

图像处理方法及装置、处理器、电子设备、存储介质

本申请要求于2019年11月27日提交中国专利局、申请号为201911182723.7、发明名称为“图像处理方法及装置、处理器、电子设备、存储介质”，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及图像处理技术领域，尤其涉及一种图像处理方法及装置、处理器、电子设备、存储介质。

背景技术

当公共场所出现人流量过大的情况时，易发生诸如踩踏之类的公共事件。因此如何对公共场所进行人群计数具有重大意义。

传统方法基于深度学习技术可对公共场所的图像进行处理，提取出图像中的特征信息，并依据该特征信息可确定与公共场所的图像对应的人群密度图像，进而可依据人群密度图像确定该公共场所的图像种的人数，实现人群计数。

发明内容

本申请提供一种图像处理方法及装置、处理器、电子设备、存储介质。

第一方面，提供了一种图像处理方法，所述方法包括：

获取待处理图像、第一卷积核和第二卷积核，所述第一卷积核的感受野与所述第二卷积核的感受野不同；

使用所述第一卷积核对所述待处理图像进行卷积处理获得第一特征图像，使用所述第二卷积核对所述待处理图像进行卷积处理获得第二特征图像；

对所述第一特征图像和所述第二特征图像进行融合处理，获得第一人群密度图像。

在该方面中，通过使用感受野不同的第一卷积核和第二卷积核分别对待处理图像进行卷积处理，以提取出不同尺度下的描述待处理图像的内容的信息，分别获得第一特征图像和第二特征图像。通过对第一特征图像和第二特征图像进行融合处理，以利用不同尺度下的描述待处理图像的内容的信息，进而提高获得的与待处理图像对应的人群密度图像的精度。

在一种可能实现的方式中，在所述对所述第一特征图像和所述第二特征图像进行融合处理，获得第一人群密度图像之前，所述方法还包括：

对所述待处理图像进行第一特征提取处理，获得第一自注意力图像，对所述待处理图像进行第二特征提取处理，获得第二自注意力图像，所述第一自注意力图像和所述第二自注意力图像均用于表征所述待处理图像的尺度信息，且所述第一自注意力图像所表征的尺度信息与所述第二自注意力图像所表征的尺度信息不同；

依据所述第一自注意力图像确定所述第一特征图像的第一权重，依据所述第二自注意力图像确定所述第二特征图像的第二权重；

所述对所述第一特征图像和所述第二特征图像进行融合处理，获得第一人群密度图像，包括：

依据所述第一权重和所述第二权重对所述第一特征图像和所述第二特征图像进行融合处理，获得所述第一人群密度图像。

在该种可能实现的方式中，通过对待处理图像分别进行第一特征提取处理和第二特征提取处理以提取不同尺度下的待处理图像的信息，获得第一自注意力图像和第二自注意力图像。依据第一自注意力图像确定第一特征图像的第一权重，依据第二自注意力图像确定第二特征图像的第二权重，并依据第一权重和第二权重对第一特征图像和第二特征图像进行融合处理，可提高获得的第一人群密度图像的精度。

在另一种可能实现的方式中，所述依据所述第一权重和所述第二权重对所述第一特征图像和所述第二特征图像进行融合处理，获得所述第一人群密度图像，包括：

确定所述第一权重与所述第一特征图像之间的点积，获得第三特征图像；

确定所述第二权重与所述第二特征图像之间的点积，获得第四特征图像；

对所述第三特征图像和所述第四特征图像进行融合处理，获得所述第一人群密度图像。

在又一种可能实现的方式中，所述依据所述第一自注意力图像确定所述第一特征图像的第一权重，依据所述第二自注意力图像确定所述第二特征图像的第二权重，包括：

对所述第一自注意力图像和所述第二自注意力图像进行归一化处理，获得所述第一自注意力图像对应的第三自注意力图像和所述第二自注意力图像对应的第四自注意力图像；

将所述第三自注意力图像作为所述第一权重，将所述第四自注意力图像作为所述第二

权重。

在该种可能实现的方式中，通过对第一自注意力图像和第二自注意力图像进行归一化处理，可使第一自注意力图像与第二自注意力图像中相同位置的像素点的像素值的和为1。再通过将第一自注意力图像作为第一权重、将第二自注意力图像作为第二权重对第一特征图像和第二特征图像进行融合处理，可实现对待处理图像中不同图像区域执行不同感受野的卷积处理，进而提高获得的第一人群密度图像的精度。

在又一种可能实现的方式中，在所述使用所述第一卷积核对待处理图像进行卷积处理获得第一特征图像，使用所述第二卷积核对待处理图像进行卷积处理获得第二特征图像之前，所述方法还包括：

对所述待处理图像进行第三特征提取处理，获得第五特征图像；

所述使用所述第一卷积核对待处理图像进行卷积处理获得第一特征图像，使用所述第二卷积核对待处理图像进行卷积处理获得第二特征图像，包括：

使用所述第一卷积核对待第五特征图像进行卷积处理获得所述第一特征图像，使用所述第二卷积核对待第五特征图像进行卷积处理获得所述第二特征图像；

所述对所述待处理图像进行第一特征提取处理，获得第一自注意力图像，对所述待处理图像进行第二特征提取处理，获得第二自注意力图像，包括：

对所述第五特征图像进行所述第一特征提取处理，获得所述第一自注意力图像，对所述第五特征图像进行所述第二特征提取处理，获得所述第二自注意力图像。

在该种可能实现的方式中，在使用第一卷积核对待处理图像进行卷积处理获得第一特征图像，使用第二卷积核对待处理图像进行卷积处理获得第二特征图像之前，对待处理图像进行第三特征提取处理，以提取出待处理图像的特征信息，获得第五特征图像。使用第一卷积核对待第五特征图像进行卷积处理获得第一特征图像，使用第二卷积核对待第五特征图像进行卷积处理获得第二特征图像。这样可从待处理图像中提取出更丰富的特征信息。

在又一种可能实现的方式中，所述第一卷积核和所述第二卷积核均为空洞卷积核，且所述第一卷积核的大小与所述第二卷积核的大小相同，且所述第一卷积核的权重与所述第二卷积核的权重相同，且所述第一卷积核的扩张率与所述第二卷积核的扩张率不同。

在该种可能实现的方式中，在第一卷积核和第二卷积核均为空洞卷积核的情况下，可将第一卷积核的权重与第二卷积核的权重取为相同，且可使第一卷积核的感受野与第二卷积核的感受野不同。这样，使用第一卷积核对待处理图像进行卷积处理获得的第一特征图像包含的信息和使用第二卷积核对待处理图像进行卷积处理获得的第二特征图像包含的信息仅存在尺度上的差异。在对第一特征图像和第二特征图像进行融合处理时，可更好的利用不同尺度下待处理图像的信息提高获得的第一人群密度图像的精度。

在又一种可能实现的方式中，所述第一卷积核或所述第二卷积核的扩张率为参考值。

在该种可能实现的方式中，通过将第一卷积核或第二卷积核的扩张率设为0（即参考值），可在第一卷积核或第二卷积核对待处理图像进行卷积处理时实现对待处理图像进行感受野为1的卷积处理，以更好的提取出待处理图像中尺度小的图像区域的信息。

在又一种可能实现的方式中，所述方法还包括：确定所述第一人群密度图像中的像素值的和，获得所述待处理图像中的人数。

在该种可能实现的方式中，依据第一人群密度图像可确定待处理图像中的人数。

在又一种可能实现的方式中，所述方法应用于人群计数网络；

所述人群计数网络的训练过程包括：

获取样本图像；

使用所述人群计数网络对所述样本图像进行处理，获得第二人群密度图像；

依据所述样本图像与所述第二人群密度图像之间的差异，获得网络损失；

基于所述网络损失调整所述人群计数网络的参数。

在该种可能实现的方式中，使用训练后的人群计数网络对待处理图像进行处理，可获得与待处理图像对应的人群密度图像。

在又一种可能实现的方式中，在所述依据所述样本图像与所述第二人群密度图像之间的差异，获得网络损失之前，所述方法还包括：

依据冲击函数、高斯核以及所述样本图像，获得所述样本图像的真实人群密度图像；

所述依据所述样本图像与所述第二人群密度图像之间的差异，获得网络损失，包括：

依据所述真实人群密度图像与所述第二人群密度图像之间的差异，获得所述网络损失。

在该种可能实现的方式中，将样本图像的真实人群密度图像作为人群计数网络的监督数据，依据真实人群密度图像与第二人群密度图像之间的差异，确定人群计数网络的训练损失，可提高获得的网络损失的精度，进而提升对人群计数网络的训练效果。

