**计算机视觉报告**

**小组成员及分工**

**王文子正：算法研究、代码实现、视频录制**

**朱浩然：算法研究、PPT制作、视频录制**

**王晨曦：代码实现协助、报告编写、视频录制**

**一、介绍**

高动态范围成像（HDR）是一组用于实现比普通数位图像技术更大曝光动态范围的技术。HDR允许显示更大的亮度范围，从最暗到最亮，使成像效果更接近人眼所见。HDR通常通过拍摄多个不同曝光的画面合并成完美的图像来实现。

现有的HDR重建方法可以大致分为三种：1）基于运动检测的方法。这类方法提出在多个LDR图像中检测可信的运动区域，然后在融合中取出这些区域[1]-[4]。然而，该方法的先决条件是输入中所涉及的运动足够小。当输入中遇到大的运动时，在融合中会有大量的像素被移除，会在运动区域上造成不可避免的巨大信息损失，从而产生失真。2）基于对齐的方法。与基于运动检测的方法相比，基于对齐的方法通过估计光流来对准输入的多个LDR图像[5]-[7]。该方法可以应对一些大运动的情况，但其性能受到光流估计精度的限制。在一些情况下（如：运动区域过饱和或欠饱和），仍然会产生明显的效果。3）基于深度学习的方法。在受到深度卷积神经网络在各种任务中取得巨大成功的启发，人们提出了许多基于DCNNs的HDR重建方法，以学习从多个LDR图像到潜在HDR图像的复杂映射函数[8]。得益于DCNNs强大的表达能力。往往比上述两种方法有更好的重建效果。但是，这些方法未能显示地除去鬼影（ghosting artifacts）并充分利用输入中的有用信息。对于基于多次曝光的HDR重建来说，去除重影伪影仍然是一个具有挑战性的问题。

为了更好地解决这一问题，文章提出了一种有效地非局部高动态范围重建网络（NHDRR-net）。该方法与上述方法相比，文章中提出通过在深度特征空间中利用输入多幅LDR图像的非局部相关性来去除重影伪影。与现有的仅使用多个残差块进行特征粗融合不同[9]，该方法采用了双分支策略。一个分支是去除重影伪影和不对齐的区域，另一个分支尝试从LDR图像中收集局部细节。该模型首先采用Unet架构将所有输入集成到深度特征空间中，然后，开发一个新的全局非局部模块，并将其纳入网络以处理由此产生的深度特征。全局非局部模块由自适应池化层、非局部处理层和自适应解释层组成。在全局非局部模块中，通过加权平均所有其他像素来重构每个像素，其权重由它们的对应关系决定。通过这种方式，所提出的网络能够充分利用输入图像的优势。而且，它可以选择有用的信息，并将其调制成合适的形式进行重建。此外，为了获得更强大的深度特征表示，有利于回复HDR图像的丰富细节，文章进一步采用三通道残差块来获取输入的局部特征，然后将得到的局部特征和来自全局非局部模块的非局部特征连接起来，重建最终的HDR图像。该文章主要的突出贡献有以下四个方面：提出了一种有效的NHDRRnet，用于基于多次曝光的HDR重建；提出了一种全新的全局非局部模块，可以灵活地插入任何现代DCNN架构；提出了融合局部和全局非局部的重建融合模块；拥有最先进的性能。

