激光与光电子学进展 先进成像

基于特征融合和反馈细化的光场图像显著性检测

梁晓[1,2]，邓慧萍[1,2*]，向森[1,2]，吴谨[1,2]

[1]武汉科技大学信息科学与工程学院，湖北 武汉 430081；

[2]武汉科技大学冶金自动化与检测技术教育部工程研究中心，湖北 武汉 430081

摘要 现有的光场图像显著性检测算法不能有效地衡量聚焦度信息，从而影响了检测目标的完整性，造成信息的冗余和边缘模糊。考虑到焦堆栈不同的图像及全聚焦图像对于显著性预测发挥着不同的作用，提出有效通道注意力(ECA)网络和卷积长短期记忆模型(ConvLSTM)网络组成特征融合模块，在不降低维度的情况下自适应地融合焦堆栈图像和全聚焦图像的特征；然后由交互特征模块(CFM)组成的反馈网络细化信息，消除特征融合之后产生的冗余信息；最后利用ECA网络加权高层特征，更好地突出显著性区域，从而获得更加精确的显著图。所提网络在最新的数据集中，F-measure和平均绝对误差(MAE)分别为0.871和0.049，表现均优于现有的红、绿、蓝(RGB)图像、红、绿、蓝和深度(RGB-D)图像以及光场图像的显著性检测算法。实验结果表明，提出网络可以有效分离焦堆栈图像的前景区域和背景区域，获得较为准确的显著图。

关键词 图像处理；显著性检测；深度学习；光场图像；卷积神经网络

中图分类号 TP391 **文献标志码** A **DOI**：10.3788/LOP202259.2210006

Saliency Detection of Light Field Image Based on Feature Fusion and Feedback Refinement

Liang Xiao[1,2], Deng Huiping[1,2*], Xiang Sen[1,2], Wu Jin[1,2]

[1]*School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan* 430081, *Hubei, China*;

[2]*Engineering Research Center for Metallurgical Automation and Measurement Technology of Ministry of Education, Wuhan University of Science and Technology, Wuhan* 430081, *Hubei, China*

Abstract The existing light field image saliency detection algorithms cannot effectively measure the focus information, resulting in an incomplete salient object, information redundancy, and blurred edges. Considering that different slices of the focal stack and the all-focus image play different roles in saliency prediction, this study combines the efficient channel attention (ECA) network and convolutional long short-term memory model (ConvLSTM) network to form a feature fusion network that adaptively fuse the features of the focal stack slices and all-focus images without reducing the dimension; then the feedback network composed of the cross feature module refines the information and eliminates the redundant information generated after the feature fusion; finally, the ECA network is used for weighing the high-level features to better highlight the saliency area to obtain a more accurate saliency map. The network proposed has F-measure and mean absolute error (MAE) of 0.871 and 0.049, respectively, in the most recent data set, which are significantly better than the existing red, green, and blue (RGB) images, red, green, blue, and depth (RGB-D) images, and light field images saliency detection algorithms. The experimental results show that the proposed network can effectively separate the foreground and background regions of the focal stack slices and produce a more accurate saliency map.

Key words image processing; saliency detection; deep learning; light field image; convolutional neural network

收稿日期：2021-07-28；**修回日期**：2021-08-26；**录用日期**：2021-10-13

基金项目：国家自然科学基金(61702384,61502357)

通信作者：*denghuiping@wust.edu.cn

1 引 言

显著性检测是从图像中获取最感兴趣和最吸引人眼球的区域或目标[1]。近年来，人们用机器学习的方法实现了显著性检测，并把它应用到各个领域中，例如图像检索[2]、图像语义分割[3]和图像压缩[4]等。传统的显著性检测算法[5-10]根据颜色、纹理和亮度等线索，由背景先验、前景先验和对比度先验等检测出显著图，但是这种基于人工特征提取的方法无法获取全局上下文语义信息，并且太过于依赖低层特征，无法取得很好的效果。

基于深度学习的显著性检测算法是把图像输入到卷积神经网络中，通过一系列的卷积操作和激活函数，自主地提取图像的各阶特征，从而得到显著图。这使得预测结果既获取了高层语义信息又有低层细节信息，在性能上取得了较大的提升。根据输入数据的不同，可分为红、绿、蓝(RGB)图像[11-13]、红、绿、蓝和深度(RGB-D)图像[14-15]以及光场图像的显著性检测。基于RGB图像的深度学习显著性检测算法采用端到端的方法，由于只在RGB图像中提取特征，包含的信息有限，因而在很多复杂场景如前景和背景颜色相似/纹理相似或背景杂乱的情况下，往往会出现背景难抑制、检测对象不完整和边缘模糊等问题。基于RGB-D图像的深度学习显著性检测算法同样采用端到端的方法，加入了深度图作为深度信息并将其补充到显著性检测线索中。该算法虽然在精度上得到了提升，但显著性检测的精度太过依赖于深度图的质量，且深度图的质量又受到技术条件的限制，当深度图的质量较差时，显著性检测的精度就会降低。