在又一种可能实现的方式中，在所述经所述人群计数网络对所述样本图像进行处理，获得第二人群密度图像之前，所述方法还包括：

对所述样本图像进行预处理, 获得至少一张预处理后的图像;

所述经所述人群计数网络对所述样本图像进行处理, 获得第二人群密度图像, 包括:

使用所述人群计数网络对所述至少一张预处理后的图像进行处理, 获得至少一张第三人群密度图像, 所述预处理后的图像与所述第三人群密度图像一一对应;

所述依据所述样本图像与所述第二人群密度图像之间的差异, 获得网络损失, 包括:

依据所述至少一张预处理后的图像中的目标图像和与所述目标图像对应的第三人群密度图像之间的差异, 获得所述网络损失。

在该种可能实现的方式中, 在将样本图像输入至人群计数网络之前, 通过对样本图像进行预处理, 获得至少一张预处理后的图像, 并将上述至少一张预处理后的图像作为训练数据输入至人群计数网络。这样, 可达到扩充人群计数网络的训练数据集的效果。

在又一种可能实现的方式中, 所述预处理包括: 从所述样本图像中截取预定尺寸的图像、对所述样本图像或所述预定尺寸的图像进行翻转处理中的至少一种。

第二方面, 提供了一种图像处理装置, 所述装置包括:

获取单元, 用于获取待处理图像、第一卷积核和第二卷积核, 所述第一卷积核的感受野与所述第二卷积核的感受野不同;

卷积处理单元, 用于使用所述第一卷积核对所述待处理图像进行卷积处理获得第一特征图像, 使用所述第二卷积核对所述待处理图像进行卷积处理获得第二特征图像;

融合处理单元, 用于对所述第一特征图像和所述第二特征图像进行融合处理, 获得第一人群密度图像。

在一种可能实现的方式中, 所述装置还包括:

特征提取处理单元, 用于在所述对所述第一特征图像和所述第二特征图像进行融合处理, 获得第一人群密度图像之前, 对所述待处理图像进行第一特征提取处理, 获得第一自注意力图像, 对所述待处理图像进行第二特征提取处理, 获得第二自注意力图像, 所述第一自注意力图像和所述第二自注意力图像均用于表征所述待处理图像的尺度信息, 且所述第一自注意力图像所表征的尺度信息与所述第二自注意力图像所表征的尺度信息不同;

第一确定单元, 用于依据所述第一自注意力图像确定所述第一特征图像的第一权重, 依据所述第二自注意力图像确定所述第二特征图像的第二权重;

所述融合处理单元用于:

依据所述第一权重和所述第二权重对所述第一特征图像和所述第二特征图像进行融合处理, 获得所述第一人群密度图像。

在另一种可能实现的方式中, 所述融合处理单元具体用于:

确定所述第一权重与所述第一特征图像之间的点积, 获得第三特征图像;

确定所述第二权重与所述第二特征图像之间的点积, 获得第四特征图像;

对所述第三特征图像和所述第四特征图像进行融合处理, 获得所述第一人群密度图像。

在又一种可能实现的方式中, 所述第一确定单元用于:

对所述第一自注意力图像和所述第二自注意力图像进行归一化处理, 获得所述第一自注意力图像对应的第三自注意力图像和所述第二自注意力图像对应的第四自注意力图像;

将所述第三自注意力图像作为所述第一权重, 将所述第四自注意力图像作为所述第二权重。

在又一种可能实现的方式中, 所述特征提取处理单元, 还用于在所述使用所述第一卷积核对所述待处理图像进行卷积处理获得第一特征图像, 使用所述第二卷积核对所述待处理图像进行卷积处理获得第二特征图像之前, 对所述待处理图像进行第三特征提取处理, 获得第五特征图像;

所述卷积处理单元用于:

使用所述第一卷积核对所述第五特征图像进行卷积处理获得所述第一特征图像, 使用所述第二卷积核对所述第五特征图像进行卷积处理获得所述第二特征图像;

所述特征提取处理单元还用于:

对所述第五特征图像进行所述第一特征提取处理, 获得所述第一自注意力图像, 对所述第五特征图像进行所述第二特征提取处理, 获得所述第二自注意力图像。

在又一种可能实现的方式中, 所述第一卷积核和所述第二卷积核均为空洞卷积核, 且所述第一卷积核的大小与所述第二卷积核的大小相同, 且所述第一卷积核的权重与所述第二卷积核的权重相同, 且所述第一卷积核的扩张率与所述第二卷积核的扩张率不同。

在又一种可能实现的方式中, 所述第一卷积核或所述第二卷积核的扩张率为参考值。

在又一种可能实现的方式中, 所述装置还包括: 第二确定单元, 用于确定所述第一人群密度图像中的像素值的和, 获得所述待处理图像中的人数。

在又一种可能实现的方式中, 所述装置执行的图像处理方法应用于人群计数网络;

所述装置还包括: 训练单元, 用于对所述人群计数网络进行训练, 所述人群计数网络的训练过程包括:

获取样本图像；
使用所述人群计数网络对所述样本图像进行处理，获得第二人群密度图像；
依据所述样本图像与所述第二人群密度图像之间的差异，获得网络损失；
基于所述网络损失调整所述人群计数网络的参数。
在又一种可能实现的方式中，所述训练单元还用于：
在所述依据所述样本图像与所述第二人群密度图像之间的差异，获得网络损失之前，
依据冲击函数、高斯核以及所述样本图像，获得所述样本图像的真实人群密度图像；
依据所述真实人群密度图像与所述第二人群密度图像之间的差异，获得所述网络损失。
在又一种可能实现的方式中，所述训练单元还用于：
在所述经所述人群计数网络对所述样本图像进行处理，获得第二人群密度图像之前，
对所述样本图像进行预处理，获得至少一张预处理后的图像；
使用所述人群计数网络对所述至少一张预处理后的图像进行处理，获得至少一张第三人群密度图像，所述预处理后的图像与所述第三人群密度图像一一对应；
依据所述至少一张预处理后的图像中的目标图像和与所述目标图像对应的第三人群密度图像之间的差异，获得所述网络损失。
在又一种可能实现的方式中，所述预处理包括：从所述样本图像中截取预定尺寸的图像、对所述样本图像或所述预定尺寸的图像进行翻转处理中的至少一种。
第三方面，提供了一种处理器，所述处理器用于执行如上述第一方面及其任意一种可能实现的方式的方法。
第四方面，提供了一种电子设备，包括：相互连接的处理器和存储器，所述存储器用于存储计算机程序代码，所述计算机程序代码包括计算机指令，当所述处理器执行所述计算机指令时，所述电子设备执行如上述第一方面及其任意一种可能实现的方式的方法。
第五方面，提供了一种计算机可读存储介质，所述计算机可读存储介质中存储有计算机程序，所述计算机程序包括程序指令，所述程序指令当被电子设备的处理器执行时，使所述处理器执行如上述第一方面及其任意一种可能实现的方式的方法。
第六方面，提供了一种包含指令的计算机程序产品，当所述计算机程序产品在计算机上运行时，使得计算机执行上述第一方面及其任一种可能的实现方式的方法。
应当理解的是，以上的一般描述和后文的细节描述仅是示例性和解释性的，而非限制本公开。

附图说明

为了更清楚地说明本申请实施例或背景技术中的技术方案，下面将对本申请实施例或背景技术中所需要使用的附图进行说明。

此处的附图被并入说明书中并构成本说明书的一部分，这些附图示出了符合本公开的实施例，并与说明书一起用于说明本公开的技术方案。

- 图 1 为本申请实施例提供了一种图像处理方法的流程图示意图；
- 图 2a 为本申请实施例提供了一种卷积核的示意图；
- 图 2b 为本申请实施例提供了一种卷积核的权重的示意图；
- 图 3 为本申请实施例提供了一种相同位置的元素的示意图；
- 图 4 为本申请实施例提供了一种人群图像示意图；
- 图 5 为本申请实施例提供的另一种图像处理方法的流程图示意图；
- 图 6a 为本申请实施例提供了一种空洞卷积核的示意图；
- 图 6b 为本申请实施例提供的另一种空洞卷积核的示意图；
- 图 7 为本申请实施例提供的又一种空洞卷积核的示意图；
- 图 8 为本申请实施例提供了一种人群计数网络的结构示意图；
- 图 9 为本申请实施例提供了一种尺度感知型卷积层的结构示意图；
- 图 10 为本申请实施例提供了一种图像处理装置的结构示意图；
- 图 11 为本申请实施例提供了一种图像处理装置的硬件结构示意图。

