可加的超分模块

Adaptive Super-Resolution for Person Re-Identifification

with Low-Resolution Images

Abstract

使用低分辨率查询图像和高分辨率画廊图像对人员进行re- Id具有挑战性。为了解决分辨率不匹配的问题，许多方法通过指定单个比例因子对低分辨率查询执行超分辨率（SR）。但是，**使用单个SR模块** （无论指定了哪个比例因子），在恢复和标识身份信息时总是带来利弊。**较大的比例因子可恢复更多细节，但会产生过多的伪像，而较小的则相反**。为了利用它们的**互补特性**来进行更健壮的恢复和识别，我们提出了自适应行人超分辨率（APSR）模型。 APSR根据它们生成的视觉内容联合训练和融合多个SR模块，以端到端的方式完全补偿和学习互补的身份特征。为了提高融合过程中对伪影的鲁棒性，我们的模型通过在线划分和整合生成的身体区域进一步学习了富含信息特征。大量的实验证明了我们方法的最新性能。

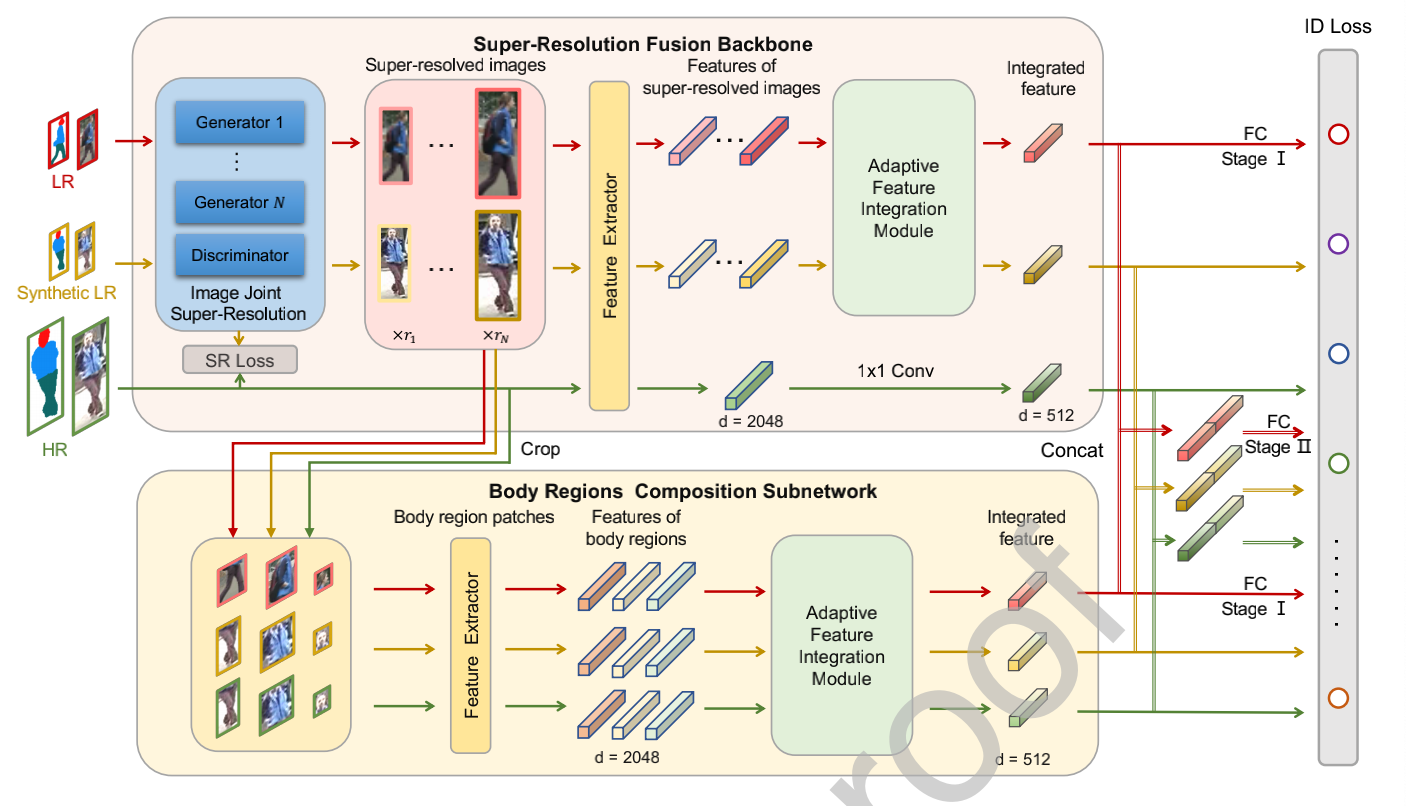
**Introduction**

如图1所示，**较大的超分辨率图像通常可以恢复更多的外观细节，例如颜色，纹理和徽标**，这对识别非常有帮助。但是，由于产生了太多的尖锐或嘈杂的伪影，例如人体区域失真和背景噪声，它们也对识别产生不利影响。**尽管较小的超分辨率图像包含较少的伪像，并且可以更好地保留原始信息，但是它们无法恢复所需的高频细节**。一个简单的主意是**联合使用具有不同比例因子的多个SR模块对LR图像进行超分辨**，然后正确融合所有这些超分辨结果，但是如何在融合过程中平衡它们的相对重要性却是一个挑战。