光场图像包含光线的位置和方向，并提供了更高维度和更灵活的场景信息，使得光场图像的显著性检测结果更加准确。现有的基于深度学习的光场图像显著性检测可以根据输入的不同分为两类：1)子孔径多视点图像；2)焦堆栈图像和全聚焦图像。Piao等[16]提出了基于卷积神经网络(CNN)的多视点图像显著性检测网络(DLFD)，以光场的多视点数据作为中间桥梁，将单张中心视点图进行光场合成并恢复水平方向和垂直方向的视图，再通过卷积网络对所得视图进行显著性检测，根据深度信息将多视角显著图变换到中央视角下，通过注意力网络加权，进而得到最终的显著图。Zhang等[17]提出角度变化模块(MAC)算法，以多个视点的微透镜图像作为网络的输入，把通过视觉几何组(VGG-19)网络得到的特征图输入到改进后的空间金字塔池化(ASPP)模型中，获得图像的多尺度信息和空间信息，最后经过融合得到预测显著图。用子孔径多视点图像作为输入，只是利用视角特性和位置信息，无法准确地区分目标物体和背景，容易造成边缘模糊。Wang等[18]提出了以焦堆栈图像和全聚焦图像作为输入的算法，并新建了一个大规模的光场图像显著性检测数据库。该算法采用两个网络分支，一个分支用注意力网络和卷积长短期记忆模型(ConvLSTM)网络组成的循环注意力机制自适应地融合焦堆栈图像的特征，另一个分支用来提取全聚焦图像的特征，最后将两个分支的显著图进行融合得到显著图。Zhang等[19]设计了一个有记忆力的光场存储器定向解码器(MOLF)算法，深入探索焦堆栈图像的内部相关性。该算法对焦堆栈图像和全聚焦图像进行特征提取，再用注意力网络和ConvLSTM网络组成的循环注意力机制将丰富的聚焦信息和空间信息融合，并使用ConvLSTM网络逐步细化空间信息，进而得到显著图。Zhang等[20]提出光场融合网络(LF-Net)算法，通过残差的方法充分提取焦堆栈图像和全聚焦图像的高层语义信息，再经过细化模块以递归的方式提炼光场特征，最后输入到由注意力网络和ConvLSTM网络组成的光场整合模块中，检测出显著性目标。Piao等[21]提出非对称输入和非对称结构的双流网络(ER-Net)算法，用学生网络学习全聚焦图像特征，用教师网络学习焦堆栈图像的特征，同时将所有的聚焦度信息转移到学生网络中进行融合，使焦堆栈图像和全聚焦图像之间的信息互补，从而得到显著图。

现有的以焦堆栈图像和全聚焦图像为输入的光场显著性检测模型大多采用注意力机制，赋予焦堆栈图像从0到1的不同权重，使聚焦在显著性物体上的图像赋予大的权重，聚焦在背景区域的图像赋予较小的权重，从而更有效地关注聚焦在显著性物体上的焦堆栈图像，并使用ConvLSTM网络逐步精细化空间信息，使特征充分融合。但注意力机制多采用降维方式会导致特征通道与权重之间的对应是间接的，会引起信息的丢失，不能有效地对聚焦度信息进行衡量，从而导致显著性检测目标不完整；此外，特征融合之后产生的信息冗余和低层噪声，未进行自下而上/自上而下的后处理，使得预测显著图的边界模糊。

在复杂场景中，如何利用焦堆栈图像的聚焦度信息有效分离光场图像的前景区域和背景区域是提高显著性检测精度的核心问题。受上述问题的引导和启发，本文提出了基于特征融合和反馈细化的光场图像显著性检测网络。该网络以光场图像的焦堆栈图像和全聚焦图像作为输入，采用不降维的有效通道注意力(ECA)[22]网络与ConvLSTM网络组成特征融合模块，使聚焦在不同焦堆栈图像上的不同信息充分融合，以强调有用的特征，抑制不必要的特征；以Wei等[13]提出的交互特征模块(CFM)为基础，组成自上而下和自下而上的反馈细化模块，消除特征冗余和差异，从而得到显著图，进一步提高显著性检测的性能。本文提出的框架有以下优点：

1) 采用ECA网络在不降低维度的情况下去赋予焦堆栈图像不同的权重，从而可以更有效地关注聚焦在显著性物体上的聚焦图像，并用较少的参数给光场图像显著性检测带来了较大的性能提升；通过

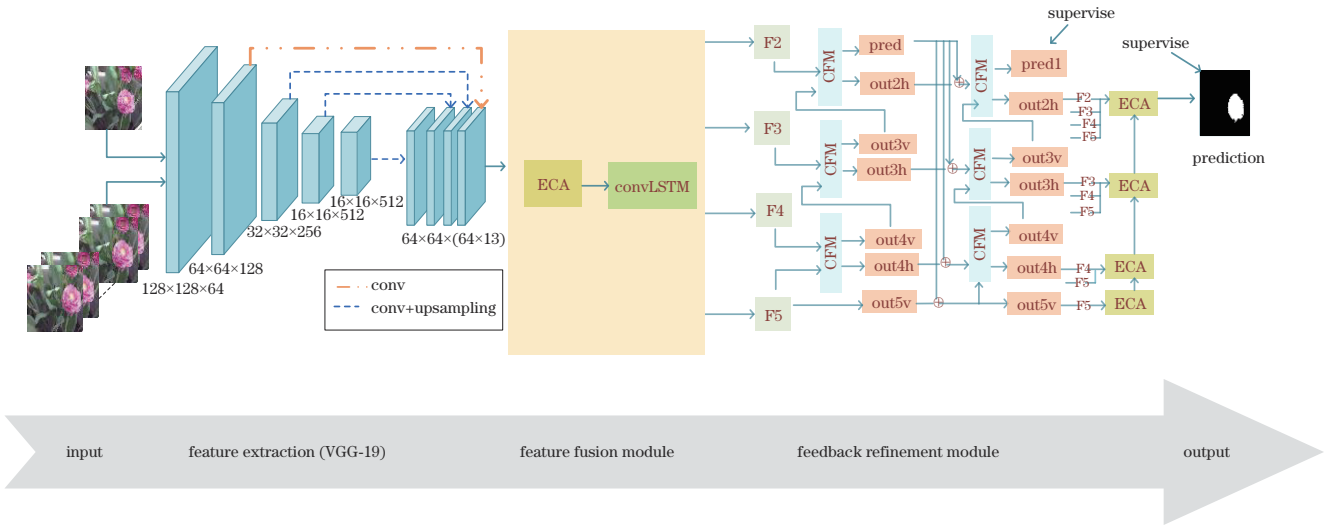
ConvLSTM 网络的循环，让信息交互，使聚焦在不同图像上的信息充分融合，突出聚焦在显著性物体上的聚焦图像，提取更加有效的特征表达，提高了显著性检测的准确性。

2）用 CFM 组成一个自上而下和自下而上的反馈细化网络，迭代地校正和细化融合之后的冗余信息，并对特征进行补充，消除特征之间的差异，再通过 ECA 网络校正，更加突出显著性，进而得到较精确的显著图。

2　光场显著性目标检测算法

2.1　整体网络构架

本文提出了一种基于特征融合和反馈细化的光场图像显著性检测算法。该算法的整体网络框架如图 1 所示，先对焦堆栈图像和全聚焦图像进行特征提取，然后利用 ECA 网络和 ConvLSTM 网络组成的光场特征融合模块自适应地让聚焦在不同图像上的信息充分融合，最后经过反馈细化模块有效地解决融合之后的信息冗余，得到更加精确的显著图。