**二、主要技术**

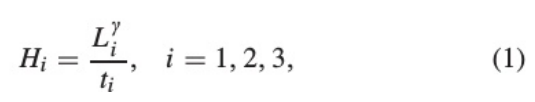
下面将对文章的研究及其相关技术进行详细的介绍。

**1.非局部网络**

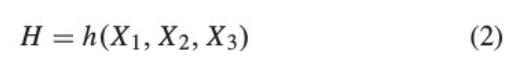
非局部操作是一种传统的滤波算法，最初用于图像去噪[10]、[11]、修复[12]、纹理合成[13]。非局部操作的工作原理是基于其他图像位置上所有像素的加权平均来计算响应位置的值。其优点是它可以建立关于空间域的长程依赖关系。在深度神经网络中，获取长时记忆至关重要。对于序列数据，递归运算是长时记忆建模的主要解决方案。对于图像数据，长时记忆建模则依靠大型感受野，后者多是多层卷积运算堆叠的结果。卷积核递归运算处理的都是局部邻域，可以是空间局部邻域，也可以是时间局部邻域，只有不断重复这些运算，逐步在数据中传播信号，才能获取长时记忆。而这会使得计算效率地下，并产生一些优化问题。为了解决上述问题，人们将非局部运算作为一个高效、简单核通用的模块，用于获取深度神经网络的长时记忆，是计算机视觉中经典的非局部均值运算的一种泛化结果。该运算将某一处位置的响应作为输入特征映射中所有位置的特征的加权核来进行计算。这些位置可以是空间位置，也可以是时间位置和时空位置，这意味着该计算适用于图像、序列和视频问题。与递归和卷积运算的渐进的操作相比，非本局运算直接通过计算任意两个位置之间的交互来获取长时记忆，可以不用管其间的距离。文章将非局部操作引入到HDR成像中，观察到非局部块可以帮助网络在不同曝光之间建立依赖关系，这对于饱和和遮挡非常有帮助。

**2.NHDRRnet架构**

该网络（如图1所示）通过一种新的Unet架构直接解决HDR成像问题。HDR成像得目标是获取一组动态场景得LDR图像作为输入，并生成与参考图像对齐的无鬼影HDR图像。在文章中所设置的训练数据集每个训练样本只含有三张LDR图像，因此将中间曝光图像作为参考图像。HDR图像的内容要么直接由LDR图像输入生成，要么在目标区域被遮挡或饱和时通过非局部策略产生幻觉细节。获取HDR图像*H1*，*H2*，*H3*是通过伽马校正。这些HDR图像通过以下方法从LDR图像中获得：



其中的*ti*表示第*i*张图像的曝光时间*Li*，*γ*是伽马校正参数。*H*表示最终的图像，而*Hi*表示*Li*的HDR图像。将Li和Hi连接成*Xi*={*Li*，*Hi*，*i* = 1,2,3}，共由6个通道。因此，过程可以写成：



其中*h*表示HDR图像与其LDR输入的复映射函数，通过使用所提出的网络在足够的训练数据的帮助下学习该函数。*Li*、*Hi*和*H*的值在0和1之间有界。NHDRRnet的输入是未经对齐预处理的原始LDR图像。

该架构可以在输入的非局部相关性的问题下，去除HDR成像中的重影伪影。由于CNN的学习能力受到架构和损失函数的正确选择的严重影响，而Unet是一个广泛用于许多任务的网络。其具有跳过连接的编码器-解码器框架。编码器层捕获不同尺度的上下文信息，底层特征获取HDR图像的细节，高层特征有利于去除重影伪影。解码器层允许网络将全局上下文信息传播到更高分辨率层。跳跃连接将编码器层的特征和相应解码器层的输入连接起来。由于上述优点，文章采用Unet作为基准。因此，如图1所示，NHDRRnet的整体框架由编码器、合并和解码器三个部分组成。

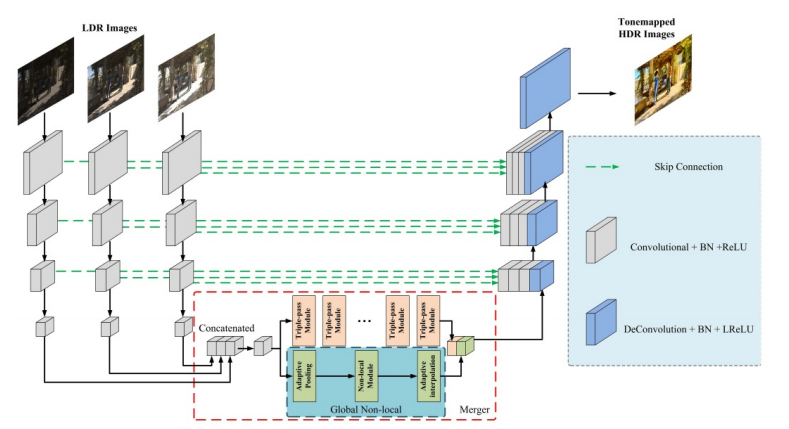


图1 NHDRRnet架构

如同上述描述，文章使用的输入是多张LDR图像。因此，可以为不同的曝光输入采用单独的编码器。这些单独的编码器从输入中学习独立的特征。而输出的特征图将通过合并块进行组合。此外，文章强制编码器学习更注重捕捉输入的互补特征的过滤器。这种方法有助于重建所需的HDR图像。文章中所提出的合并块包括两个模块：三遍残差模块和全局非局部模块。通过这两种模块来提取局部和全局信息。