另外，仅在超分辨图像之间执行融合可能是不够的，因为**较大的图像通常同时包含更多恢复良好的细节和不想要的伪影**，在融合过程中始终会对其进行平均加权。例如，在图1中，query A的紫色衬衫和蓝色裤子的纹理在×4超分辨图像中很好地合成，而头部严重扭曲。 一些恢复良好的身体区域有助于重新识别，而失真的区域和背景噪音则无助，甚至会大大降低最终性能。 **因此，合理的做法是使不同的恢复区域对识别具有不同的重要性**。 **我们应该更多地利用那些信息丰富的区域而不是伪像**，尤其是对于最大的超分辨率图像。 但是，如何在融合多个SR模块的情况下实现此目标尚未探索。

为了解决上述问题，我们为LR re-ID提出了一种新颖的自适应行人超分辨率（APSR）模型。这是一个高效的端到端网络，可以根据生成的视觉内容联合训练和融合具有不同比例因子的多个SR模块。为了提高不同SR模块的完备性，**我们提出了端到端自适应特征集成模块**，**自动平衡不同超分辨内容的相对重要性**。该模块可以自适应地为每个输入特征向量提供预测的权重向量，该权重向量由所有输入特征通过它们的交互共同确定。考虑到较大的超分辨图像同时包含恢复良好的细节和许多伪像，因此我们还学习了对伪像具有鲁棒性的判别性noise-free特征。为此，我们将**最大的超分辨图像划分为多个语义主体区域**，即头部，上半身和下半身，然后再次使用自适应特征集成模块将它们在线集成到信息丰富的特征中。最后，多分辨率全局特征和无噪声局部特征构成了更强大的身份表示，可以有效缩小LR和HR图像之间的分辨率差距。结果，我们的模型在四个可公开获得的LR re-ID数据集（包括MLR-VIPeR，MLR-SYSU，MLR-CUHK03和CAVIAR）上超越了现有的最新技术。

**Methods**



为了解决使用单个SR模块进行检测恢复和识别的问题，我们提出了用于LR re-ID的自适应person超分辨率（APSR）模型。 如图2所示，**它由两个主要部分组成**。为了实现不同SR模块在恢复身份信息方面的互补优势，超分辨率融合主干根据生成的视觉内容**自适应融合**具有不同比例因子的**SR模块**。 为了在融合过程中进一步提高对伪像的鲁棒性，**人体区域成分子网络**通过在线划分和自动整合生成的人体区域来学习信息特征。请注意，这两个部分紧密相关。**子网是从**在网络上生成超分辨身体区域的**主干网发展而来的**，身体区域的组成成分可以帮助抑制伪像的产生。在以下各节中，我们将详细介绍这两个部分。

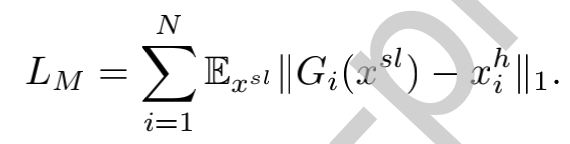
**Super-Resolution Fusion for LR re-ID**

Image joint super-resolution

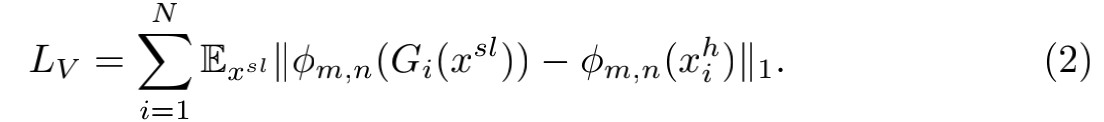
给定LR图像，我们旨在将其放大为多个高分辨率图像，并自适应地整合它们的互补特征。 为此，我们首先设计一个图像联合超分辨率模块，以并行恢复多个超分辨图像。