ECA: efficient channel attention; conv: convolution; ConvLSTM: convolution long-short term memory; CFM: cross feature module

图 1　提出网络的整体框架

Fig. 1　Overall architecture of proposed network

2.2　特征提取模块

本文以 VGG-19 网络为主干架构，删除最后一个池化层和全连接层，并保留 5 个卷积层以更好地适用于显著性检测。将一张全聚焦图像和 12 张焦堆栈图像分别输入到 VGG-19 网络的 5 个卷积层中，并提取不同层次的特征。由于 VGG-19 网络的第 1 层卷积太浅，提取不到有用的信息且容易造成信息冗余，不能进行可靠的预测。因此只保留后 4 层的输出，分别用 $f_{\frac{0}{m}}(m=2,3,4,5)$ 表示全聚焦图像支路的每个卷积层输出；用 $\left[f_{\frac{i}{m}},(m=2,3,4,5)\right]_{i=1}^{12}$ 表示焦堆栈图像支路的每个卷积层输出，其中 m 表示第 m 个卷积特征层。各卷积层输出图像的空间分辨率分别为原始图像的 1/2、1/4、1/8 和 1/16。为了减少不同尺度间的语义信息和细节信息的干扰，采用卷积和采样将焦堆栈图像提取的特征与全聚焦图像提取的特征降维到 64 个通道，并进行层间级联，得到既有高层语义信息又有低层细节信息的特征，输出的特征图为 64×64×(64×13)。特征提取模块的输出细节如表 1 所示（表中 $k\times k$-n 中 $k\times k$ 表示卷积层的尺寸，n 表示通道个数，S 表示步长）。

2.3　特征融合模块

人们通常对感兴趣的对象进行聚焦拍摄，而光场具有独特的重聚焦能力，可以生成聚焦在不同深度层的焦堆栈图像。处于聚焦深度的区域比其他深度的区域相对清晰，聚焦度值相对较大，更可能为显著区域，而聚焦深度较远的区域大多属于背景区域。由于注意力机制具有强大的选择特征能力，非常适合光场图像显著性检测。因此本文通过注意力机制自适应地整合特征以准确地定位和识别显著性物体，让聚焦在显著性物体上的图像赋予大的权重，聚焦在背景区域的图像赋予较小的权重，从而可以更有效地关注聚焦在显著性物体上的聚焦图像，强调有用的特征，抑制不必要的特征。

与以往的压缩和激活的通道注意力网络（SENet）[23] 不同，本文采用一个如图 2 所示的局部跨信道交互 ECA 网络，在不降低维度的情况下进行通道间的交互，提高了注意力机制的学习能力，有效实现了给聚焦在显著性物体上的图像提供大的权重，给聚焦在背景区域物体上的图像提供较小权重的目的，且降低了模型的复杂度。

把 VGG-19 网络输出的 13 张特征图通过级联变

表 1　特征提取模块的网络参数

Table 1　Network parameters of feature extraction module

VGG-19					Dimensionality reduction				
Layer	$k\times k-n$	Input size	Output size	S	Layer	$k\times k-n$	Input size	Output size	S
Conv1	3×3−64	256×256×3	256×256×64	1	Conv2	3×3−64	64×64×128	64×64×64	1
Maxpool	2×2	256×256×64	128×128×64	2	Conv3	3×3−64	32×32×256	32×32×64	1
Conv2	3×3−128	128×128×64	128×128×128	2	Conv4	3×3−64	16×16×512	16×16×64	1
Maxpool	2×2	128×128×128	64×64×128	2	Conv5	3×3−64	16×16×512	16×16×64	1
Conv3	3×3−256	64×64×128	64×64×256	1	Conv2		64×64×64	64×64×64	
Maxpool	2×2	64×64×256	32×32×256	2	Conv3	Upsampling	32×32×64	64×64×64	1
Conv4	3×3−512	32×32×256	32×32×512	1	Conv4	Upsampling	16×16×64	64×64×64	1
Maxpool	2×2	32×32×512	16×16×512	2	Conv5	Upsampling	16×16×64	64×64×64	1
Conv5	3×3−512	16×16×256	16×16×512	1	Conv2-Conv5	Concat	64×64×64	64×64×(64×13)	

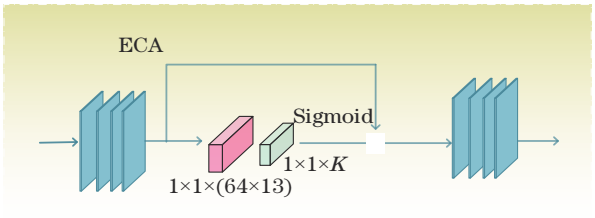


图 2　ECA 网络模块

Fig. 2　ECA network module

成 64×64×(64×13)的特征图，在不降低通道维数的情况下，经过 ECA 网络的全局平均池化得到 1×1×(64×13)的通道特征图，再通过一个 1×1×K 的卷积得到特征图像所对应的权重，最后使用 Sigmoid 函数进行归一化和激活操作，得到加权的特征图 $\bar{f}_i$。其中 K 为卷积核大小，表示局部跨通道间的覆盖范围，即有多少邻域参与一个通道的注意力预测，K 的大小可通过一个与特征图通道个数 C 相关的函数自适应确定，K 的定义为

$$K=\Phi(C)=\left|\frac{\log_2(C)}{2}\right|_{\text{odd}}, \tag{1}$$

式中，Φ 表示 K 和通道 C 之间的映射关系，K 的取值为 $\left|\frac{\log_2(C)}{2}\right|$ 的相邻最近奇数。

ECA 网络可以学习到焦堆栈图像里每个聚焦图像的特征权重，但无法做到信息交互，导致网络获取显著信息不充分。如何有效地将信息融合一直是光场图像显著性检测的一项难题，如果只是简单地将这些信息串联起来，可能会破坏信息之间的关联性，增加噪声的产生。因此把加权后的特征图 $\bar{f}_i$ 传入到 ConvLSTM 网络中去学习特征内部各个信息之间的空间依赖关系，让信息充分融合。把加权过后的 13 张特征图 $\bar{f}_{\frac{i}{t}}$ 和 ConvLSTM 网络的上一层的隐藏状态 h_{t-1} 作为输入，用长短期记忆神经网络(LSTM)堆叠进行感知，得到输出的结果 h_t 为

$$h_t=O_t\circ\tanh(C_t), \tag{2}$$

式中，C_t 表示在 t 次 ConvLSTM 网络循环时的输出，用来存储历史信息：

$$C_t=\bar{f}_{\frac{i}{t}}\circ C_{t-1}+i_t\circ\tan h\left(W_{xC}*\bar{f}_{\frac{i}{t}}+W_{hC}*h_{t-1}+b_C\right), \tag{3}$$