具体实施方式

为了使本技术领域的人员更好地理解本申请方案，下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

本申请的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别不同对象，而不是用于描述特定顺序。此外，术语“包括”和“具有”以及它们任何变形，意图在于覆盖不排他的包含。例如包含了一系列步骤或单元的过程、方法、系统、产品或

设备没有限定于已列出的步骤或单元，而是可选地还包括没有列出的步骤或单元，或可选地还包括对于这些过程、方法、产品或设备固有的其他步骤或单元。

在本文中提及“实施例”意味着，结合实施例描述的特定特征、结构或特性可以包含在本申请的至少一个实施例中。在说明书中的各个位置出现该短语并不一定均是指相同的实施例，也不是与其它实施例互斥的独立的或备选的实施例。本领域技术人员显式地和隐式地理解的是，本文所描述的实施例可以与其它实施例相结合。

在公共场所（例如广场、超市、地铁站、码头等地方）中，有时会存在人流量过多的情况，进而导致人群过于密集的情况发生。这时易发生一些公共事故，例如踩踏事件。因此，如何对公共场所进行人群计数就变得非常有意义。

随着深度学习技术的发展，基于深度学习的方法可确定图像中的人数，实现人群计数。传统的深度学习方法通过使用一个卷积核对整张图像进行卷积处理以提取出图像中的特征信息，并依据特征信息确定图像中的人数。由于一个卷积核的感受野是固定不变的，若使用一个卷积核对整张图像进行卷积处理，即相当于对图像中不同尺度的内容进行相同感受野的卷积处理，而不同人物在图像中的尺度不同，这将导致不能有效提取出图像中的尺度信息，进而导致确定的人数的误差。

本申请中，图像中近处的人物对应的图像尺度大，图像中远处的人物对应的图像尺度小。本申请实施例中的“远”指与图像中人物对应的真实人物与采集上述图像的成像设备之间的距离远，“近”指与图像中人物对应的真实人物与采集上述图像的成像设备之间的距离近。

在卷积神经网络中，感受野（receptive field）的定义是卷积神经网络每一层输出的特征图（feature map）上的像素点在输入图片上映射的区域大小。本申请中，卷积核的感受野即为使用该卷积核对图像进行卷积处理感受野。

本申请实施例提供的技术方案可提取出图像中的尺度信息，进而提升确定的人数的精度。

下面结合本申请实施例中的附图对本申请实施例进行描述。

请参阅图1，图1是本申请实施例（一）提供的一种图像处理方法的流程示意图。

101、获取待处理图像、第一卷积核和第二卷积核，上述第一卷积核的感受野与上述第二卷积核的感受野不同。

本申请实施例的执行主体可以是服务器、手机、电脑、平板电脑等终端硬件。本申请实施例提供的方法也可通过处理器运行计算机可执行代码的方式执行。上述待处理图像可以是任意图像。例如，待处理图像可以包含人物对象，其中，待处理图像可以只包括人脸，并无躯干、四肢（下文将躯干和四肢称为人体），也可以只包括人体，不包括人脸，还可以只包括下肢或上肢。本申请对待处理图像具体包含的人体区域不做限定。又例如，待处理图像可以包含动物。再例如，待处理图像可以包含植物。本申请对待处理图像中包含的内容不做限定。

在进行接下来的阐述之前，首先对本申请实施例中的卷积核的权重的含义进行定义。

本申请实施例中，通道为1的卷积核以 $n \times n$ 的矩阵的形式存在，该矩阵中包含 $n \times n$ 个元素，每个元素均有一个取值，该矩阵中元素的取值即为卷积核的权重。在图2a所示的 3×3 的卷积核中，若元素a的取值为44、元素b的取值为118、元素c的取值为192、元素d的取值为32、元素e的取值为83、元素f的取值为204、元素g的取值为61、元素h的取值为174、元素i的取值为250，则该 3×3 的卷积核的权重为图2b所示的 3×3 的矩阵。

本申请实施例中，在满足第一卷积核的感受野与第二卷积核的感受野不同的情况下，第一卷积核和第二卷积核均可任意大小的卷积核，且第一卷积核的权重和第二卷积核的权重均可任意自然数，本实施例对第一卷积核的大小、第二卷积核的大小、第一卷积核的权重以及第二卷积核的权重均不做限定。

获取待处理图像的方式可以是接收用户通过输入组件输入的待处理图像，也可以是接收终端发送的待处理图像。获取第一卷积核的方式可以是接收用户通过输入组件输入的第一卷积核，也可以是接收终端发送的第一卷积核。获取第二卷积核的方式可以是接收用户通过输入组件输入的第二卷积核，也可以是接收终端发送的第二卷积核。上述输入组件包括：键盘、鼠标、触控屏、触控板和音频输入器等。上述终端包括手机、计算机、平板电脑、服务器等。

102、使用上述第一卷积核对上述待处理图像进行卷积处理获得第一特征图像，使用上述第二卷积核对上述待处理图像进行卷积处理获得第二特征图像。

由于第一卷积核的感受野与第二卷积核的感受野不同，使用第一卷积核对待处理图像进行卷积处理和使用第二卷积核对待处理图像进行卷积处理相当于以不同的感受野“观察”

图像, 实现获得不同尺度下的图像信息。即第一特征图像和第二特征图像均包含用于描述待处理图像的内容的信息, 但第一特征图像包含的信息的尺度与第二特征图像包含的信息的尺度不同。

103、对上述第一特征图像和上述第二特征图像进行融合处理, 获得第一人群密度图像。

本申请实施例中, 人群密度图像包含人群密度信息。人群密度图像中的每个像素点的像素值表征在该像素点处的人数。举例来说, 人群密度图像中的像素点 A 的像素值为 0.05, 则像素点 A 处有 0.05 个人。

需要理解的是, 由于一个人覆盖的图像区域包含至少一个像素点, 当一个人覆盖的图像区域为 1 个像素点时, 该像素点对应的像素值为 1, 当一个人覆盖的图像区域为至少两个像素点时, 该至少两个像素点的像素值的和为 1。因此, 人群密度图像中的像素值的取值范围: 大于或等于 0 且小于或等于 1。举例来说, 人物 A 覆盖的图像区域包含像素点 a、像素点 b 和像素点 c, 则像素点 a 的像素值+像素点 b 的像素值+像素点 c 的像素值=1。

上述第一人群密度图像为与待处理图像对应的人群密度图像, 可表征待处理图像中的人群密度分布。第一人群密度图像的尺寸与待处理图像的尺寸相同。本实施例中图像的尺寸指图像的宽和高。第一人群密度图像中的第一像素点的像素值可用于表征待处理图像中的第二像素点处的人数。其中, 第一像素点在第一人群密度图像中的位置与第二像素点在待处理图像中的位置相同。

本申请实施例中, 两张图像中相同位置的像素点可参见图 3, 如图 3 所示, 像素点 A₁₁ 在图像 A 中的位置与像素点 B₁₁ 在图像 B 中的位置相同, 像素点 A₁₂ 在图像 A 中的位置与像素点 k 在图像 B₁₂ 中的位置相同, 像素点 A₁₃ 在图像 A 中的位置与像素点 B₁₃ 在图像 B 中的位置相同, 像素点 A₂₁ 在图像 A 中的位置与像素点 B₂₁ 在图像 B 中的位置相同, 像素点 A₂₂ 在图像 A 中的位置与像素点 B₂₂ 在图像 B 中的位置相同, 像素点 A₂₃ 在图像 A 中的位置与像素点 B₂₃ 在图像 B 中的位置相同, 像素点 A₃₁ 在图像 A 中的位置与像素点 B₃₁ 在图像 B 中的位置相同, 像素点 A₃₂ 在图像 A 中的位置与像素点 B₃₂ 在图像 B 中的位置相同, 像素点 A₃₃ 在图像 A 中的位置与像素点 B₃₃ 在图像 B 中的位置相同。

若像素点 x 在图像 X 中的位置与像素点 y 在图像 Y 中的位置相同, 为简洁表述, 下文将像素点 x 称为图像 X 中与像素点 y 位置相同的像素点, 或将像素点 y 称为图像 Y 中与像素点 x 位置相同的像素点。

