sup>3</sup>
- [21] Alessandro Foi, Mejd Trimeche, Vladimir Katkovnik, and Karen Egiazarian. 实用的poissonian-gaussian噪声调制和单幅图像原始数据的拟合。 *IEEE Transactions on Image Processing*, 17(10):1737-1754, 2008.<sup>3,4</sup>
- [22] Nic Ford, Justin Gilmer, Nicolas Carlini, and Dogus Cubuk. 对抗性例子是噪声中测试错误的自然结果。 *arXiv 预印本 arXiv:1901.10513*, 2019.<sup>2</sup>
- [23] Robert Geirhos, Jörn-Henrik Jacobsen, Claudio Michaelis, Richard Zemel, Wieland Brendel, Matthias Bethge, and Felix A Wichmann. 深度神经网络中的捷径学习。 *arXiv 预印本 arXiv:2004.07780*, 2020.<sup>1</sup>
- [24] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A Wichmann, and Wieland Brendel. Imagenet训练的cnns偏向于纹理: 增加形状偏向可以提高准确性和鲁棒性。 *arXiv 预印本 arXiv:1811.12231*, 2018.<sup>2,5</sup>
- [25] Majed El Helou, Ruofan Zhou, Johan Bartheas, and Sabine Suessstrunk. Vidit: 用于照明转换的虚拟图像数据集, *arXiv 预印本 arXiv:2005.05460*, 2020.<sup>3</sup>
- [26] Dan Hendrycks, Steven Basart, Norman Mu, Saurav Kadavath, Frank Wang, Evan Dorundo, Rahul Desai, Tyler Zhu, Samyak Parajuli, Mike Guo, 等. The many faces of robustness: 对分布外泛化的批判性分析。在 *IEEE/CVF 计算机视觉国际会议论文集* 中, 第8340-8349页, 2021.<sup>2,5</sup>
- [27] Dan Hendrycks and Thomas Dietterich. *arXiv preprint arXiv:1903.12261*, 2019.<sup>1,2,5,8</sup>

- [28] Dan Hendrycks, Kimin Lee, and Mantas Mazeika. 使用预训练可以提高模型的鲁棒性和不确定性。在*国际机器学习会议上*, 第2712-2721页。PMLR, 2019.[2](#)
- [29] Dan Hendrycks, Norman Mu, Ekin D Cubuk, Barret Zoph, Justin Gilmer, and Balaji Lakshminarayanan. Augmix: 一个简单的数据处理方法, 以提高鲁棒性和不确定性。 *arXiv预印本 arXiv:1912.02781*, 2019.[2](#)
- [30] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, and Pheng-Ann Heng. 用于单幅图像除雨的深度注意力特征。 *IEEE/CVF 计算机视觉与模式识别会议论文集*, 第8022-8031页, 2019.[3](#)
- [31] Jason Jo和Yoshua Bengio. 测量cnns学习表面统计规律性的趋势。 *arXiv预印本 arXiv:1711.11561*, 2017.[1](#)
- [32] Christoph Kamann和Carsten Rother. 语义分割模型的稳健性基准。在*IEEE/CVF 计算机视觉和模式识别会议上*, 第8828-8838页, 2020.[2](#)。
- [33] Sanjay Kariyappa and Moinuddin K Qureshi. 用多样性训练提高集合的对抗性稳健性。 *arXiv预印本 arXiv:1901.09981*, 2019.[2](#)
- [34] Deepa Kundur 和 Dimitrios Hatzinakos. 盲目的图像去卷积。 *IEEE 信号处理杂志*, 13(3) : 43- 64, 1996.[3](#)
- [35] Alexey Kurakin, Ian Goodfellow, and Samy Bengio. Adversarial machine learning at scale. *arXiv preprint arXiv:1611.01236*, 2016.[3,5](#)
- [36] Alfred Laugros, Alice Caplier, and Matthieu Ospici. 使用合成腐败来衡量对自然分布转移的鲁棒性。 *arXiv预印本 arXiv:2107.12052*, 2021.[8](#)
- [37] Guillaume Leclerc, Hadi Salman, Andrew Ilyas, Sai Vempala, Logan Engstrom, Vibhav Vineet, Kai Xiao, Pengchuan Zhang, Shibani Santurkar, Greg Yang, 等. 3db: *arXiv preprint arXiv:2106.03805*, 2021.[2](#)
- [38] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: 上下文中的常见对象。在*欧洲计算机视觉会议上*, 第740-755页。Springer, 2014.[4,5,7](#)
- [39] Raphael Gontijo Lopes, Dong Yin, Ben Poole, Justin Gilmer, and Ekin D Cubuk. 在不牺牲精度的情况下用补丁高斯增强来提高鲁棒性。 *arXiv预印本 arXiv:1906.02611*, 2019.[2](#)
- [40] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*, 2017.[2,3](#)
- [41] Julien Mairal, Francis Bach, Jean Ponce, Guillermo Sapiro, and Andrew Zisserman. 用于图像修复的非局部稀疏模型。在*2009年IEEE第12届计算机视觉国际会议上*, 第2272-2279页。IEEE, 2009.[3](#)
- [42] Claudio Michaelis, Benjamin Mitzkus, Robert Geirhos, Evgenia Rusak, Oliver Bringmann, Alexander S Ecker,