式中，O_t 为门状态，表示抛弃不必要的冗余信息。

$$O_t=\sigma\left(W_{xO}*\bar{f}_{\frac{i}{t}}+W_{hO}*h_{t-1}+W_{CO}\circ C_t+b_O\right), \tag{4}$$

式中，i_t、f_t 为门状态，它们分别表示保留多少有用的输入信息，输出多少有用的信息：

$$i_t=\sigma\left(W_{xi}*\bar{f}_{\frac{i}{t}}+W_{hi}*h_{t-1}+W_{Ci}\circ C_{t-1}+b_i\right), \tag{5}$$

$$f_t=\sigma\left(W_{xf}*\bar{f}_{\frac{i}{t}}+W_{hf}*h_{t-1}+W_{Cf}\circ C_{t-1}+b_f\right), \tag{6}$$

式中：t 代表 ConvLSTM 网络的循环步骤数，$t=\{1,2,3,\cdots,13\}$；$\bar{f}_{\frac{i}{t}}$ 表示在 ConvLSTM 网络循环到 t 步骤时第 i 张特征图的 ConvLSTM 网络的输入；所有的 b 和 W 表示模型需要学习的参数，其中 b 表示卷积层的偏置，W 表示焦堆栈对应的权重；*表示卷积操作；∘表示矩阵逐像素相乘；σ 表示 Sigmoid 激活函数；tan$h$ 的 h 表示隐藏层信息；W_{xC} 的下标 xC 表示存储的输入信息；W_{hC} 的下标 hC 表示存储的隐藏层信息；W_{xO} 的下标 xO 表示抛弃的输入信息；W_{hO} 的下标 hO 表示抛弃的隐藏层信息；W_{CO} 的下标 CO 表示抛弃的历史存储信息；W_{xi} 的下标 xi 表示保留的有用的输入信息；W_{hi} 的下标 hi 表示保留的有用的隐藏层信息；W_{Ci} 的下标 Ci 表示保留的有用的历史存储信息；W_{xf} 的下标 xf 表示输出的有用的输入信息；W_{hf} 的下标 hf 表示输出的有用的隐藏层信息；W_{Cf} 的下标 Cf 表示输出的有用的历史存储信息。

通过 ConvLSTM 网络的循环，对不同层的特征进行整合实现信息交互，更好地将全局上下文感知信息的高层特征保留并赋予大的权重，且被用来指导低层

局部空间细节，逐步完善其空间信息。利用特征融合模块深入探索焦堆栈图像和全聚焦图像的内部相关性与差异性，突出聚焦在显著性物体上的聚焦图像，使高层语义信息和低层细节信息充分融合，避免有效信息的丢失。

2.4 反馈细化模块

特征融合之后会产生来自低层的噪声和高层粗糙边界的冗余信息，这些信息可能会破坏原始特征导致显著性检测结果不精确。本文以 Wei 等[13]提出的如图 3 所示的 CFM 为基础，用 CFM 组成的反馈细化网络去提取高层和低层的信息，从而迭代地校正和细化特征，消除特征之间的差异，有效避免因引入过多的冗余信息而破坏原有特征，造成边界的模糊。

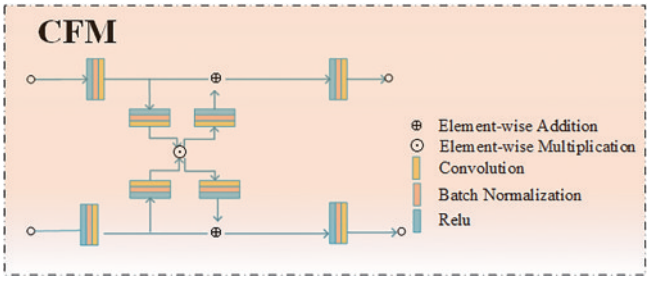


图 3 CFM 网络模块[13]

Fig. 3 CFM network module[13]

如图 3 所示，以高层和低层信息作为 CFM 的输入，用高层信息指导低层信息并作为互补信息，抑制融合之后的冗余信息，避免特征间的污染从而得到精确的高层和低层特征。与单独的特征相加或相乘不同，CFM 用加法和乘法把相邻两层特征结合到一起，先对低层和高层特征分别做两次卷积、正则化和激活操作，然后两者进行相乘，再分别与做了一次卷积、正则化和激活操作的原特征图相加，最后通过卷积、正则化、激活操作有效的细化特征和锐化边界，去除背景噪声。

由于不同层次的特征经过下采样，冗余的部分信息和噪声可能经过一次 CFM 的细化无法到达最优，影响最终的性能。因此本文做了一个如图 1 所示的反馈体系去细化和校正，分为自上而下和自下而上。对于自上而下的流程，通过 CFM 将特征从高层逐步聚集到低层，由聚集的特征反馈给下一个 CFM 的高层特征，让高层指导低层并进行细化和校正；对于自下而上的流程，由上一个 CFM 输出的特征直接输入到下一个 CFM 中进行再次细化和校正。

在卷积过程中，深层特征包含更多的高层语义信息，而浅层特征则可以保留丰富的细节信息，由于 CFM 组成的反馈体系使用大量重复的卷积操作会大大降低预测结果的分辨率，容易导致显著目标的边界模糊。为了抑制无用信息同时准确地定位显著目标，并且获取更加清晰的边界，最后使用 ECA 网络在不降维的情况下对各层的上下文信息进行逐步编码，自适应地整合特征以准确地定位和识别显著性物体，提取更加有效的特征表达，进而得到较精确的显著图。

