摘要：

红外目标检测拥有很广泛的应用价值，但是由于红外图像的特性，红外目标很容易淹没在复杂的背景中。现如今的检测算法并不能满足真实场景下虚警率和检测率的需求。为此我们设计了含有大尺寸感受野的红外弱小目标检测算法。它使用了大尺寸卷积层构成的残差网络，因此可以拥有更大的有效感受野，可以提高模型的鲁棒性。此外通过transformer结构的注意力机制，可以使网络更好地定位目标所在区域，提高模型的检测效果。实验结果显示我们所提出的算法在各项评价指标均超过了其他先进算法。

引言：

红外目标检测技术由于其不受天气影响和电磁干扰等特性在军事和民用均有着广泛的应用。为了能够及时发现目标，探测器需要能够探测很大的范围，因此探测器一般距离目标很远，因此目标在图像中呈现小而弱的点状形态。此类目标被称为红外弱小目标。

与可见光相比，红外图像缺少纹理信息，目标与背景的边界不是非常突出，因此难以在复杂背景中识别出红外弱小目标，导致检测率下降。并且由于红外探测器盲元和噪点的现象，会导致虚警率上升。因此在实际应用中急需高检测率和低虚警率的红外弱小目标检测算法。

现如今的红外弱小目标检测算法大致可以分为传统算法和深度学习算法。

传统的算法依据一定的先验知识进行目标检测，其大致可以分为基于滤波器的方法，基于人类视觉系统的方法和基于数据结构的方法。基于滤波器的算法通过设计特定的滤波器，去除背景信息，以检测目标；其计算量比较少，实现简单，但是其对噪声和杂波干扰极为敏感，只能适合背景简单的场景。基于人类视觉系统的方法使用对比度进行区分目标和背景，其能够增强目标并且抑制背景，但是其对物体的边缘极为敏感，只能适合目标与背景特征差异较大的场景。基于数据结构的算法主要利用目标的稀疏性和背景的低秩性，将目标检测转化为数学求解问替，其效果较好，适应于大多场景，但是其大量的矩阵计算会导致实现困难，并且比较耗时。

近些年来，随着深度学习技术的发展，越来越多的深度学习算法被应用到了红外弱小目标检测领域。由于红外目标的尺寸极小，因此直接使用可见目标的检测网络极容易导致红外目标的丢失；且由于红外弱小目标在图像中大多呈现点状形态，因此其对定位框的偏移极为敏感，因此大多深度学习红外弱小目标检测算法都是基于图像分割的。

戴等首先提出了基于图像分割的红外弱小目标检测算法ACM，之后在此基础上增加了局部对比机制，设计出来了ALC-Net。之后一些算法使用对抗生成网络，使用不同的生成器和辨别器来平衡红外目标检测中的虚警和漏检平衡问题。一些算法使用反复地跳跃连接，以保留特征图中红外弱小目标的细节信息，更好地检测红外弱小目标。

但是现如今的红外弱小目标检测算法仍存在一定的问题。传统算法大多依赖于一定的先验知识，当真实场景于其不符合时，传统算法的表现就会很差。而基于深度学习的算法相对于传统算法有很大的提升，但是它们仍存在一定的不足。它们大多不能同时取得高的检测率和低的虚警率，并且一些算法的通用性不足，不能满足真实场景的检测需求。

在本文中，我们提出一种新的基于卷积网络的红外弱小目标检测算法：基于大卷积核和transformer结构的红外弱小目标检测算法。它使用了大尺寸卷积核构成的Res-Net，可以提高模型的感受野，提高模型的鲁棒性；此外通过使用全卷积构成的类似transformer结构，可以避免transformer的难以训练的缺点。实验证明，我们的算法相较于其他先进算法，取得了最好的。

总的来说，本文所作的贡献如下所示;