Matthias Bethge, and Wieland Brendel. 物体检测中的罗盘基准。 *arXiv preprint arXiv:1907.07484*, 2019.[2](#)



- [43] John P Miller, Rohan Taori, Aditi Raghunathan, Shiori Sagawa, Pang Wei Koh, Vaishaal Shankar, Percy Liang, Yair Carmon, and Ludwig Schmidt. 准确度在行。关于分布外和分布内泛化之间的强关联性。在*国际机器学习会议上*, 第7721-7735页。PMLR, 2021.[2](#)
- [44] Eric Mintun, Alexander Kirillov, and Saining Xie. On interaction between augmentations and corruptions in natural corruption robustness. *arXiv preprint arXiv:2102.11273*, 2021.[2](#)
- [45] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. 用于动态场景去模糊化的深度多尺度卷积神经网络。In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3883-3891, 2017.[3](#)
- [46] Seungjun Nah, Sanghyun Son, Suyoung Lee, Radu Timofte, and Kyoung Mu Lee. Ntire 2021关于图像去模糊化的挑战。在*CVPR研讨会上*, 第149-165页, 2021年6月。[3](#)
- [47] Muzammal Naseer, Kanchana Ranasinghe, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. 视觉变换器的诱人特性。 *arXiv预印本 arXiv:2105.10497*, 2021.[2,4](#)
- [48] Simon Niklaus, Long Mai, Jimei Yang, and Feng Liu. 来自单一图像的3D肯烧伤效果。 *ACM Transactions on Graphics (TOG)*, 38(6):1-15, 2019.[3](#)
- [49] A Emin Orhan. facebook's resnext wsl models. *arXiv preprint arXiv:1907.07640*, 2019.[2](#) [50] Tianyu Pang, Kun Xu, Chao Du, Ning Chen, and Jun Zhu. 通过促进集合多样性提高对抗性稳健性。 *arXiv预印本 arXiv:1901.08846*, 2019.[2](#)
- [51] Amandalynne Paullada, Inioluwa Deborah Raji, Emily M Bender, Emily Denton, and Alex Hanna. 数据和它的(dis)内容。 a survey of dataset development and use in machine learning research. *arXiv preprint arXiv:2012.05345*, 2020.[2](#)
- [52] Michael Potmesil和Indranil Chakravarty. 用于合成图像生成的镜头和光圈相机模型。 *ACM SIGGRAPH Computer Graphics*, 15(3):297-305, 1981.[3](#)
- [53] Ren e' Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. 密集预测的维度变换器。在*IEEE/CVF国际计算机视觉会议论文集*, 第12179-12188页, 2021.[5,6](#)
- [54] Ren e' Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. 迈向强大的单眼深度估计。 *arXiv预印本 arXiv:1907.01341*, 2019.[5,7,8](#)
- [55] Sashank J Reddi, Satyen Kale, and Sanjiv Kumar. 论adam的收敛性及其他。 *arXiv预印本 arXiv:1904.09237*, 2019.[5](#)
- [56] Jaesung Rim, Haeyun Lee, Jucheol Won, and Sunghyun Cho. 用于学习和衡量去模糊算法的真实世界模糊数据集。在*欧洲计算机视觉会议上*, 第184-201页。Springer, 2020.[3](#)
- [57] Mike Roberts, Jason Ramapuram, Anurag Ranjan, Atulit Kumar, Miguel Angel Bautista, Nathan Paczan, Russ Webb, and Joshua M