任何典型的SR模块，例如SRCNN [5]，SRGAN [9]和ESRGAN [3]，都可以用作图像联合超分辨率的基本生成器。在这里，我们使用**ESRGAN**是因为它在生成逼真的图像方面具有竞争优势。图像联合超分辨率模块由**N个基本生成器Gi（**i = 1、2，···，N）组成，其上采样比例因子为{r1，r2，···，rN}和**共享的判别器D**。受[1]的启发，我们的模型采用从训练集中随机选择的一组图像作为输入，即HR，LR和由HR降采样的合成LR图像。我们将它们分别表示为{xh，xl，xsl}。图像集的每个图像具有相同的标识，但是扮演不同的角色。xsl和xh创建了一对用于训练SR模块的输入（LR）和标签（HR），而xl只是通过SR模块向前传递，不需要相应的HR ground truth（在实践中也不可用）。 xsl对于使SR模块在监督下可学习，并使SR和re-ID两个原本相互独立的任务共同桥接和培训至关重要。注意，合成LR图像xsl从HR以最大比例因子**rN**向下采样得到。这可以确保每个超分辨图像具有对应的ground truth，**这是通过将xh下采样到相应大小而获得的，**因此可以在监督下训练每个SR模块Gi。例如，如果xh的大小为H∗W（H和W分别为xh的高度和宽度），并且xsl为1/rN H ∗ 1/rNW，则当我们通过比例因子ri将xsl超解析为Gi（xsl）时，**则通过将xh下采样至ri /rN H ∗ ri /rN W的大小来获取相应的地面实况xhi。**

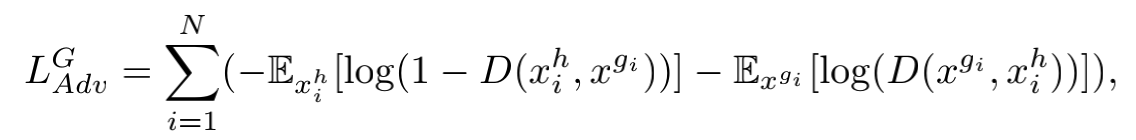
生成器Gi输出xl和xsl的超分辨图像Gi（xl），Gi（x sl），以混淆区分图像是真实图像还是伪图像的鉴别器D。 一些损失函数被用来训练图像联合超分辨率模块。 首先，为了使超分辨图像在像素水平上接近ground truth，我们利用均方误差（MSE）损失LM：

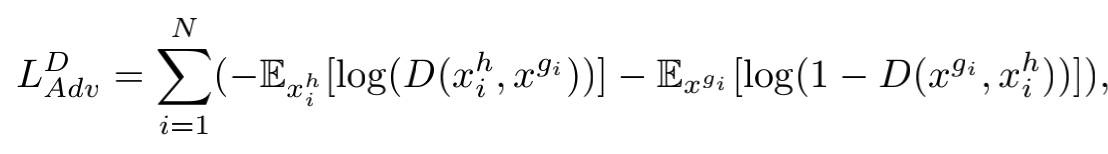


但是，仅在训练过程中最小化MSE损失LM会导致图像过于平滑，而缺少高频细节。因此，我们在特征图上进一步添加了VGG损失，以帮助生成感知上更精细的纹理。VGG损失计算超分辨图像和ground truth的特征图之间的1范数距离，这些特征图是从预先训练的VGG19 [27]网络中提取的。 在这里，我们将φm，n表示为VGG19的第m个最大池化层之前的第n个卷积的特征图（本文中m，n分别设置为5,4），并且将VGG损耗LV计算为：

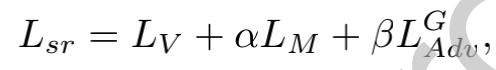


除了上述针对图像内容设计的损失外，当生成器和鉴别器玩游戏时，我们还利用对抗性损失。 在ESRGAN之后，我们采用相对应的生成器对抗损失LG Adv和鉴别器对抗损失LD Adv。 我们将它们表示为：





总的超分辨率损失：

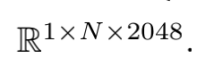


**Adaptive feature integration module**

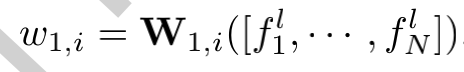
多个生成的图像可恢复冗余和唯一身份信息。为了利用他们的互补性得到原始LR图像的信息量更多的特征表示，我们通过自适应特征集成模块在它们之间建立了基于特征的融合。

对于超分辨图像G1（xl），···，GN（xl），G1（xsl），···，GN（xsl）（分别表示为xlg1，···，xlgN，xsl g1，xsl gN，为简化起见），我们通过ResNet50 [28]特征提取器提取其特征。 除去ResNet50的完全连接层，我们在全局平均池之后将2048维特征向量作为输出，并将它们表示为{fl1，···，flN，fsl1，···，fsl N }作为自适应特征集成模块的输入。如图3所示，我们的自适应特征集成模块构建了从多个单独的输入特征向量到更具信息量向量的映射的映射，这可以通过以下两个分支来实现。