**3.三遍残差模块**

三遍残差模块（如图2所示）用于输入中学习更有效的特征。因为不同的感受野包含了不同的信息，所以聚合这些信息是增强细节的更好方式。文章在三遍残差模块中使用了几种不同类型的核，以增加网络的宽度和网络对多尺度处理的适应性。小的核尺寸致力于提取局部特征，大的核尺寸可以覆盖更广泛的接受层区域。

卷积神经网络通过堆叠非常深的层来获得优异的性能，这些层主要考虑了局部感受野，并且没有显示地利用图像内部的自相似性。文章中表示，利用输入的清晰区域来重建鬼影区域使解决鬼影问题的有效方法。而这时非局部策略的关键概念。因此文章中提出采用非局部操作来去除鬼影，而全局上下文有助于去除鬼影伪影。但是因为传统的非局部网络受限于计算量，只能在有限的区域内进行计算。通过计算特征域中的非局部信息，其策略允许非局部操作有效地全局工作。先采用自适应池化在全局非局部模块中生成大小固定（32×32）的特征，然后采用非局部处理层来消除鬼影并自动产生幻觉，其中全局非局部模块的输出覆盖了整个输入的信息。之后，将特征映射上采样到三遍分支的输出大小。最终的HDR图像由几个解码器层生成。

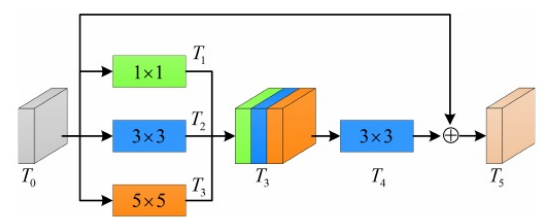
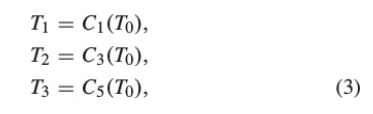
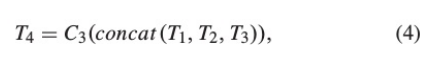


图2 三遍残差模块架构

如图2所示，三遍残差模块中有几种不同类型的核。使用了小尺寸的滤波器（1×1）、中尺寸的滤波器（3×3）和大尺寸的滤波器（5×5）。小尺寸的滤波器能够提取特征的局部细节，大尺寸的滤波器可以覆盖接受层的较大区域。所有的特征都以通道级联的方式进行拼接，并通过卷积层进行压缩。与一般残差模块不同，三遍残差模块可以嵌入不同的卷积核。图2中的*T0*表示三遍残差模块的输入。将*T0*与几个不同的和进行卷积，得到特征*T1* ,*T2*, *T3*：



其中*C1*, *C3*和*C5*分别表示核大小为1×1，3×3和5×5的卷积层。所有的卷积层后面都跟着一个ReLU激活函数。每个卷积层的通道数为256。*T1*、*T2*、*T3*的维度相同，并将它们的通道进行级联，采用3×3的卷积层来更稀疏地表示级联地特征：



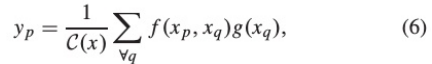
这一步有利于降低特征图的厚度。输出特征*T5*是通过学习输入特征*T0*的残差生成的：