由于第一特征图像包含描述待处理图像的图像内容的信息的尺度和第二待处理图像包含描述待处理图像的图像内容的信息的尺度不同, 通过对第一特征图像和第二特征图像进行融合处理 (例如对应位置的像素值加权处理等), 可利用不同尺度下的描述待处理图像的图像内容的信息生成待处理图像对应的人群密度图像, 即第一人群密度图像。这样, 可提高获得的与待处理图像对应的人群密度图像的精度, 进而提升获得的待处理图像中人数的精度。

需要理解的是, 本实施例阐述了通过两个感受野不同的卷积核 (即第一卷积核和第二卷积核) 分别对待处理图像进行卷积处理, 获得两个尺度下的描述待处理图像的图像内容的信息。但在实际使用中, 也可通过三个或三个以上感受野不同的卷积核分别对待处理图像进行卷积处理, 以获得三个或三个以上尺度下的描述待处理图像的图像内容的信息, 并将该三个或三个以上尺度下的描述待处理图像的图像内容的信息进行融合, 获得与待处理图像对应的人群密度图像。

可选的, 在获得第一人群密度图像后, 可通过确定第一人群密度图像中所有像素点的像素值的和, 得到待处理图像中的人数。

本实施例通过使用感受野不同的第一卷积核和第二卷积核分别对待处理图像进行卷积处理, 以提取出不同尺度下的描述待处理图像的内容的信息, 分别获得第一特征图像和第二特征图像。通过对第一特征图像和第二特征图像进行融合处理, 以利用不同尺度下的描述待处理图像的内容的信息, 提高获得的与待处理图像对应的人群密度图像的精度, 进而提升获得的待处理图像中人数的精度。

在图像中, 近处的人物覆盖的图像区域的面积比远处的人物覆盖的图像区域的面积大。例如, 图 4 中人物 A 相较于人物 B 为近处的人物, 且人物 A 覆盖的图像区域的面积比人物 B 覆盖的图像区域的面积大。而近处的人物覆盖的图像区域的尺度大, 远处的人物覆盖的图像区域的尺度小。因此, 人物覆盖的图像区域的面积与人物覆盖的图像区域的尺度呈正相关。显然, 当卷积处理的感受野与人物覆盖的图像区域的面积相同时, 通过卷积处理获得的人物覆盖的图像区域的信息最丰富 (下文将可获得人物覆盖的图像区域的最丰富的信息的感受野称为人物覆盖区域的最佳感受野)。也就是说, 人物覆盖的图像区域的尺度与人物覆盖区域的最佳感受野呈正相关。

虽然实施例 (一) 通过使用感受野不同的第一卷积核和第二卷积核分别对待处理图像进行卷积处理获得不同尺度下的描述待处理图像的内容的信息。但第一卷积核的感受野和第二卷积核的感受野均为固定的, 而待处理图像中不同的图像区域的尺度不同, 因此分别

使用第一卷积核和第二卷积核对待处理图像进行卷积处理无法获得待处理图像中每个图像区域的感受野,即无法使获得的待处理图像中不同图像区域的信息均为最丰富。为此,本申请实施例还提供了一种通过在对第一特征图像和第二特征图像进行融合处理时为第一特征图像和第二特征图像赋予权重,以实现对待处理图像中不同尺度的图像区域进行不同感受野的卷积处理,进而获得更丰富的信息。

请参阅图5,图5是本申请实施例(二)提供的另一种图像处理方法的流程示意图。

501、对上述待处理图像进行第一特征提取处理,获得第一自注意力图像,对上述待处理图像进行第二特征提取处理,获得第二自注意力图像,上述第一自注意力图像和上述第二自注意力图像均用于表征上述待处理图像的尺度信息,且上述第一自注意力图像所表征的尺度信息与上述第二自注意力图像所表征的尺度信息不同。

本申请实施例中,特征提取处理可以是卷积处理,也可以是池化处理,还可以是卷积处理和池化处理的结合。本申请对第一特征提取处理的实现方式和第二特征提取处理的实现方式不做限定。

在一种可能实现的方式中,依次通过多层卷积层对待处理图像进行逐级卷积处理,实现对待处理图像的第一特征提取处理,获得第一自注意力图像。同理,可依次通过多层卷积层对待处理图像进行逐级卷积处理,实现对待处理图像的第二特征提取处理,获得第二自注意力图像。

可选的,在使用第一卷积核对待处理图像进行卷积处理获得第一特征图像,使用第二卷积核对待处理图像进行卷积处理获得第二特征图像之前,可对待处理图像进行第三特征提取处理,以提取出待处理图像的特征信息,获得第五特征图像。使用第一卷积核对第五特征图像进行卷积处理获得第一特征图像,使用第二卷积核对待第五特征图像进行卷积处理获得第二特征图像。这样可从待处理图像中提取出更丰富的特征信息。

上述第一自注意力图像的尺寸和上述第二自注意力图像的尺寸均与待处理图像的尺寸相同。上述第一自注意力图像和上述第二自注意力图像均可用于表征待处理图像的尺度信息(即待处理图像中不同图像区域的尺度),且第一自注意力图像所表征的尺度信息与第二自注意力图像所表征的尺度信息不同。本申请实施例中,图像(包括:上述第一特征图像、上述第二特征图像、上述第一自注意力图像、上述第二自注意力图像、下文将要提及的第三自注意力图像等)的尺度与对待处理图像进行特征提取处理(包括上述第一特征提取处理、上述第二特征提取处理以及上述第三特征提取处理)时所使用的卷积核的感受野匹配。例如,使用大小为 3×3 的卷积核对待处理图像进行卷积处理得到的图像的尺度为 a ,使用大小为 5×5 的卷积核对待处理图像进行卷积处理得到的图像的尺度为 b ,那么使用大小为 3×3 的卷积核对待处理图像进行特征提取处理得到的自注意力图像的尺度为 a (即该自注意力图像可表征待处理图像在尺度 a 的信息),使用大小为 5×5 的卷积核对待处理图像进行特征提取处理得到的特征图像的尺度为 b 。

举例来说(例1),第一自注意力图像表征待处理图像在尺度 a 下的信息,第二自注意力图像表征待处理图像在尺度 b 下的信息,其中,尺度 a 大于尺度 b 。

第一自注意力图像中的像素点的像素值和第二自注意力图像中的像素点的像素值的取值范围均为:大于或等于0,且小于或等于1。第一自注意力图像(或第二自注意力图像)中的某个像素点的像素值越接近于1,表征在待处理图像中与该像素点位置相同的像素点的最佳尺度与第一自注意力图像(或第二自注意力图像)所表征的尺度越接近。本申请实施例中,最佳尺度即为与该像素点的最佳感受野对应的尺度。

接着例1继续举例,像素点 a 和像素点 b 为第一自注意力图像中的两个不同的像素点,像素点 c 为待处理图像中与像素点 a 在第一自注意力图像中的位置相同的像素点,像素点 d 为待处理图像中与像素点 b 在第一自注意力图像中的位置相同的像素点。若像素点 a 的像素值为0.9,像素点 b 的像素值为0.7。则像素点 c 的最佳尺度与尺度 a 之间的差异小于像素点 d 的最佳尺度与尺度 a 之间的差异。

502、依据上述第一自注意力图像确定上述第一特征图像的第一权重,依据上述第二自注意力图像确定上述第二特征图像的第二权重。

可选的,上述第一自注意力图像所表征的尺度与第一特征图像的尺度相同,上述第二自注意力图像所表征的尺度与第二特征图像的尺度相同。则第一自注意力图像中的像素点的像素值与1越接近表征第一特征图像中与该像素点在第一自注意力图像中的位置相同的像素点的最佳尺度与第一特征图像的尺度越接近,第二自注意力图像中的像素点的像素值与1越接近表征第二特征图像中与该像素点在第二自注意力图像中的位置相同的像素点的最佳尺度与第二特征图像的尺度越接近。

因此,可依据第一自注意力图像确定第一特征图像的第一权重,以调整第一特征图像中的像素点的尺度,使第一特征图像中的像素点更接近最佳尺度。同理,可依据第二自注意力图像确定第二特征图像的第二权重,以调整第二特征图像中的像素点的尺度,使第二特征图像中的像素点更接近最佳尺度。

在一种可能实现的方式中,可对第一自注意力图像和第二自注意力图像进行归一化处理,获得第一自注意力图像对应的第三自注意力图像和第二自注意力图像对应的第四自注意力图像。将第三自注意力图像作为上述第一权重,将第四自注意力图像作为上述第二权重。