1. 我们设计了一个具有大尺寸卷积核的残差网络，使模型拥有了更大的感受野，提高了模型的鲁棒性，并且消融实验证明了增加大尺寸卷积核的有效性。
2. 我们设计了一个用全卷积组成的类似transformer结构，在利用了transformer结构的优点的前提下，避免了该结构难以训练的缺点。
3. 与其他算法相比，我们的算法取得了更好的检测率和虚警率。

相关工作

红外弱小目标检测：

基于滤波器的方法是最开始用来进行红外弱小目标检测的算法，其可以根据作用域的不同，可以分为频域滤波和空间域滤波。空间域滤波主要利用目标和背景直接的灰度差异来进行检测，其大致有空域高通滤波，均值滤波，双边滤波和数学形态学等方法。频域滤波的方法主要利用目标和背景在频域中成分不同，通过去除背景部分来实现小目标检测，这类方法大致有频域高通滤波，小波变换滤波等。频率滤波的方法相比于空间域滤波能够更好地区分目标和背景，因此检测效果更好，但是其计算复杂度也更高。总的来说基于滤波器的方法实现简单，计算资源消耗很少，在背景均匀的场景能够取得不错的效果，但是在背景比较复杂的场景，其效果不能满足检测要求。

基于人类视觉系统的检测算法模仿人眼定位方式，使用对比度区分目标和背景，其主要有局部对比机制和多特征融合等方法。基于局部对比机制的方法主要利用目标和背景之间地对比度不同，将某一区域乘以其的灰度值和周围灰度值之比，从而增大目标区域和背景区域的差值，从而可以检测出目标。基于多特征融合的方式通常需要考虑多个特征对小目标局部特性进行表征, 使用不同的特征来去除具有不用的背景信息，如使用局部强度去除噪声和盲元干扰，使用梯度信息来消除物体边缘干扰，使用局部灰度残差信息来抑制平缓的背景等。基于HSV方法能够增强目标并且抑制背景，但是其容易被背景的起伏干扰，造成虚警率升高，此外此类方法容易造成目标细节信息丢失，给后续辨别带来麻烦。

基于图像数据结构的算法主要利用背景的自相似性和目标的稀疏性，其主要有红外图像块模型和稳健主成分分析等方法。红外图像块模型将目标检测问题转化为恢复低秩和稀疏矩阵的优化问题，而稳健主成分分析则转化为凸优化问题。这类算法能够能够在目标数量较少的情景检测到目标，但是当目标数据较多时，目标的稀疏性会降低，会出现一定的漏检现象。且该算法需要大量矩阵计算，实时性不足。

由于深度学习算法在鲁棒性和泛化能力均超越了传统算法，越来越多的研究人员使用深度学习算法进行红外弱小目标检测。

王等人设计了基于生成对抗网络网络的红外弱小目标检测算法，其使用了两个不同生成器，将目标检测分为降低虚警和漏检的两个子任务，以同时达到低虚警和低漏检。但是其辨别器中含有大量全连接，因此其在训练时需要输入固定的图像尺寸，会造成一定的麻烦。

侯等人使用了ResNeSt,将特征图在通道维度上分为不同的组，因此每个组训练后可以获得不同的权重，因此可以保留不同的目标信息。此外其引入了全连接，以增加网络的非线性能力，但是其也因此需要输入固定尺寸的图像。

张等人设计了注意力金字塔网络，通过注意力机制来提取局部和全局范围内的目标信息，并且通过融合机制来融合不同层级提取出的目标信息。但是其中大量的注意力机制，会带来大量的矩阵计算，会消耗大量的计算资源。

大尺寸卷积核：

现如今的深度学习模型大都小而深，既堆叠大量小尺寸的卷积层。但是随着transformer的爆火，研究人员也在探索为什么transformer能够比cnn表现更好。ViT和swint通过使用全局或者局部自注意机制，可以获得远远大于普通卷积的感受野，一些学者认为这是transformer可以在计算机视觉领域超过cnn的一部分原因。