Susskind. Hypersim: 一个用于整体室内场景理解的逼真的协同数据集。在*IEEE/CVF计算机视觉国际会议论文集*中, 第10912-10922页, 2021.[7](#)

- [58] Olaf Ronneberger, Philipp Fischer, and Thomas Brox.U-net:用于生物医学图像分割的卷积网络。在*国际医学图像合成和计算机辅助干预会议上*, 第234-241页。Springer, 2015.[5,6](#)
- [59] Evgenia Rusak, Lukas Schott, Roland Zimmermann, Julian Bitterwolf, Oliver Bringmann, Matthias Bethge, and Wieland Brendel.通过玩噪音游戏增加dnn对图像损坏的鲁棒性。2020.[2](#)
- [60] Evgenia Rusak, Lukas Schott, Roland S Zimmermann, Julian Bitterwolf, Oliver Bringmann, Matthias Bethge, and Wieland Brendel.使神经网络对各种图像损坏具有鲁棒性的简单方法。In *European Conference on Computer Vision*, pages 53-69.Springer, 2020.[2](#)
- [61] Christos Sakaridis, Dengxin Dai, and Luc Van Gool.Semantic foggy scene understanding with synthetic data.*International Journal of Computer Vision*, 126(9):973-992, 2018.[3,4](#)
- [62] Nithya Sambasivan, Shivani Kapania, Hannah Highfill, Diana Akrong, Praveen Paritosh, and Lora M Aroyo."每个人都想做模型工作, 而不是数据工作"。高风险AI中的数据串联。在*2021年CHI计算系统中人的因素会议的论文集*中, 第1-15页, 2021.[2](#)
- [63] Rulin Shao, Zhouxing Shi, Jinfeng Yi, Pin-Yu Chen, and Cho-Jui Hsieh.论视觉转换器的对抗性鲁棒性。*arXiv预印本arXiv:2103.15670*, 2021.[2](#)
- [64] Julian Straub, Thomas Whelan, Lingni Ma, Yufan Chen, Erik Wijmans, Simon Green, Jakob J. Engel, Raul Mur-Artal, Carl Ren, Shobhit Verma, Anton Clarkson, Mingfei Yan, Brian Budge, Yajie Yan, Xiaqing Pan, June Yon, Yuyang Zou, Kimberly Leon, Nigel Carter, Jesus Briales, Tyler Gillingham, Elias Mueggler, Luis Pesqueira, Manolis Savva, Dhruv Batra, Hauke M. Strasdat, Renzo De Nardi, Michael Goesele, Steven Lovegrove, and Richard Newcombe.Replica数据集。a digital replica of indoor spaces. *arXiv preprint arXiv:1906.05797*, 2019.[4,7](#)
- [65] Jiachen Sun, Qingzhao Zhang, Bhavya Kailkhura, Zhiding Yu, Chaowei Xiao, and Z Morley Mao.对照常见的corruptions, 对三维点云识别的robustness进行基准测试。*arXiv preprint arXiv:2201.12296*, 2022.[2](#)
- [66] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus.神经网络的诱人特性。*arXiv预印本arXiv:1312.6199*, 2013.[1,3](#)
- [67] Rohan Taori, Achal Dave, Vaishaal Shankar, Nicholas Carlini, Benjamin Recht, and Ludwig Schmidt.测量图像分类中的自然分布偏移的robustness。*arXiv预印本arXiv:2007.00644*, 2020.[2](#)
- [68] Maxime Tremblay, Shirsendu Sukanta Halder, Raoul de Charette, and Jean-Francois Lalonde.用于评估和提高对恶劣天气的鲁棒性的雨景渲染。In *International Journal of Computer Vision*, 129(2):341-360, 2021.
- [69] Alexander Von Bernuth, Georg Volk, and Oliver Bringmann.在现有图像上模拟照片般真实的雪和雾, 用于增强Cnn训练和评估。In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 41-