在上述可能实现的方式中,通过对第一自注意力图像和第二自注意力图像进行归一化处理,可使第一自注意力图像与第二自注意力图像中相同位置的像素点的像素值的和为1。举例来说,像素点a在第一自注意力图像中的位置与像素点b在第二自注意力图像中的位置相同,则对第一自注意力图像和第二自注意力图像进行归一化处理后像素点a的像素值和像素点b的像素值的和为1。如像素点c在第三自注意力图像中的位置与像素点a在第一自注意力图像中的位置相同,像素点d在第四自注意力图像中的位置与像素点b在第二自注意力图像中的位置相同,则像素点c的像素值与像素点d的像素值的和为1。

可选的,上述归一化处理可通过将第一自注意力图像和第二自注意力图像分别输入至softmax函数实现。需要理解的是,若第一自注意力图像和第二自注意力图像均包含多个通道的图像,则将第一自注意力图像与第二自注意力图像中相同通道的图像分别输入至softmax函数。例如,第一自注意力图像和第二自注意力图像均包含2个通道的图像,则在对第一自注意力图像和第二自注意力图像进行归一化处理时,可将第一自注意力图像中第一个通道的图像和第二自注意力图像中第一个通道的图像输入至softmax函数,获得第三自注意力图像中第一个通道的图像以及第四自注意力图像中第一个通道的图像。

503、依据上述第一权重和上述第二权重对上述第一特征图像和上述第二特征图像进行融合处理,获得上述第一人群密度图像。

由于获得第一特征图像的卷积处理的感受野和获得第二特征图像的卷积处理的感受野不同。通过将第三自注意力图像作为第一特征图像的第一权重,将第四自注意力图像作为第二特征图像的第二权重对第一特征图像和第二特征图像进行融合处理,可对待处理图像中的不同图像区域进行最佳感受野下的卷积处理。这样,可充分提取待处理图像中不同图像区域的信息,使获得的与待处理图像对应的人群密度图像的精度更高。

在一种依据第一权重和第二权重对第一特征图像和第二特征图像进行融合处理,获得第一人群密度图像的实现方式中,计算第一权重与第一特征图像之间的点积,获得第三特征图像,计算第二权重与第二特征图像之间的点积,获得第四特征图像。通过对第三特征图像和第四特征图像进行融合处理(例如相同位置的像素值相加),可获得第一人群密度图像。

本实施例通过对待处理图像分别进行第一特征提取处理和第二特征提取处理以提取不同尺度下的待处理图像的信息,获得第一自注意力图像和第二自注意力图像。依据第一自注意力图像确定第一特征图像的第一权重,依据第二自注意力图像确定第二特征图像的第二权重,并依据第一权重和第二权重对第一特征图像和第二特征图像进行融合处理,可提高获得的第一人群密度图像的精度。

在实施例(一)和实施例(二)中的第一卷积核的权重和第二卷积核的权重不同时,使用第一卷积核对待处理图像进行卷积处理提取出的特征信息的侧重点与使用第二卷积核对待处理图像进行卷积处理提取出的特征信息的侧重点不同。例如,使用第一卷积核对待处理图像进行卷积处理侧重于提取出待处理图像中人物的属性特征(如衣服颜色、裤子长度),而使用第二卷积核对待处理图像进行卷积处理侧重于提取出待处理图像中人物的轮廓特征(该轮廓特征可用于识别待处理图像中是否包含人物)。再考虑到第一卷积核的感受野和第二卷积核的感受野的不同。这样,在后续对提取出的第一特征图像和第二特征图像进行融合处理时,需要将不同尺度下的不同特征信息进行融合(如将尺度a下的属性特征与尺度b下的轮廓特征融合),这将给尺度信息的融合带来困难。

为此,本申请实施例还提供了一种技术方案,将第一卷积核的权重和第二卷积核的权重取为相同,以减小对第一特征图像和第二特征图像进行融合处理时非尺度信息的融合,提高尺度信息融合的效果,进而提高获得的第一人群密度图像的精度。

由于若第一卷积核和第二卷积核为常规卷积核,在第一卷积核的感受野与第二卷积核的感受野不同的情况下,第一卷积核的权重与第二卷积核的权重不可能相同。因此,在接下来阐述的技术方案中第一卷积核和第二卷积核均为空洞卷积核,且第一卷积核的大小与第二卷积核的大小相同,且第一卷积核的权重与第二卷积核的权重相同,且第一卷积核的扩张率与第二卷积核的扩张率不同。

举例来说,如图6a、图6b所示的两个空洞卷积核,上述两个空洞卷积核的大小均为 3×3 ,其中,图6a所示的空洞卷积核和图6b所示的空洞卷积核中的黑色区域表示有参数,白色部分表示没有参数(即参数为0)。可选的,可将图6a所示的空洞卷积核的权重与图6b所示的空洞卷积核的权重取为相同。此外,从图中可以看出,由于图6a所示的空洞卷积核的扩张率为2,图6b所示的空洞卷积核的扩张率为1,图6a所示的空洞卷积核的感受野与图6b所示的空洞卷积核的感受野不同,具体的,图6a所示的空洞卷积核的感受野(5×5)

比图 6b 所示的空洞卷积核的感受野 (3*3) 大。

在第一卷积核和第二卷积核均为空洞卷积核的情况下, 可将第一卷积核的权重与第二卷积核的权重取为相同, 且可使第一卷积核的感受野与第二卷积核的感受野不同。这样, 使用第一卷积核对待处理图像进行卷积处理获得的第一特征图像包含的信息和使用第二卷积核对待处理图像进行卷积处理获得的第二特征图像包含的信息仅存在尺度上的差异。在对第一特征图像和第二特征图像进行融合处理时, 可更好的利用不同尺度下待处理图像的信息提高获得的第一人群密度图像的精度。

可选的, 可通过使第一卷积核和第二卷积核共享同一组权重的方式使第一卷积核的权重与第二卷积核的权重相同, 这样, 在后续分别使用第一卷积核和第二卷积核对待处理图像进行卷积处理时, 可减少所需处理的参数的数量。

在空洞卷积核的大小一定的情况下, 空洞卷积核的感受野与空洞卷积核的扩张率呈正相关。当空洞卷积核的扩张率为 1 时, 空洞卷积核的感受野与相同大小的常规卷积核的感受野相同, 如: 图 6b 所示的空洞卷积核的扩张率为 1, 此时该空洞卷积核的感受野与大小为 3*3 的常规卷积核的感受野相同。

考虑到待处理图像中存在最佳尺度较小的像素区域, 这些尺度较小的图像区域需要使用较小的感受野的卷积处理才能提取出更丰富的信息。为此本申请实施例还提供了一种将空洞卷积核的扩张率设为 0 (即参考值), 使空洞卷积核的感受野小于常规卷积核的感受野, 以更好的提取出待处理图像中尺度较小的图像区域的信息。

下面将从理论上推导扩张率为 0 的空洞卷积核如何实现。

假设使用一个大小为 3*3, 扩张率为 d 的空洞卷积核对待处理图像进行卷积处理, 则该卷积处理的过程满足下式:

$$O_{(x,y)} = \sum_{i=-1}^1 \sum_{j=-1}^1 I_{(x+i+d,y+j+d)} * w_{(1+i,1+j)} + b \dots \text{式 (1)}$$

其中, x 和 y 分别为空洞卷积核滑动至待处理图像上某个像素点时空洞卷积核的中心像素点的位置。($x+i, y+j$) 为待处理图像中的采样点在待处理图像中的坐标, $w_{(1+i,1+j)}$ 为空洞卷积核的权重, b 为空洞卷积核的偏差。 I 为待处理图像, O 为使用空洞卷积核对待处理图像进行卷积处理获得的特征图像。

当 $d = 0$ 时, 式 (1) 可转化为下式:

$$\begin{aligned} O_{(x,y)} &= \sum_{i=-1}^1 \sum_{j=-1}^1 I_{(x+i+0,y+j+0)} * w_{(1+i,1+j)} + b \\ &= \sum_{i=-1}^1 \sum_{j=-1}^1 I_{(x+i,y+j)} * w_{(1+i,1+j)} + b = \sum_{i=-1}^1 \sum_{j=-1}^1 \left(I_{(x+i,y+j)} * w_{(1+i,1+j)} + \frac{b}{9} \right) \\ &= \sum_{k=1}^9 I_{(x,y)} * w'_k + b'_k \dots \text{式 (2)} \end{aligned}$$