参考文献1设计了超大卷积核架构，通过增加卷积核的大小，增加网络的感受野，并且通过使用深度可分离卷积来避免参数过多，其在语义分割和目标检测领域超过了使用swint的模型。

参考文献2通过对比四个不同尺寸的卷积层的实验结果，通过增加卷积核的大小，该模型的准确率提高了，而计算量没有太多增加。此外当前深度学习框架以及硬件更加适配大尺寸卷积层，因此其对速度的影响并不大。

注意力机制：

注意力机制是模仿人类注意力视觉和认知系统的方法，它可以给特征图中的不同部分赋予不同的权重，从而更加突出特征图中更重要的信息。通过引入注意力机制，网络可以自主学习并且关注到更重要的信息，从而可以提高模型的泛化能力和鲁棒性。其中在计算机视觉中常使用的注意力机制大致可以分为通道注意力机制，空间注意力机制和自注意力机制。

通道注意力机制是在通道维度给不同的通道赋予不同的权重，其中最为经典的为SE-Net。SE-Net对特征图在空间维度进行池化，使空间分辨率变为1\*1，之后经过全连接，获得通道权重，最后与初始特征图相乘。

空间注意力机制是在空间维度上给特征图的不同区域赋予不同的权重。空间变换神经网络(STN)是空间注意力机制的经典模型。它通过局部网络和网络生成器可以预测出物体的变形，并且关注到图像中的重点区域，可以通过变形后的图像获得变形前的物体。

自注意力机制原本是应用到自然语言处理等序列信息处理的，随着ViT的爆火，自注意力机制也被应用到了计算机视觉领域。自注意力机制通过计算特征图内部不同元素的相关程度，可以捕获到长距离的信息，因此可以拥有全局的感受野。但是其需要大量的矩阵计算，会大量计算和存储资源消耗，此外其大量的全连接会使模型的参数巨大，给训练带来困难。

方法：

整体结构：

作为图像分割领域的基线模型，U-Net由于其效果卓越，被应用到语义分割的各个领域。U-Net使用对称的编解码结构，并且通过跳跃连接连接。编码器对输入图像进行下采样；解码器对图像进行上采样，并与跳跃连接中的原始图像进行拼接，以恢复图像中的细节信息。这种结构使网络能够拥有不同大小的感受野，而且保留了图像中的细节信息，因此，我们在深度学习网络中使用U-Net结构。

网络整体结构如图所示。编码器和解码器的结构是对称的。前几层编码器由大尺寸卷积层构成的残差网络和fct组成，而最后一层由并行的inception模块和fct组成。

大尺寸卷积核：

用大量小尺寸卷积层代替少量大尺寸卷积层可以实现相同的理论感受野，并且降低模型的参数量。多个大小为k的卷积层的理论感受野rf和卷积层的数量n成线性关系，其可以用公式1描述，此时需要的参数量w为n\*k\*k。所以当rf固定时，参数量和卷积层的数量成线性关系，和卷积层大小的平方成线性关系。所以使用多个小尺寸的卷积层代替大尺寸的卷积层确实能够减少模型的参数量。

Rf = 1+n(k-1)

但是，堆叠大量的小尺寸的卷积层的有效感受野实际是小于少量的大尺寸卷积层的。根据有效感受野的特性，感受野内的影响因子随着和感受野中心的距离迅速衰减。感受野边缘的元素需要乘以多个卷积层的参数才能影响最终的结果，并且越靠近边缘，需要乘以的参数越多。此外感受野边缘对最终结果影响的路径很少，而感受野中心的元素可以通过多个路径影响到最终的结果。因此大量小尺寸卷积层的有效感受野是远远小于大尺寸卷积层的。