46.IEEE, 2019.[3,4](#)

- [70] Neal Wadhwa, Rahul Garg, David E Jacobs, Bryan E Feldman, Nori Kanazawa, Robert Carroll, Yair Movshovitz-Attias, Jonathan T Barron, Yael Pritch, and Marc Levoy.用单镜头手机合成景深。 *ACM Transactions on Graphics (ToG)*, 37(4):1-13, 2018.[3](#)
- [71] 王立军, 沈晓辉, 张建明, Oliver Wang, 林哲, 谢志尧, Sarah Kong, 和Huchuan Lu. Deeplens:来自单一图像的浅景深。 *arXiv预印本arXiv:1810.08100*, 2018.[3](#)
- [72] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli.图像质量评估:从错误可见性到结构相似性。 *IEEE Transactions on Image Processing*, 13(4):600-612, 2004.[4](#)
- [73] 魏凯旋, 傅莹, 杨蛟龙, 和黄华。基于物理学的极端低照度原始去噪模型。在*IEEE/CVF 计算机视觉和模式识别会议论文集中*, 第2758-2767页, 2020.[3](#)
- [74] Qizhe Xie, Minh-Thang Luong, Eduard Hovy, and Quoc V Le.带有噪声的学生的自我训练提高了图像网络的分类。 *IEEE/CVF 计算机视觉和模式识别会议论文集*, 第10687-10698页, 2020.[2](#)
- [75] Zexiang Xu, Kalyan Sunkavalli, Sunil Hadap, and Ravi Ramamoorthi.基于深度图像的最佳稀疏样本的再照明。 *ACM Transactions on Graphics (ToG)*, 37(4):1-13, 2018.[3](#)
- [76] 杨焕瑞, 张景阳, 董洪亮, Nathan Inkawhich, Andrew Gardner, Andrew Touchet, Wesley Wilkes, Heath Berry, 和李海。Dverge:为增强集合体的稳健生成而使脆弱性多样化。 *神经信息处理系统进展*, 33, 2020.[2](#)
- [77] Teresa Yeo, Oğuzhan Fatih Kar, and Amir Zamir.通过跨域组合的Robustness.在*IEEE/CVF国际计算机视觉会议 (ICCV) 论文集*, 第12189-12199页, 10月2021.[2](#), [5](#)
- [78] Chenyu Yi, Siyuan Yang, Haoliang Li, Yap-peng Tan, and Alex Kot.*arXiv preprint arXiv:2110.06513*, 2021.[2](#),[4](#).
- [79] Dong Yin, Raphael Gontijo Lopes, Jon Shlens, Ekin Dogus Cubuk, and Justin Gilmer.计算机视觉中模型稳健性的傅里叶视角。In *Advances in Neural Information Processing Systems*, pages 13276-13286, 2019.[2](#)
- [80] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo.Cutmix:用正则化策略训练具有可定位特征的强分类器。In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6023-6032, 2019.[2](#)
- [81] Amir R Zamir, Alexander Sax, William Shen, Leonidas J Guibas, Jitendra Malik, and Silvio Savarese.Taskonomy:分离任务转移学习。在*IEEE 计算机视觉和模式识别会议上*, 第3712-3722页, 2018.[4](#),[5](#),[6](#)
- [82] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup:Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017.[2](#)



- [83] 张凯，左王蒙，陈云锦，孟德宇，和张磊。超越高斯去噪器。用于图像去噪的深度Cnn的剩余学习。 *IEEE transactions on image processing*, 26(7):3142-3155, 2017.3
- [84] 张凯，左王蒙，顾书航，和张磊。学习用于图像修复的深度Cnn去噪器先验。在 *IEEE 计算机视觉和模式识别会议* 上，第3929-3938页，2017.3