其中, w'_k 表示大小为 1*1 的常规卷积核的权重, b'_k 表示大小为 1*1 的常规卷积核的偏差。从式 (2) 可以看出使用一个大小为 3*3、扩张率为 0 的空洞卷积核对待处理图像进行卷积处理等价于使用 9 个大小为 1*1 的常规卷积核分别对待处理图像进行卷积处理。因此, 扩张率为 0 的空洞卷积核可使用 9 个 1*1 的常规卷积核代替, 即扩张率为 0 的空洞卷积核中所有权重均位于空洞卷积核上的同一个位置。图 7 所示为大小为 3*3、扩张率为 0 的空洞卷积核, 图 6 所示的空洞卷积核中的黑色区域即为权重所在的位置。从图 6 所示的空洞卷积核可以看出, 扩张率为 0 的空洞卷积核的感受野为 1。

本申请实施例中, 在第一卷积核为空洞卷积核的情况下, 通过将第一卷积核的扩张率设为 0, 可在使使用第一卷积核对待处理图像进行卷积处理时实现对待处理图像进行感受野为 1 的卷积处理, 以更好的提取出待处理图像中尺度小的图像区域的信息。

本申请实施例还提供了一种人群计数网络, 可用于实现前文所提及的技术方案。请参阅图 8, 图 8 为本申请实施例提供的一种人群计数网络的结构示意图。如图 8 所示, 人群计数网络中的网络层依次串联, 共包含 11 层卷积层和 9 层池化层和 6 层尺度感知型卷积层。

将待处理图像输入至人群计数网络, 经第一层卷积层对待处理图像进行处理获得第一

层卷积层输出的图像，第一层卷积层输出的图像经第二层卷积层的处理获得第二层卷积层输出的图像，第二层卷积层输出的图像经第一层池化层的处理获得第一层池化层输出的图像，...，第十层卷积层输出的图像经第一层尺度感知型卷积层的处理获得第一层尺度感知型卷积层输出的图像，...，第九层池化层输出的图像经第十层卷积层的处理获得第一人群密度图像。

可选的，人群计数网络中除上述第十层卷积层之外的所有卷积层中的卷积核的大小均可为 3×3 ，第十层卷积层中的卷积核的大小为 1×1 。第一层卷积层中卷积核的数量和第二层卷积层中卷积核的数量均可为 64，第三层卷积层中卷积核的数量和第四层卷积层中卷积核的数量均可为 128，第五层卷积层中卷积核的数量、第六层卷积层中卷积核的数量以及第七层卷积层中卷积核的数量均可为 256，第八层卷积层中卷积核的数量、第九层卷积层中卷积核的数量以及第十层卷积层中卷积核的数量均可为 512，第十层卷积层中卷积核的数量为 1。

人群计数网络中的池化层可以为最大池化层，也可以是平均池化层，本申请对此不做限定。

尺度感知型卷积层的结构示意图可参见图 9。如图 9 所示，尺度感知型卷积层包括三个空洞卷积核、一个自注意力模块。上述三个空洞卷积核的结构可参见图 6a、图 6b 和图 7，此处将不再赘述。上述自注意力模块包含 3 个并联的卷积层。

尺度感知型卷积层的输入图像分别经 3 个不同感受野的空洞卷积核的处理，分别获得第六特征图像、第七特征图像和第八特征图像。

尺度感知型卷积层的输入图像分别经自注意力模块中的 3 个卷积层的卷积处理，分别获得第五自注意力图像、第六自注意力图像和第七自注意力图像。

第六特征图像的尺度与第五自注意力图像的尺度相同，第七特征图像的尺度与第六自注意力图像的尺度相同，第八特征图像的尺度与第七自注意力图像的尺度相同。通过将第五自注意力图像作为第六特征图像的权重，将第六自注意力图像作为第七特征图像的权重，将第七自注意力图像作为第八特征图像的权重，对第六特征图像、第七特征图像和第八特征图像进行融合处理，获得尺度感知型卷积层的输出图像。即将第五自注意力图像与第六特征图像进行点乘获得第九特征图像，将第六自注意力图像与第七特征图像进行点乘获得第十特征图像，将第七自注意力图像与第八特征图像进行点乘获得第十一特征图像。对第九特征图像、第十特征图像和第十一特征图像进行融合处理，获得尺度感知型卷积层的输出图像。可选的上述融合处理可以是进行融合处理的两张图像中相同位置的像素点的像素值相加。

需要理解的是，图 8 所示的人群计数网络中网络层的具体数量仅为一个示例，不应对本申请构成限定。

在应用图 8 所示的人群计数网络对待处理图像执行人群计数任务之前，需对人群计数网络进行训练。为此，本申请还提供了一种人群计数网络的训练方法。该训练方法可包括以下步骤：获取样本图像。经人群计数网络对样本图像进行处理，获得第二人群密度图像。依据样本图像与第二人群密度图像之间的差异，获得网络损失。基于网络损失调整人群计数网络的参数。

上述样本图像可以是任意数字图像。例如，样本图像可以包含人物对象，其中，样本图像可以只包括人脸，并无躯干、四肢（下文将躯干和四肢称为人体），也可以只包括人体，不包括人脸，还可以只包括下肢或上肢。本申请对样本图像具体包含的人体区域不做限定。又例如，样本图像可以包含动物。再例如，样本图像可以包含植物。本申请对样本图像中包含的内容不做限定。

经人群计数网络对样本图像的处理获得与样本图像对应的第二人群密度图像后，可依据样本图像与第二人群密度图像之间的差异确定人群计数网络的网络损失。上述差异可以是样本图像与第二人群密度图像中相同位置的像素点的像素值之间的差异。本申请实施例中样本图像中像素点的像素值可用于表征像素点处是否有人物，例如，人物 A 在样本图像中所覆盖的图像区域包含像素点 a，像素点 b，像素点 c，那么像素点 a 的像素值、像素点 b 的像素值和像素点 c 的像素值均为 1。若样本图像中的像素点 d 不属于人物覆盖的图像区域，则像素点的像素值为 0。

在确定人群计数网络的网络损失后，可基于该网络损失通过反向梯度传播的方式调整人群计数网络的参数，直至人群计数网络收敛，完成对人群计数网络的训练。

由于样本图像中的像素点的像素值非 0 即 1，而第二人群密度图像中的像素点的像素值为大于或等于 0 且小于或等于 1 之间的数值。因此，依据用样本图像与第二人群密度图像之间的差异确定人群计数网络的网络损失存在较大的差异。

由于真实人群密度图像中像素点的像素值的取值范围也为大于或等于 0 且小于或等于 1 之间的数值，可选的，可将样本图像的真实人群密度图像作为监督信息，依据真实人群密度图像与第二人群密度图像之间的差异确定人群计数网络的网络损失，以提高获得的网

络损失的精度。

在一种可能实现的方式中，依据脉冲函数、高斯核以及样本图像，可获得上述样本图像的真实人群密度图像。

在该种可能实现的方式中，可依据冲击函数获得样本图像的人物标签图像，该人物标签图像中像素点的像素值用于表征像素点是否属于人物覆盖的图像区域。上述人物标签图像满足下式：

$$H(x) = \sum_{i=1}^N \delta(x - x_i) \cdots \text{公式 (3)}$$

N 为样本图像中的总人数。 x_i 为物体覆盖的图像区域的中心在样本图像中的位置，用于表示该人物。 $\delta(x - x_i)$ 为样本图像中物体覆盖的图像区域的中心在样本图像中的位置的冲击函数。若样本图像中的 x 处有人物，则 $\delta(x)$ 等于 1，若样本图像中的 x 处没有人物，则 $\delta(x)$ 等于 0。

使用高斯核对上述人物标签图像进行卷积处理，可获得样本图像的真实人群密度图像，该过程满足下式：

$$F(x) = \sum_{i=1}^N \delta(x - x_i) * G_{\sigma_i}(x), \text{ 其中, } \sigma_i = \beta d_i \cdots \text{公式 (4)}$$

$$d_i = \frac{1}{m} \sum_{j=1}^m d_j^i \cdots \text{公式 (5)}$$

上述 $G_{\sigma_i}(x)$ 为高斯核， σ_i 为该高斯核的标准差。 β 为正数。 d_i 为距离人物 x_i 最近的 m 个人物与 x_i 之间的距离的平均值。显然， d_i 越大，与 d_i 对应的人物覆盖的图像区域的人群密度也越大。由于样本图像中远处的人物的 d_i 比近处的人物的 d_i 小，通过使高斯核的标准差满足 $\sigma_i = \beta d_i$ ，可使高斯核的标准差与物体覆盖的图像区域的尺度呈正相关，即样本图像中不同图像区域对应的高斯核的标准差不同。这样，通过使用高斯核对样本图像进行卷积处理获得的真实人群密度图像的精确度更高。