为此，我们设计了具有大尺寸卷积层的残差网络。为了避免大尺寸卷积层的参数量过大，我们使用深度可分离卷积，将卷积核大小大于5的大尺寸卷积层分为深度卷积和逐点卷积。其具体结构如图所示。残差网络中由多个不同尺寸的卷积层组成，其卷积核大小是逐渐变小的，这样小尺寸卷积层能够使用之前大尺寸卷积层处理后的数据，在一定程度上能够增大小尺寸卷积层的感受野。

如图左边所示，输入特征图经过1\*1大小的卷积层改变通道数量后，获得通道数量为输出通道数量的特征图，之后将其输入到i+1个不同尺寸的卷积块，之后再与其相加，最后输出到下一层。如图右边所示，卷积块内部为由两个卷积和一个池化层一个激活函数组成的残差网络。卷积块中由深度可分离卷积块和普通卷积块组成，卷积核大于5的卷积层使用深度可分离卷积，卷积核小于等于5的使用普通卷积。

Inception:

Inception使用不同大小的卷积核并行计算，可以使网络提取到不同信息，从而使网络能够处理含有不同大小目标的场景。此外，当网络比较深的时候，梯度传播比较困难，容易出现梯度消失的现象。而inception并行的结构可以给梯度传播增加新的路径，在一定程度上可以缓解梯度消失的现象。

因此我们在u型网络的底层使用了inception模块，其结构图如图所示。Inception模块使用了四个不同的分支。第一个分支为大小为3\*3，膨胀系数为2的卷积层，其可以实现5\*5的感受野。第二个分支为一个大小为1\*1和一个大小为3\*3的卷积层，其可以实现3\*3的感受野。第三个分支为两个大小为3\*3的卷积层，其可以实现5\*5的感受野。第四个分支为池化层和一个大小为1\*1的卷积层，通过池化操作，可以提取到不同的信息。此外，为了避免增加分支导致的参数量过多，我们将每个分支的通道数量降为输出通道数量的1/4.

Fully Convolutional Transformer：

Transformer使用自注意力机制和多层感知机，可以获得较大的感受野，但是其大量的矩阵计算和全连接会带来极大的计算量和模型参数。我们设计了全部基于卷积层的Transfomer。其与transformer的结构相同，均由注意力机制和多层感知机组成。注意力机制由串联的通道注意力和空间注意力机制组成。通道注意力在空间方向进行平均池化和最大池化，然后经过1维卷积，生成通道注意力权重。空间注意力在通道方向进行池化，然后经过二维卷积，生成空间注意力权重。通过增大注意力机制中的卷积层大小，可以实现较大的感受野，但是计算量和模型参数量会大大降低。

Transformer中输入特征图经过变形后，通道维度为最后的维度，所以MLP中的全连接在通道维度进行的，而这与1\*1大小的卷积层是相同的，因此我们使用1\*1大小的卷积层代替了全连接。虽然这样参数量不会改变，但是这样不需要对张量进行变形操作，因此可以节省模型推理时间。此外我们使用了逆瓶颈结构，在经过第一个卷积层后，通道数量变为原来的mlp倍，之后经过第二个卷积层时通道数量又恢复成原来的通道数量。通过逆瓶颈结构，可以增强信息的传播，时网络能够捕获到更多的信息。

Transformer中使用的归一化为layer 归一化，其在layer维度上进行归一化，更适合序列数据。而batch norm在batch维度进行归一化，更适合图像等数据。此外，layernorm对于输入特征图尺寸敏感，其只能输入固定大小的图像。因此我们并没有使用transformer中的layer norm，将其替代为卷积网络中常用的batch norm。

实验：

消融实验

为了验证我们所提出算法的有效性，我们在上述的数据集上进行了消融实验。

首先我们对比了使用不同的大小的卷积层模型的性能，通过改变残差网络中的最大卷积核大小，可以改变模型的卷积层大小。然后我们对比了在最底层使用Inception和残差网络的模型性能，以验证在底层引入多个分支对模型性能的影响。最后我们对比了使用transformer结构注意力和仅使用通道和空间注意力模型的性能，以验证transformer对模型性能的影响。在进行消融实验时，我们使模型的其他部分保持不变。