3 实验及结果分析

3.1 实验设置

为了评估网络的性能，本文在两个公共的光场图像数据集 LFSD[5]和 DUT-LF[18]中进行了实验。其中 LFSD 数据集是由 Lytro 相机捕获的 100 个光场样本组成。DUT-LF 数据集包含 1462 个样本，1000 个样本用于训练，其余 462 个样本用于测试。由于 DUT-LF 数据集中包含了许多挑战性场景，如显著物体较大、背景杂乱、前景与背景颜色相似等，因此本文选择 DUT-LF 数据集中的训练集来训练所提出的模型，用 DUT-LF 和 LFSD 数据集中的的测试集来测试本文的网络。本文的 VGG-19 网络在 ImageNet[24]数据集上进行预训练以初始化参数，其余模型参数进行

随机初始化。本文提出的网络是在 Pytorch0.4 框架上实现的，在配置为 GeForce RTX 1080-Ti GPU 的计算机上执行。本文采用了一些数据增强操作：翻转、裁剪和旋转等，把训练图像扩展为原始的 11 倍。具体来说，使用了水平翻转和垂直翻转；分别从上、下、左、右和中裁剪部分图像；使图像分别旋转 90°、180°和 270°。整个网络训练采用端到端的方式，损失函数选用 Softmax 交叉熵损失函数，使用随机梯度下降(SGD)优化。权重衰减、动量、学习率分别设置为 0.0005、0.99、1×10^{-10}。本文所有训练图像和测试图像的大小均为 256 pixel×256 pixcel×3 pixel，batch size 设为 1。该网络训练 9 天，在 15 个 epoch 后收敛，每个 epoch 的时间为 3.5 h。

为了验证本文所提方法的有效性和先进性，将本文的实验结果与最新的传统显著性检测方法 LFS 算法[6]、RDFD 算法[9]、FPM 算法[10]；基于深度学习的 RGB 图像显著性检测方法 MWS 算法[11]；基于深度学习的 RGB-D 图像显著性检测方法 S2MA 算法[14]和基于深度学习的光场显著性检测方法 DLSD 算法[16]、MAC 算法[17]、MOLF 算法[19]、LF-Net 算法[20]和 ER-Net 算法[21]进行了全面的比较。为了公平比较，本文使用作者直接公布的显著性结果或者使用他们公开的参数和代码进行了测试。

3.2 定量评价与分析

本文采用 5 种评估指标，其中包括广泛使用的精确召回率(PR)曲线来评估显著性算法的性能；F-measure 衡量整个网络的总体性能；平均绝对误差(MAE)评估手工标注的真实显著性图与预测结果图的相近程度；S-measure 评估捕获图像空间级的结构相似性；E-measure 通过同时考虑局部像素和全局整体评估预测结果和真值图的相似性。

在 DUT-LF 和 LFSD 两个数据集中，本文方法和

其他方法对比的PR曲线，如图4所示。基于深度学习的显著性检测算法在两个数据集上基本都优于传统的显著性检测算法；基于深度学习的光场图像显著性检测算法基本都优于基于深度学习的RGB图像和RGB-D图像的显著性检测算法。在DUT-LF数据集中，本文提出方法的最小召回率高于其他方法，在LFSD数据集中所提算法的最小召回率为次优。因为LFSD数据集图像的分辨率为360 pixel×360 pixel×3 pixel，而本文网络输入图像的分辨率为256 pixel×256 pixel×3 pixel，将图像下采样之后图像分辨率会减小，图像所包含的特征信息也会相对减小；此外本文网络输入的焦堆栈图像为12张，而LFSD数据集为每个场景提供的焦堆栈图像为1～12张不等。因此将下采样之后的LFSD数据集的图片输入到本文网络中进行测试时，PR曲线效果没有达到最优。

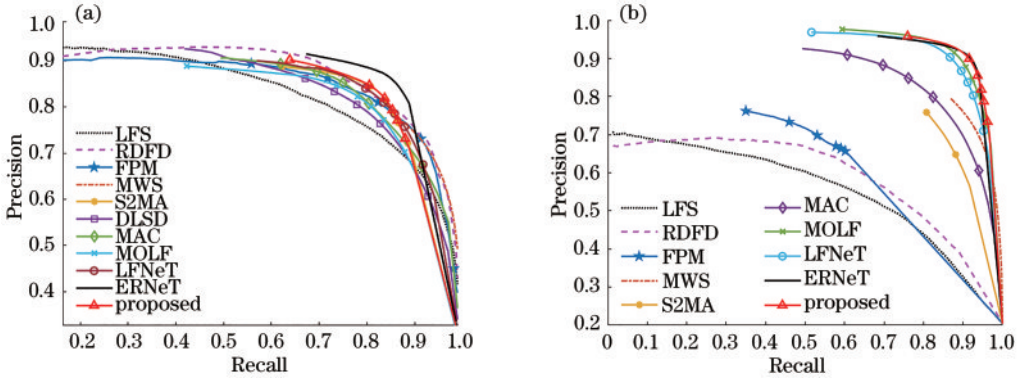


图4　不同算法在(a)LFSD数据集和(b)DUT-LF数据集中的PR曲线结果对比

Fig. 4　Comparison of PR curve results of differrent algorithms in (a) LFSD data set and (b) DUT-LF data set

在DUT-LF和LFSD两个数据集中，本文提出方法和其他方法在F-measure，MAE、S-measure和E-measure 4个指标上的对比结果，如表2所示（表中黑体为最优，斜体为次优）。从表中可以看出，所提算法在两个数据集中都有很好的结果。在DUT-LF数据集中，所提算法的F-measure、MAE、S-measure 3个指标都是次优，仅次于ER-Net算法。在LFSD数据集中，所提算法的S-measure指标最优，MAE指标和MOLF算法一样为次优。由此可以看出，MOLF算法和LF-Net算法虽然采用了注意力机制和ConvLSTM网络，但没有进行自上而下的细化处理，难以获得精度更高的显著图，而所提算法进行了自上而下和自下而上的反馈细化网络，从而获得了精度更高的显著图。

表2　不同算法在DUT-LF数据集和LFSD数据集中的指标结果对比

Table 2　Comparison of index results of different algorithms in DUT-LF data set and LFSD data set