举例来说，公式 (3) 中的 x_i 为样本图像中人物的头部覆盖的图像区域的中心（下文将称为人头区域的中心）在样本图像中的位置， $\delta(x - x_i)$ 为样本图像中人头区域的中心的位置的冲击函数。若样本图像中的 x 处有人头，则 $\delta(x)$ 等于 1，若样本图像中的 x 处没有人头，则 $\delta(x)$ 等于 0。基于公式 (4) 使用高斯核对上述人物标签图像进行卷积处理，得到样本图像的真实人群密度图像。对人物标签图像中的第 i 个人头进行卷积处理所使用的高斯核的标准差满足 $\sigma_i = \beta d_i$ ，其中， d_i 为物体标签图像中的第 i 个人头的中心与 m 个目标人头的中心（此处的目标人头为物体标签图像中距离第 i 个人头最近的人头）之间的平均距离，通常情况头部的大小与两个相邻的人在拥挤的场景中的中心之间的距离有关， d_i 在人群较密的情况下近似等于人头大小。由于物体标签图像中“近”处的人头覆盖的图像区域的面积比“远”出的人头覆盖的图像区域的面积大，也就是说，物体标签图像中“近”处的两个人头的中心之间的距离比“远”出的两个人头的中心之间的距离大，通过使高斯核的标准差满足 $\sigma_i = \beta d_i$ ，可达到使高斯核的标准差与人物的头部覆盖的图像区域的尺度呈正相关的效果。

在获得样本图像的真实人群密度图像后，可依据真实人群密度图像中与第二人群密度图像中相同位置的像素点的像素值之间的差异，确定人群计数网络的网络损失。例如将真实人群密度图像中与第二人群密度图像中所有的相同位置的像素点的像素值之间的差异的和作为人群计数网络的网络损失。

可选的，在将样本图像输入至人群计数网络之前，可对样本图像进行预处理，获得至少一张预处理后的图像，并将上述至少一张预处理后的图像作为训练数据输入至人群计数网络。这样，可达到扩充人群计数网络的训练数据集的效果。

上述预处理包括从样本图像中截取预定尺寸的图像、对样本图像或所述预定尺寸的图像进行翻转处理中的至少一种。其中，预定大小可以为 $64*64$ 。对样本图像进行翻转处理

包括：水平镜面翻转处理。

例如，分别沿样本图像的水平中轴线和竖直中轴线对样本图像进行划分，可获得4张预处理后的图像。同时从样本图像中随机截取5张预定尺寸的图像，可获得5张预处理后的图像。至此，已获得9张预处理后的图像。对该9张预处理后的图像进行水平镜面翻转处理，可获得9张翻转后的图像，即另外9张预处理后的图像。这样即可获得18张预处理后的图像。

通过将至少一张预处理后的图像输入至人群计数网络，可获得至少一张第三人群密度图像，其中，每一张预处理后的图像均对应有一张第三人群密度图像。例如（例2），将图像A、图像B、图像C这3张预处理后的图像分别输入至人群计数网络，将分别获得与图像A对应的人群密度图像a，与图像B对应的人群密度图像b，图像C对应的人群密度图像c。其中，人群密度图像a、人群密度图像b、人群密度图像c均可称为第三人群密度图像。

依据至少一张预处理后的图像中的目标图像和与目标图像对应的第三人群密度图像之间的差异，可获得人群计数网络的网络损失。接着例2继续举例，依据图像A与图像a之间的差异可获得第一差异，依据图像B与图像b之间的差异可获得第二差异，依据图像C与图像c之间的差异可获得第三差异。对第一差异、第二差异和第三差异求和可获得人群计数网络的网络损失。

本实施例提供了一种人群计数网络，使用该人群计数网络对待处理图像进行处理，可获得与待处理图像对应的人群密度图像，进而可确定待处理图像中的人数。

基于本申请实施例提供的技术方案，本申请实施例还提供了几种可能实现的应用场景：

场景A：如上所述，在公共场所常因人流量过多导致人群过于密集的情况的发生，进而发生一些公共事故，如何对公共场所进行人群计数就具有非常大的意义。

目前，为了增强工作、生活或者社会环境中的安全性，会在各个公共场所内安装监控摄像设备，以便根据视频流信息进行安全防护。利用本申请实施例提供的技术方案对监控摄像设备采集到的视频流进行处理，可确定公共场所的人数，进而可有效预防公共事故的发生。

举例来说，监控摄像设备的视频流处理中心的服务器可执行本申请实施例提供的技术方案，该服务器可与至少一个监控摄像头相连。服务器在获取到监控摄像头发送的视频流后，可采用本申请实施例提供的技术方案对视频流中的每一帧图像进行处理，以确定视频流中的每一帧图像中的人数。在图像中的人数大于或等于人数阈值的情况下，服务器可向相关设备发送指令，以进行提示或报警。例如，服务器可向采集该图像的摄像头发送指令，该指令用于指示采集该图像的摄像头进行报警。又例如，服务器可向采集该图像的摄像头所在的区域的管控人员的终端发送指令，该指令用于提示该终端输出人数超过人数阈值的提示信息。

场景B：商场中不同区域的人流量不同，将主推商品放置于人流量多的区域进行展示可有效提高主推商品的销量，因此，如何准确确定商场不同区域的人流量对商家来说具有非常重要的意义。例如，商场中有区域A、区域B和区域C，其中区域B的人流量最大，基于此，商家可将主推商品放置于区域B进行展示，以提高主推商品的销量。

商场的监控摄像头的视频流的管控中心的服务器可执行本申请实施例提供的技术方案，该服务器可与至少一个监控摄像头相连。服务器在获取到监控摄像头发送的视频流后，可采用本申请实施例提供的技术方案对视频流中的每一帧图像进行处理，以确定视频流中的每一帧图像中的人数。依据每一帧图像中的人数可确定不同摄像头监控的区域在某一时间段内的人流量，进而可确定商场内的不同区域的人流量。例如，商场中有区域A、区域B、区域C，摄像头A、摄像头B和摄像头C，其中，摄像头A监控区域A，摄像头B监控区域B，摄像头C监控区域C。服务器使用本申请实施例提供的技术方案对摄像头A采集到的视频流中的图像进行处理，确定区域A在过去一个星期内平均每天的人流量为900，确定区域B在过去一个星期内平均每天的人流量为200，确定区域C在过去一个星期内平均每天的人流量为600。显然，区域A的人流量最多，因此商家可将主推商品放置于区域A内进行展示，以提高主推商品的销量。

本领域技术人员可以理解，在具体实施方式的上述方法中，各步骤的撰写顺序并不意味着严格的执行顺序而对实施过程构成任何限定，各步骤的具体执行顺序应当以其功能和可能的内在逻辑确定。

上述详细阐述了本申请实施例的方法，下面提供了本申请实施例的装置。

请参阅图10，图10为本申请实施例提供了一种图像处理装置的结构示意图，该装置1包括：获取单元11、卷积处理单元12、融合处理单元13、特征提取处理单元14、第一确定单元15、第二确定单元16以及训练单元17。其中：

获取单元11，用于获取待处理图像、第一卷积核和第二卷积核，所述第一卷积核的感受野与所述第二卷积核的感受野不同；

卷积处理单元 12, 用于使用所述第一卷积核对所述待处理图像进行卷积处理获得第一特征图像, 使用所述第二卷积核对所述待处理图像进行卷积处理获得第二特征图像;

融合处理单元 13, 用于对所述第一特征图像和所述第二特征图像进行融合处理, 获得第一人群密度图像。

在一种可能实现的方式中, 所述装置 1 还包括:

特征提取处理单元 14, 用于在所述对所述第一特征图像和所述第二特征图像进行融合处理, 获得第一人群密度图像之前, 对所述待处理图像进行第一特征提取处理, 获得第一自注意力图像, 对所述待处理图像进行第二特征提取处理, 获得第二自注意力图像, 所述第一自注意力图像和所述第二自注意力图像均用于表征所述待处理图像的尺度信息, 且所述第一自注意力图像所表征的尺度信息与所述第二自注意力图像所表征的尺度信息不同;

第一确定单元 15, 用于依据所述第一自注意力图像确定所述第一特征图像的第一权重, 依据所述第二自注意力图像确定所述第二特征图像的第二权重;

所述融合处理单元 13 用于:

依据所述第一权重和所述第二权重对所述第一特征图像和所述第二特征图像进行融合处理, 获得所述第一人群密度图像。

在另一种可能实现的方式中, 所述融合处理单元 13 具体用于:

确定所述第一权重与所述第一特征图像之间的点积, 获得第三特征图像;

确定所述第二权重与所述第二特征图像之间的点积, 获得第四特征图像;

对所述第三特征图像和所述第四特征图像进行融合处理, 获得所述第一人群密度图像。

在又一种可能实现的方式中, 所述第一确定单元 15 用于:

对所述第一自注意力图像和所述第二自注意力图像进行归一化处理, 获得所述第一自注意力图像对应的第三自注意力图像和所述第二自注意力图像对应的第四自注意力图像;

将所述第三自注意力图像作为所述第一权重, 将所述第四自注意力图像作为所述第二权重。

在又一种可能实现的方式中, 所述特征提取处理单元 14, 还用于在所述使用所述第一卷积核对所述待处理图像进行卷积处理获得第一特征图像, 使用所述第二卷积核对所述待处理图像进行卷积处理获得第二特征图像之前, 对所述待处理图像进行第三特征提取处理, 获得第五特征图像;

所述卷积处理单元 12 用于:

使用所述第一卷积核对所述第五特征图像进行卷积处理获得所述第一特征图像, 使用所述第二卷积核对所述第五特征图像进行卷积处理获得所述第二特征图像;

所述特征提取处理单元 14 还用于:

对所述第五特征图像进行所述第一特征提取处理, 获得所述第一自注意力图像, 对所述第五特征图像进行所述第二特征提取处理, 获得所述第二自注意力图像。

在又一种可能实现的方式中, 所述第一卷积核和所述第二卷积核均为空洞卷积核, 且所述第一卷积核的大小与所述第二卷积核的大小相同, 且所述第一卷积核的权重与所述第二卷积核的权重相同, 且所述第一卷积核的扩张率与所述第二卷积核的扩张率不同。

在又一种可能实现的方式中, 所述第一卷积核或所述第二卷积核的扩张率为参考值。

在又一种可能实现的方式中, 所述装置 1 还包括: 第二确定单元 16, 用于确定所述第一人群密度图像中的像素值的和, 获得所述待处理图像中的人数。

在又一种可能实现的方式中, 所述装置 1 执行的图像处理方法应用于人群计数网络;

所述装置 1 还包括: 训练单元 17, 用于对所述人群计数网络进行训练, 所述人群计数网络的训练过程包括:

获取样本图像;

使用所述人群计数网络对所述样本图像进行处理, 获得第二人群密度图像;

依据所述样本图像与所述第二人群密度图像之间的差异, 获得网络损失;

基于所述网络损失调整所述人群计数网络的参数。

在又一种可能实现的方式中, 所述训练单元 17 还用于:

在所述依据所述样本图像与所述第二人群密度图像之间的差异, 获得网络损失之前, 依据冲击函数、高斯核以及所述样本图像, 获得所述样本图像的真实人群密度图像;

依据所述真实人群密度图像与所述第二人群密度图像之间的差异, 获得所述网络损失。

在又一种可能实现的方式中, 所述训练单元 17 还用于:

在所述经所述人群计数网络对所述样本图像进行处理, 获得第二人群密度图像之前, 对所述样本图像进行预处理, 获得至少一张预处理后的图像;

使用所述人群计数网络对所述至少一张预处理后的图像进行处理, 获得至少一张第三人群密度图像, 所述预处理后的图像与所述第三人群密度图像一一对应;

依据所述至少一张预处理后的图像中的目标图像和与所述目标图像对应的第三人群密度图像之间的差异, 获得所述网络损失。

在又一种可能实现的方式中，所述预处理包括：从所述样本图像中截取预定尺寸的图像、对所述样本图像或所述预定尺寸的图像进行翻转处理中的至少一种。

本实施例通过使用感受野不同的第一卷积核和第二卷积核分别对待处理图像进行卷积处理，以提取出不同尺度下的描述待处理图像的内容的信息，分别获得第一特征图像和第二特征图像。通过对第一特征图像和第二特征图像进行融合处理，以利用不同尺度下的描述待处理图像的内容的信息，提高获得的与待处理图像对应的人群密度图像的精度，进而提升获得的待处理图像中人数的精度。

在一些实施例中，本公开实施例提供的装置具有的功能或包含的模块可以用于执行上文方法实施例描述的方法，其具体实现可以参照上文方法实施例的描述，为了简洁，这里不再赘述。

图 11 为本申请实施例提供的一种图像处理装置的硬件结构示意图。该图像处理装置 2 包括处理器 21，存储器 22，还可以包括输入装置 23，输出装置 24。该处理器 21、存储器 22、输入装置 23 和输出装置 24 通过连接器相耦合，该连接器包括各类接口、传输线或总线等等，本申请实施例对此不作限定。应当理解，本申请的各个实施例中，耦合是指通过特定方式的相互联系，包括直接相连或者通过其他设备间接相连，例如可以通过各类接口、传输线、总线等相连。

处理器 21 可以是一个或多个图形处理器（graphics processing unit，GPU），在处理器 21 是一个 GPU 的情况下，该 GPU 可以是单核 GPU，也可以是多核 GPU。可选的，处理器 21 可以是多个 GPU 构成的处理器组，多个处理器之间通过一个或多个总线彼此耦合。可选的，该处理器还可以为其他类型的处理器等等，本申请实施例不作限定。

存储器 22 可用于存储计算机程序指令，以及用于执行本申请方案的程序代码在内的各类计算机程序代码。可选地，存储器包括但不限于是随机存储记忆体（random access memory，RAM）、只读存储器（read-only memory，ROM）、可擦除可编程只读存储器（erasable programmable read only memory，EPROM）、或便携式只读存储器（compact disc read-only memory，CD-ROM），该存储器用于相关指令及数据。

输入装置 23 用于输入数据和信号，以及输出装置 24 用于输出数据和信号。输入装置 23 和输出装置 24 可以是独立的器件，也可以是一个整体的器件。

可理解，本申请实施例中，存储器 22 不仅可用于存储相关指令，还可用于存储相关图像，如该存储器 22 可用于存储通过输入装置 23 获取的待处理图像，又或者该存储器 22 还可用于存储通过处理器 21 获得的第一人群密度图像等等，本申请实施例对于该存储器中具体所存储的数据不作限定。

可以理解的是，图 11 仅仅示出了图像处理装置的简化设计。在实际应用中，图像处理装置还可以分别包含必要的其他元件，包括但不限于任意数量的输入/输出装置、处理器、存储器等，而所有可以实现本申请实施例的图像处理装置都在本申请的保护范围之内。

本申请实施例还提供了一种处理器，该处理器的缓存中可存储计算机程序，当该计算机程序被该处理器执行时，该处理器可执行实施例（一）和实施例（二）所提供的技术方案、或实现已训练的人群计数网络对待处理图像的处理。

本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的系统、装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。所属领域的技术人员还可以清楚地了解到，本申请各个实施例描述各有侧重，为描述的方便和简洁，相同或类似的部分在不同实施例中可能没有赘述，因此，在某一实施例未描述或未详细描述的部分可以参见其他实施例的记载。

在本申请所提供的几个实施例中，应该理解到，所揭露的系统、装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

另外，在本申请各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。

在上述实施例中，可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时，可以全部或部分地以计算机程序产品的形式实现。所述计算机程序产品包括一个或多个计算机指令。在计算机上加载和执行所述计算机程序指令时，全部或部分地产生按照本申请实施例所述的流程或功能。所述计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。所述计算机指令可以存储在计算机可读存储介质中，或者通过所述计算机可读存储介质进行传输。所述计算机指令可以从一个网站站点、计算机、服务器或数据中心通过有线（例如同轴电缆、光纤、数字用户线（digital subscriber line, DSL））或无线（例如红外、无线、微波等）方式向另一个网站站点、计算机、服务器或数据中心进行传输。所述计算机可读存储介质可以是计算机能够存取的任何可用介质或者是包含一个或多个可用介质集成的服务器、数据中心等数据存储设备。所述可用介质可以是磁性介质，（例如，软盘、硬盘、磁带）、光介质（例如，数字通用光盘（digital versatile disc, DVD））、或者半导体介质（例如固态硬盘（solid state disk, SSD））等。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，该流程可以由计算机程序来指令相关的硬件完成，该程序可存储于易失性和非易失性计算机可读存储介质中，该程序在执行时，可包括如上述各方法实施例的流程。而前述的存储介质包括：只读存储器（read-only memory, ROM）或随机存取存储器（random access memory, RAM）、磁碟或者光盘等各种可存储程序代码的介质。