三遍残差模块通过并行执行多个卷积来学习多个特征。这种设计的结构可以更好地捕捉图像表征。有效的多尺度特征表示使网络具有较强的认知能力，生成更高质量的HDR图像。

**4.全局非局部模块**

在非局部均值操作后，一个全局非局部操作将特征*x*作为输入，并生成与*x*大小相同的输出信号*y*。



其中*p*是要计算其响应的输出位置的索引，*q*是枚举所有可能位置的索引。函数*f*表示*xp*和*xq*之间的关系。*G*计算位置*q*出输入信号的表示形式，矩阵*C(x)*表示对响应进行归一化的因子。卷积操作是在局部区域内进行，而非局部操作使用所有信息，并为自相似度量产生可靠的特征关联。对于HDR任务，有一些区域既过饱和又被遮挡。很难通过深度卷积网络对这些区域的细节进行计算。然而，非局部模块提供了恢复退化区域的机会。

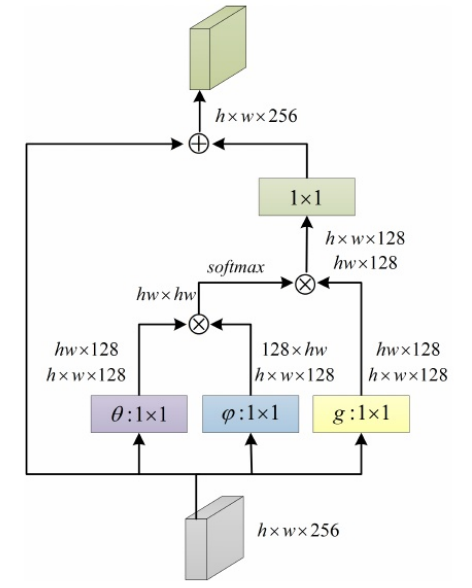


图3 全局非局部模块

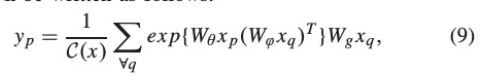
根据一般非局部形式，定义*g*的线性嵌入如下：



其中*Wg*表示一个要学习的权重矩阵，可以通过1×1卷积实现。将线性嵌入高斯核的欧式距离，作为距离度量：



其中*θ(xp) = Wθxp和φ(xq) = Wφxq*分别是两个可学习的嵌入变换。因此，非局部运算可以写成：



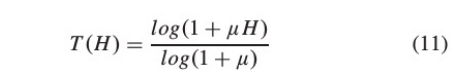
其中的*C(x) = ∀qf(xp, xq)*。为了使提出的网络覆盖更大的感受野，并对细节进行计算，使用了特征的全局信息。具体来说，为了计算*yp*，评估了整个输入特征中*xp*和每个*xq*之间的相关性。

通过添加残差连接，可以将非局部操作视为非局部模块：



其中*zp*为非本地模块的输出。如图3所示，如果*Wz*初始化为零，残差连接可以保持初始行为。

HDR图像通常是经过色调映射后显示在屏幕上的，虽然已经提出了几种鲁棒的色调映射方法，但是这些方法要么复杂，要么不可微。因此文章中使用了*μ-law*函数，其为音频处理中广泛使用的范围压缩器，同时*μ-law*是可微的。定义如下：



其中*μ* = 5000是决定压缩程度的参数，*T(H)*是HDR图像*H*的色调映射图像，其中*H*的范围为[0, 1]。在文章使用的方法中，使用MSE损失函数来最小化所提出网络的色调映射结果与地面真实HDR图像之间的距离。可被写为：