	DUT-LF data set				LFSD data set			
Algorithm	F-measure	MAE	S-measure	E-measure	F-measure	MAE	S-measure	E-measure
LFS	0.533	0.227	0.585	0.742	0.735	0.205	0.681	0.773
RDFD	0.599	0.191	0.658	0.774	0.802	0.136	0.786	0.834
FPM	0.619	0.142	0.675	0.745	0.800	0.134	0.791	0.839
MWS	0.742	0.132	0.702	0.781	0.788	0.132	0.809	0.781
S2MA	0.753	0.102	0.787	0.816	0.803	0.094	0.837	0.863
DLSD	0.684	0.087	0.786	0.839	0.779	0.117	0.786	0.852
MAC	0.717	0.092	0.752	0.789	0.793	0.118	0.789	0.839
DLFS	0.868	0.070	0.852	0.905	0.715	0.147	0.737	0.806
MOLF	0.843	0.052	0.887	*0.923*	*0.819*	**0.088**	*0.886*	0.831
LF-Net	0.833	0.055	0.878	0.913	0.805	0.092	0.820	*0.882*
ER-Net	**0.903**	*0.040*	**0.898**	**0.946**	**0.825**	*0.085*	0.822	**0.885**
Proposed	*0.871*	**0.049**	*0.890*	0.913	0.812	**0.088**	**0.889**	0.843

3.3　鲁棒性实验

在本小节中，在DUT-LF和LFSD两个测试数据集中添加均值为0、方差为0.01的高斯噪声来验证本文提出方法的鲁棒性，定量比较结果如表3所示。

从表3可以看出，此时在加噪声的DUT-LF数据集中，MOLF算法的MAE增加了0.007，F-measure减

表 3　不同算法的鲁棒性结果对比

Table 3　Comparison of robustness results of different algorithms

Algorithm	DUT-LF data set		LFSD data set	
	MAE	F-measure	MAE	F-measure
MOLF	0.059	0.801	0.088	0.786
ER-Net	0.048	0.870	0.089	0.805
Proposed	**0.049**	**0.870**	**0.086**	**0.810**

少了 0.042；ER-Net 算法的 MAE 增加了 0.008，F-measure 减少了 0.033；而本文所提算法的 MAE 增加了 0.0003，F-measure 减少了 0.001。实验结果表明，在有噪声干扰时，MOLF 算法和 ER-Net 算法的显著性检测性能会存在一定的下降，而所提算法更稳定，噪声的加入对其显著性检测性能产生的影响极其微弱，也表明在同等条件下所提算法可以提高光场显著性检测的精度，具有更好的鲁棒性。

3.4　定性评价与分析

在 DUT-LF 和 LFSD 两个数据集中，不同显著性检测算法的视觉效果如图 5 和图 6 所示。从两幅图中可以直观看出，传统的光场图像显著性检测算法不能很好地抑制背景，甚至对显著目标的检测都不完整，而基于深度学习的显著性检测算法要优于传统的显著性检测算法；基于深度学习的光场图像显著性检测算法基本都优于基于深度学习的 RGB 图像和 RGB-D 图像的显著性检测算法。

如图 5 所示，本文选取 8 幅测试图片来做比较，第 1 行和第 2 行为显著物体较大的测试图，所提算法可以有效地抑制背景，更加突出显著对象且得到较为精确的显著物体的边缘，预测图也更接近于真实显著图；第 3 行到第 5 行为背景杂乱的测试图，所提算法均能准确地突出显著对象，而且对小细节处理得更好，第 6 行到第 8 行为背景和前景相似的测试图，当相似程度较小时，其他的网络还可以检测出显著物体，但是当前景和背景基本一致时，其他的网络均不能有效地抑制背景和提取显著目标，而所提算法可以有效地抑制背景，检测出显著目标，得到基本完整且边界清晰的显著图。

不同算法在 LFSD 数据集上的显著图如图 6 所示，图中展示了 8 张复杂的自然场景图像。实验结果表明，所提算法能够有效地利用特征融合模块找到正确的显著物体，较好地检测出显著对象，有效地抑制非显著区域，同时利用反馈细化模块优化边界模糊问题，获得边界清晰的显著图，从而进一步证明该算法的有效性。

从图 5 和图 6 可以看出，虽然所提算法在客观指标上为次优，但在主观测试结果上，尤其是在复杂场景中，如显著物体较大、背景杂乱、前景与背景纹理相似等，所提算法均优于其他算法，同时可以准确地突出显著对象，抑制非显著区域，获得边界清晰的显著图。