我们对比了使用从3到13不同大小的卷积层模型的性能，它们在不同数据集上的量化指标如表所示。由于检测率是检测出的目标数量比上总的目标数量，所以检测出相同的目标数量时，检测率也是一样的；所以有几个不同的网络取得了相同的检测率。可以看到，在最大卷积核大小小于11时，模型的性能随着卷积核的增大而变好，当卷积核大小增大到13时，模型的性能有所下降，但仍要比最大卷积核大小为5和7的网络要好。说明了通过增大卷积层大小能够增大网络的有效感受野，提高模型的检测效果，并且即使将卷积核大小增加过大，仍要比使用小尺寸卷积层网络检测效果好。

我们对比了在最底层使用残差网络和inception模块的模型性能，他们在不同数据集上的量化指标如表所示。可以看到使用inception模块的检测率和虚警率均优于使用残差网络的模型。说明了在网络最底层使用并行的网络结构能够使网络提取到不同的信息，能够增加网络的梯度传播，提高网络的训练效率。

我们对比了transformer结构注意力和仅使用通道和空间注意力模型的性能，他们在不同数据集上的量化指标如表所示。可以看到使用transformer结构注意力的检测率和虚警率均优于仅使用通道和空间注意力模型的模型。这说明了TSAM中的逆瓶颈结构能够保留更多的信息，可以比CASM更好地定位目标所在位置，因此可以提高模型的检测精度。

和其他先进算法的对比实验：

我们选取了多个传统算法和深度学习算法。传统算法包括基于滤波器的算法FKRW，基于HSV的算法MPCM和基于数据结构的算法IPI。深度学习算法包括使用生成对抗网络的MDFA，使用大量跳跃连接的DNA-Net，使用分组卷积的ISTDU，使用全局注意力机制的LPNet，使用局部对比的MLCL。

量化指标对比：

这些算法在不同的数据集上的量化指标对比如表所示。其中检测率和交并比是越大越好，而虚警率是越小越好。在这些不同算法的指标中，最好的指标用红色加粗字体突出显示。总体而言，由于深度学习算法能够自主学习到最适合的权重，泛化能力更强，所以基于深度学习算法大都比传统算法表现更好。

MPCM使用局部对比方法，对不同大小的区域计算目标和背景之间的灰度差异，因此其可以适应含有不同大小目标的场景，因此其的检测率很高。但是其对变化剧烈的背景十分敏感，当某一个区域的中心灰度值比周围的灰度值高的时候没算法就会认为该区域为目标所在位置，因此其的虚警率也很高。

IPI算法将图像使用滑动窗口分割成有重叠的图像块，之后再进行重构，并利用求解稀疏和低秩矩阵的方法进行目标检测。其可以适用背景比较复杂的场景，所以在传统算法中取得了最好的量化指标。但是由于其进行目标检测时依赖目标的稀疏性，当目标较多时，检测效果会变差，所以其在目标数量较多的数据集上的表现不如其他两个数据集。

FKRW算法使用均值和facet kernel 滤波器用来去除背景，然后使用随机游走算法分割目标和背景。其能够去除大多数的背景和干扰，因此其虚警率较低；但是在此过程中，目标信息也不可避免地被同时去除了一部分，因此其的检测率比较低。

Istdu使用分组卷积，并且对不同的组分配不同的权重，因此其可以提取具有不同特征的信息，但是其使用平均均方差作为损失函数，不能够减弱红外图像中正负样本不平衡的问题，所以其的量化指标并不是很好。DNA-Net使用大量的跳跃连接，因此可以提取到丰富的上下文信息，同时能够保留更多的细节信息。但是其没有使用注意力机制，不能从大量信息中关注到更重要的目标信息，所以其的检测率并不高。LP-Net使用全局注意力机制，同时获取局部和全局信息，可以抑制复杂的背景。但是其后复杂的后处理会导致目标的细节信息丢失，因此其的交并比不高。