式中*T(H)*表示色调映射的真实HDR图像。

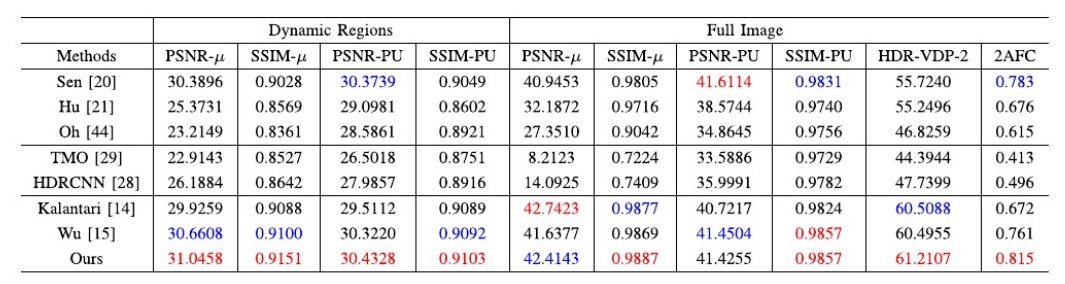
**三、实验细节**

在实验中，文章将公式(1)中的*γ*设置为2.2，*H*始终在[0, 1]范围内。NHDRRnet的输入被分别送入不同的编码器。编码层是步幅为2的四个卷积层，而解码曾是步幅为1/2的反卷积层核。为了保持特征的大小，对所有卷积采用零填充。所有编码层使用3×3核，然后是归一化和ReLU激活函数。所有解码层使用4×4核，然后是批量归一化和LeankyReLU激活函数。每层的通道数在编码期间从32增加到256，在解码期间从256减少到32。在合并块中，输入连接三个分支的特征（基于三个LDR图像），然后是一个3×3卷积核。最后的卷积层有三个通道，与输出的HDR图像的维度相匹配。文章中的方法不需要任何辐射校准或对齐。

**四、实验对比**

文章将提出的网络与几种方法在测试数据集上进行了定量比较，如表1所示：

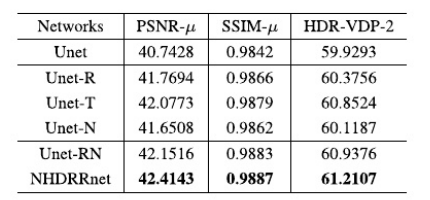
表1 比较结果



其中红色表示最佳性能，蓝色表示次优结果。

同时也通过消融实验对不同网络之间进行实验，证明方法中突出的模块的有效性。为了证明全局非局部模块是HDR成像任务的有效策略，比较了基线架构的性能。

表2 定量比较结果

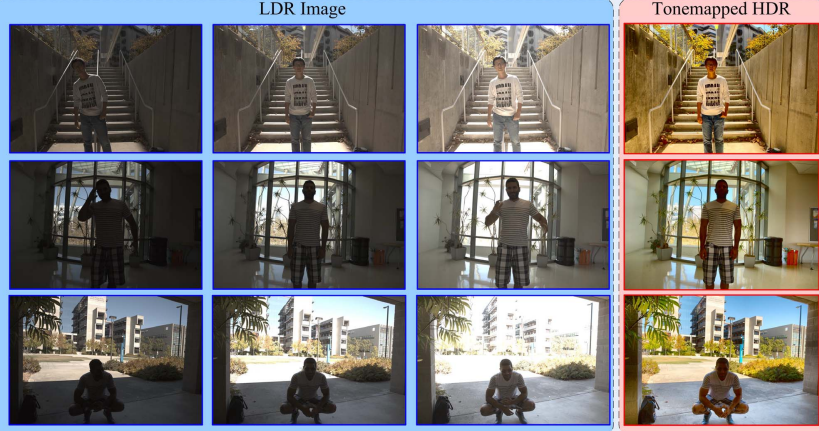


从表2中可以看出，全局非局部模块是HDR图像去重影任务的有效机制。全局非全局分支专注于提取周围信息，填充遗漏或饱和的区域。与Unet相比，Unet-N显著提升了性能。由于采用自适应池化和插值策略的Unet-N算法只能捕获全局特征，因此在处理过程中会丢失一些信息。不具有池化的Unet-R和Unet-T具有比Unet-N更高的数值结果。此外，Unet-RN和NHDRRnet分别比Unet-R和Unet-T获得更好的定量结果，也证明了全局非局部模型在整个网络中发挥着至关重要的作用。

文章中同时比较了三遍残差模块的性能。虽然已经引入了全局非局部机制，但三遍残差可以进一步提高性能。从表2中可以明显看出，Unet-T的结果与Unet-RN相当，并且优于Unet的结果，因为包含了合并过程，这有利于去除鬼影伪影，获得更好的性能。此外，使用标准残差单元证明了该模块可以恢复HDR成像的细节，并获得了满意的结果。具有三通模块的Unet-T比Unet-R获得了更好的定量结果，主要原因是不同的感受野在三遍残差模块中包含了不同的信息，因此聚合这些信息是增强细节的更好方法。