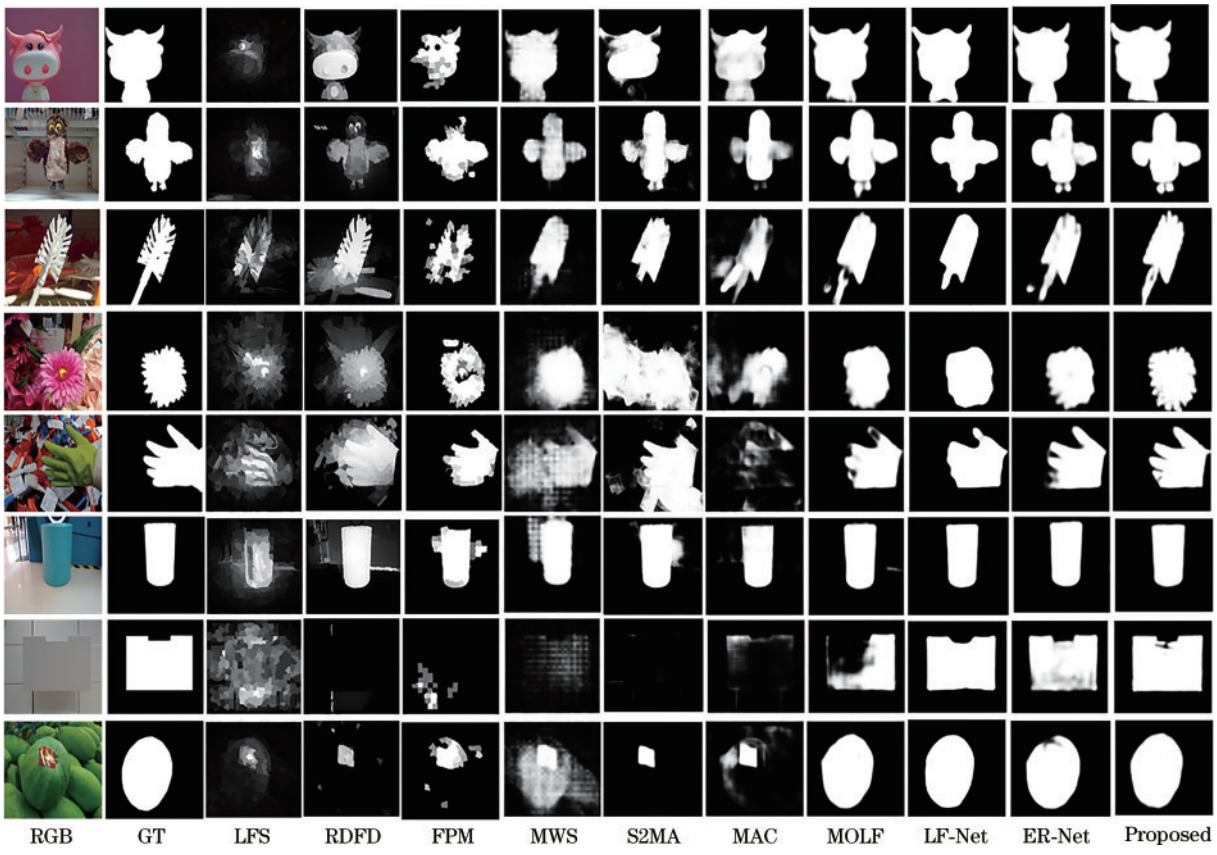


图 5　不同算法在 DUT-LF 数据集中的视觉结果对比

Fig. 5　Comparison of visual results of different algorithms in DUT-LF data set

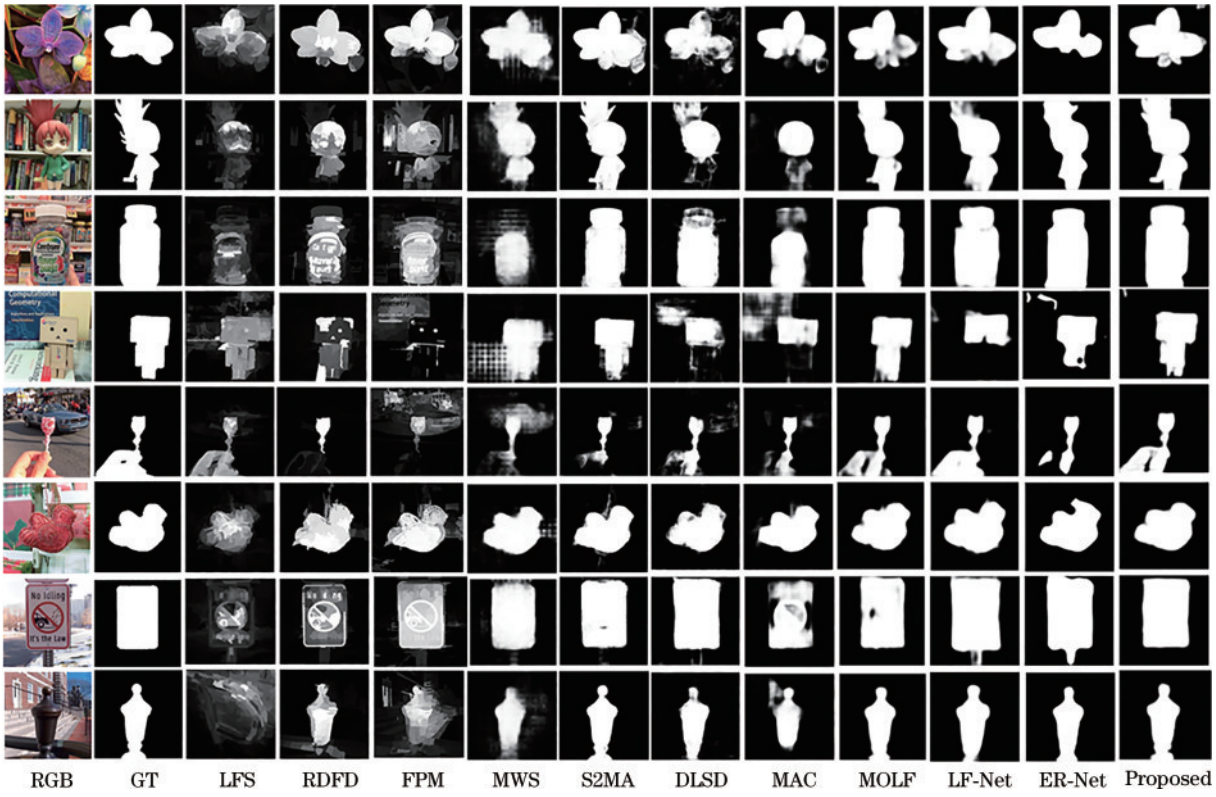


图 6　不同算法在 LFSD 数据集中的视觉结果对比

Fig. 6　Comparison of visual results of different algorithms in LFSD data set

3.5　消融实验

在本小节中，将通过多个消融实验来验证本文提出的特征融合模块和反馈细化模块在光场图像显著性目标检测过程中所起到的作用和对最终结果的贡献程度，定量比较结果如表 4 所示。

表 4　不同模块的消融实验

Table 4　Ablation experiments of different modules

Network structure	DUT-LF data set		LFSD data set	
	MAE	F-measure	MAE	F-measure
Baseline	0.110	0.822	0.132	0.753
＋SE and ConvLSTM	0.072	0.850	0.100	0.771
＋ECA and ConvLSTM	0.061	0.863	0.094	0.798
＋ECA and ConvLSTM ＋Feedback refinement module	**0.049**	**0.871**	**0.088**	**0.812**