Mdfa使用两个生成器将红外弱小目标检测分为降低虚警率和增减检测率两个子任务。其实现了很高的检测率，但是其两个生成器中的G1网络结构过于简单，并且最后输出为两个生成器的输出进行平均，所以最后输出容易被G1所影响，所以其的虚警率也很高。Mlcl使用局部对比结构生成局部对比特征图，并通过多尺度融合结构，其可以提取到丰富的多尺度信息。但是其内部使用了很多的膨胀卷积，会造成网格效应，容易使目标信息丢失，所以其的检测率和虚警率都很低。

得益于大尺寸的卷积核带来的大的有效感受野和fct带来的精准的目标定位，我们的算法在对比的算法中取得了最好的量化指标。此外，我们还对比了使用动态阈值的ROC曲线，其如图所示，可以看到我们的算法在动态阈值的测试条件下，性能也超过了其他的算法。

视觉对比：

三个数据集中的的一些视觉对比图片分别如图所示。其中图像中的黄色圈代表虚警，红色圈代表漏检。我们将目标区域放大放在了图像的角落处，当图像中含有多个目标或有虚警时，我们使用蓝色虚线将图像中的目标和放大后的图片对应起来。

在传统算法中IPI算法能够检测出所有的目标，但有一定程度的虚警，综合来看是传统算法中检测效果最好的。FKRW的虚假比较少，但是其有很多漏检，此外他在图像的底部边缘引入了额外的噪声，综合来看检测效果不太好。Mpcm对于噪声过于敏感，其图像中存在大量的虚假目标，并且难以分辨目标所在位置，该算法的检测效果是这些算法中最差的，因此该算法的检测结果没有展示。

Mlcl没有虚警，但是有很多目标没有检测出来；mdfa没有漏检，但是也存在很多的虚警，因此这两个算法检测效果并不好。ISTDU存在一定的漏检和虚警，而dna基本没有漏检，而且虚警也比较少，综合来看检测效果还可以。LPNet的虚警和漏检均很少，但是其后续的处理会使目标的细节信息丢失，造成图像模糊。而我们的算法能够在没有虚警的情况下，保证漏检也非常少，仅有一个目标没有检测出来，在这些对比的算法中取得了最好的检测效果。

讨论：

综合上述的量化指标对比和视觉对比来看，基于深度学习的算法的检测性能是明显优于传统算法的。传统算法依赖于人们之前总结的先验知识，当数据集中的图像不符合该先验知识时，检测效果就会大打折扣，所以传统算法的泛化能力不足，不能满足实际场景的要求。而基于深度学习的算法能够通过改变内部的权重以适应不同的数据集，因此泛化能力更强，检测效果比传统算法更好。

通过使用大尺寸的卷积层，我们的算法可以拥有大的有效感受野，可以提高模型的鲁棒性，提高检测效果。通过使用transformer结构的注意力机制，可以使算法能够更好地定位目标所在区域。相较于其它的算法，我们的算法取得了最好的量化指标，同时也取得了最好的检测结国。消融和对比实验证明了我们所提出的大尺寸卷积核组成的残差网络和transformer结构的注意力机制能够有效提高模型的检测性能。

结论：

在本文中，我们设计了一个具有大感受野的红外弱小目标检测算法LRF-Net。它使用了使用大尺寸卷积层。因此可以使网络具有更大的有效感受野，从而使网络能够提取到更多的信息。此外，他使用了transformer结构注意力机制，通过逆瓶颈结构，可以保留更多信息，提高模型的定位精度。我们所提出的LRF-Net在不同的数据集上的表现均超越了其它的先进算法，证明了我们所提出算法的先进性。