NHDRRnet中的合并包括两个分支，可以捕获全局（全局非局部模块）和局部信息（三通模块）。从表2可以看出，与单分支结构Unet-R、Unet-T和Unet-N相比，双分支结构Unet-RN和NHDRRnet在定量指标上获得了明显的性能改进。获得这些令人满意的结果有两个主要原因，一方面，全局非局部分支利用所有信息来恢复退化区域；另一方面，局部分支为最终的图像提供了丰富的细节。具有两分支合并的NHDRRnet更适合HDR成像任务。

最终结果如下图所示：



**五、感受与体会**

这次作业的难度很大，对我们小组来说是一个巨大的挑战。一开始，我们希望能够复现一篇无代码的论文，然而，经过一段时间的尝试和努力，我们发现这个目标难以实现，不得不放弃。因此，我们转而尝试理解并自行复现一篇带有代码的论文。这样的转变大大降低了我们的难度，同时也为我们提供了更多的学习机会。

通过此次实验，我们对高动态范围（HDR）这项技术有了更加深刻的理解和认知。我们不仅掌握了HDR技术的基本原理，还深入了解了其在实际应用中的实现方法和优化策略。通过阅读和分析论文中的代码，我们学习到了许多编程技巧和算法设计思想，这些都极大地提升了我们的技术水平和团队合作能力。

在此过程中，我们的团队成员互相配合、共同克服了一个又一个难题。最终，我们成功地复现了这篇论文。

**参考文献**

1. K. Jacobs, C. Loscos, and G. Ward, “Automatic high-dynamic range image generation for dynamic scenes,” IEEE Comput. Graph. Appl., vol. 28, no. 2, pp. 84–93, Mar. 2008.
2. T. Grosch, “Fast and robust high dynamic range image generation with camera and object movement,” in Proc. IEEE Int. Conf. Vis., Modeling Vis., Nov. 2006, pp. 277–284.
3. Q. Yan, J. Sun, H. Li, Y. Zhu, and Y. Zhang, “High dynamic range imaging by sparse representation,” Neurocomputing, vol. 269, pp. 160–169, Dec. 2017.
4. A.Srikantha and D. Sidibé, “Ghost detection and removal for high dynamic range images: Recent advances,” Signal Process., Image Commun., vol. 27, no. 6, pp. 650–662, Jul. 2012.
5. O. T. Tursun, A. O. Akyüz, A. Erdem, and E. Erdem, “The state of the art in HDR deghosting: A survey and evaluation,” Comput. Graph. Forum, vol. 34, no. 2, pp. 683–707, May 2015.
6. H. Zimmer, A. Bruhn, and J. Weickert, “Freehand HDR imaging of moving scenes with simultaneous resolution enhancement,” Comput. Graph. Forum, vol. 30, no. 2, pp. 405–414, Apr. 2011.
7. Q. Yan, Y. Zhu, and Y. Zhang, “Robust artifact-free high dynamic range imaging of dynamic scenes,” Multimed Tools Appl., vol. 78, no. 9, pp. 11487– 11505, May 2019.
8. N. K. Kalantari and R. Ramamoorthi, “Deep high dynamic range imaging of dynamic scenes,” ACM Trans. Graph., vol. 36, no. 4, pp. 1–12, Jul. 2017.
9. S. Wu, J. Xu, Y.-W. Tai, and C.-K. Tang, “Deep high dynamic range imaging with large foreground motions,” in Proc. Eur. Conf. Comput. Vis., Sep. 2018, pp. 117–132.

[10] H. C. Burger, C. J. Schuler, and S. Harmeling, “Image denoising with multi-layer perceptrons, part 2: Training trade-offs and analy-sis of their mechanisms,” 2012, arXiv:1211.1552. [Online]. Available: https://arxiv. org/abs/1211.1552

[11] A. Buades, B. Coll, and J. M. Morel, “A non-local algorithm for image denoising,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., vol. 2, Jul. 2005, pp. 60–65.

[12] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, “Patchmatch: A randomized correspondence algorithm for structural image editing,” ACM Trans. Graph., vol. 28, no. 3, p. 24, 2009.

[13] A. A. Efros and T. K. Leung, “Texture synthesis by non-parametric sampling,” in Proc. Int. Conf. Comput. Vis., 1999, pp. 1033–1038.