首先，本文利用特征提取模块组成一个编解码网络对全聚焦图像和焦堆栈图像进行预测，将这种网络定义为“基本网络”，此时 DUT-LF 数据集中 MAE 和 F-measure 分别为 0.110 和 0.822。将 SE 注意力和 ConvLSTM 网络添加到基本模型中，得到 MAE 的下降、F-measure 的上升。其次，把 ECA 网络和 ConvLSTM 网络添加到基本模型中，可以看出 ECA 网络比 SE 网络效果更好。最后，在将 ECA 网络和 ConvLSTM 网络加入到基本网络的基础上，添加反馈细化网络并获得了最佳效果，MAE 和 F-measure 分别为 0.049 和 0.871。

4　结　　论

充分利用光场数据特有的重聚焦特性和深度学习的自学习特性，提出了基于特征融合和反馈细化的光场图像显著性检测方法。用 VGG-19 网络对全聚焦图像和焦堆栈图像进行特征提取，并输入到由 ECA 网络和 ConvLSTM 网络组成的特征融合模块中充分融合光场特征，再用 CFM 形成的反馈体系细化信息，消除特征之间的差异，最后使用 ECA 网络加强高层特征，更好地突出显著性。

所提算法充分利用焦堆栈图像的聚焦度信息，有效地分离光场图像的前景区域和背景区域，提高了显著性检测精度。在显著物体较大、背景杂乱、前景与背景纹理相似等情况下，能较好地突出前景抑制背景，获得边界较为清晰的显著图。实验结果证明，所提算法比现有的 RGB 图像、RGB-D 图像和光场图像的显著性检测方法具有更好的检测效果。但是当显著目标与背景区域的深度相近且颜色对比度不高时，边缘效果不太理想。在今后的工作中，将考虑把提取边缘网络运用到光场图像显著性检测中，进一步提高检测的准确性。

参　考　文　献

[1] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis[J]. IEEE

Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254-1259.

[2] 杨锋, 魏国辉, 曹慧, 等. 基于内容的医学图像检索研究进展[J]. 激光与光电子学进展, 2020, 57(6): 060003.
Yang F, Wei G H, Cao H, et al. Research progress on content-based medical image retrieval[J]. Laser & Optoelectronics Progress, 2020, 57(6): 060003.

[3] 孟俊熙, 张莉, 曹洋, 等. 基于Deeplab v3+的图像语义分割算法优化研究[J]. 激光与光电子学进展, 2022, 59(16): 1610009.
Meng J X, Zhang L, Cao Y, et al. Research on optimization of image semantic segmentati-on algorithms based on Deeplab v3+ [J]. Laser & Optoelectronics Progress, 2022, 59(16): 1610009.

[4] 季渊, 郑志杰, 吴浩, 等. 立体视觉中心凹JND模型及其图像压缩硬件实现[J]. 光学学报, 2021, 41(12): 1210001.
Ji Y, Zheng Z J, Wu H, et al. Foveated JND model based on stereo vision and its application in image compression with hardware implementation[J]. Acta Optica Sinica, 2021, 41(12): 1210001.

[5] Li N Y, Ye J W, Ji Y, et al. Saliency detection on light field[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 2806-2813.

[6] Li N Y, Ye J W, Ji Y, et al. Saliency detection on light field[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(8): 1605-1616.

[7] Zhang J, Wang M, Lin L, et al. Saliency detection on light field[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2017, 13(3): 1-22.

[8] Piao Y R, Li X, Zhang M, et al. Saliency detection via depth-induced cellular automata on light field[J]. IEEE Transactions on Image Processing, 2019, 29: 1879-1889.

[9] 李爽, 邓慧萍, 朱磊, 等. 联合聚焦度和传播机制的光场图像显著性检测[J]. 中国图象图形学报, 2020, 25(12): 2578-2586.
Li S, Deng H P, Zhu L, et al. Saliency detection on a light field via the focusness and propagation mechanism[J]. Journal of Image and Graphics, 2020, 25(12): 2578-2586.

[10] Wang X, Dong Y Y, Zhang Q, et al. Region-based depth feature descriptor for saliency detection on light field[J]. Multimedia Tools and Applications, 2021, 80(11): 16329-16346.

[11] Zeng Y, Zhuge Y Z, Lu H C, et al. Multi-source weak supervision for saliency detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 6067-6076.

[12] Chen Z Y, Xu Q Q, Cong R M, et al. Global context-aware progressive aggregation network for salient object detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 10599-10606.

[13] Wei J, Wang S H, Huang Q M. F^3Net: fusion, feedback and focus for salient object detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12321-12328.

[14] Liu N, Zhang N, Han J W. Learning selective self-mutual attention for RGB-D saliency detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 13753-13762.

[15] Li G Y, Liu Z, Ye L W, et al. Cross-modal weighting network for RGB-D salient object detection[M]//Vedaldi A, Bischof H, Brox T, et al. Computer vision-ECCV 2020. Lecture notes in computer science. Cham: Springer, 2020, 12362: 665-681.

[16] Piao Y R, Rong Z K, Zhang M, et al. Deep light-field-driven saliency detection from a single view[C]// Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, August 10-16, 2019, Macao, China. California: International Joint Conferences on Artificial Intelligence Organization, 2019: 904-911.

[17] Zhang J, Liu Y, Zhang S, et al. Light field saliency detection with deep convolutional networks[J]. IEEE Transactions on Image Processing, 2020, 29: 4421-4434.

[18] Wang T T, Piao Y R, Lu H C, et al. Deep learning for light field saliency detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 8837-8847.

[19] Zhang M, Li J J, Wei J, et al. Memory-oriented decoder for light field salient object detection[C]//Advances in Neural Information Processing Systems 2019, December 8-14, 2019, Vancouver, BC, Canada. [S.l.: s.n.], 2019: 898-908.

[20] Zhang M, Ji W, Piao Y R, et al. LFNet: light field fusion network for salient object detection[J]. IEEE Transactions on Image Processing, 2020, 29(99): 6276-6287.

[21] Piao Y R, Rong Z K, Zhang M, et al. Exploit and replace: an asymmetrical two-stream architecture for versatile light field saliency detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 11865-11873.

[22] Wang Q L, Wu B G, Zhu P F, et al. ECA-net: efficient channel attention for deep convolutional neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11531-11539.

[23] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.

[24] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10) [2020-09-14]. https//arxiv.org/abs/1409.1556v6.