**Feature interaction branch**

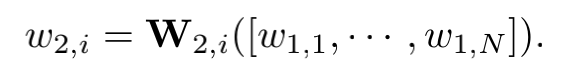
在此分支中，我们评估每个特征向量的作用，并通过每个特征的相互作用预测具有相同维度的权重向量。首先将不同的SR图片特征向量整合成一个tensor

然后，它同时经过N个具有1×N内核的独立卷积层，从而为每个输入fli（i = 1，...，N）生成一个粗略的权重向量w1，i，公式为：



W1，i是粗略权重预测的第i个卷积层。请注意，我们通过级联特征并使用1×N内核来构建此类交互，以便所有特征向量可以共同决定每个特征向量的权重。容易理解，我们可以通过参考其他特征并进行比较来更好地评估特征向量的重要性。 我们还通过实验发现，如果没有与其他对象的交互作用，特征向量就无法完全靠自身很好地判断其权重。

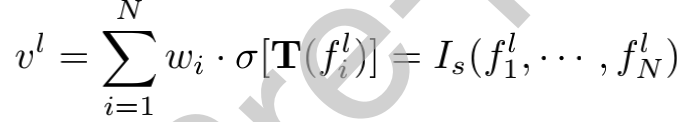
接下来，我们设计一个类似的级联和卷积过程以细化权重，公式如下：



W2，i是权重细化的第i个卷积层，而w2，i是fli的细化权重向量。 最后，我们沿每个维度在w2，i（i = 1，...，N）上使用softmax函数，从而形成特征向量fli的最终权重向量wi。 它反映了**每个特征向量每个维度对最终表示的重要性**。 实验表明，在粗略权重预测的基础上，权重细化进一步提高了输出特征的集成能力。 粗略和细化的权重向量的维数均为d，在本文中设置为512。

**Transformation and integration branch.**

在另一个分支中，输入特征向量将进一步转换以与预测的权重匹配。我们使用1×1内核设计共享卷积层，其维数也为d，以便我们可以统一特征向量和权重向量的维数。 我们在1×1卷积层后面用Sigmoid型函数以获得变换后的特征向量，最终的集成特征vl是所有变换后特征的加权和。 我们将其表示为：



其中σ是S形函数，T是转换层，vl是xl的整合特征。 我们将从单独的输入管道fl1，...，flN到输出vl总结为Is。 类似地，xsl的集成特征可以表示为vsl = Is（fsl1，...，f sl N）。

请注意，上面的图像超分辨率和特征集成过程是针对xl和xsl执行的，因此我们可以通过利用多个超解析版本的互补属性来完全补偿低分辨率，以弥补其丢失的信息。 相反，HR图像xh是内在的信息丰富的。我们通过与xl和xsl共享的特征提取器提取其特征向量fh，并使用独立的1×1卷积层将其直接投影到d维特征向量vh中。 在d维特征空间中，我们测量LR图像的集成特征与HR图像的直接投影特征之间的相似性。

**Discussion**

一些现有的集成超分辨率方法[7，8]也尝试将不同模块的超分辨结果融合在一起，但是我们提出的自适应特征集成模块在以下几个方面有明显的区别。

1）**目标**。这些集成方法的目标是将多个估计的高分辨率图像作为输入，将整体图像作为输出，以像素级复原，而我们的模块则用于特征级融合，将多个特征矢量映射为一个更强大的矢量。因此，它们的方法不能应用于我们的问题。

1. **权重预测**。尽管上述所有方法都可以视为融合策略，但是它们在权重预测方面有很大不同。两种集成方法都是手工制作的权重预测器，具体取决于额外的参考数据集[7]或下采样图像[8]，它们的适应性较弱，并且倾向于局部最优。相反，我们的模块可以通过端对端方式通过输入特征向量之间的交互来自动预测权重。此外，通过SR模块生成图像和权重预测在它们的方法中是分开的，而我们的端到端网络则经过共同培训以实现更好的兼容性。
2. **细粒度的权重**。产生的权重[7，8]是一个实数，并由整个图像的所有像素共享，而我们的权重是一个实值向量，指示特征向量的逐元素权重。因此，我们的重量具有更好的粒度和更高的精度。

*Body regions composition*

如上所述，当我们提取超分辨图像的特征时，全局平均池几乎平等地权衡了信息细节和伪像。 但是，伪影对识别具有负面影响，随着比例因子的增加，伪影会变得更加严重。 因此，合理的做法是使不同区域的识别具有不同的重要性，特别是对于较大的超分辨图像。 为此，我们提出了人体区域组成子网络，**以学习对伪像具有鲁棒性的无噪声特征**，以获得最大的超分辨图像。

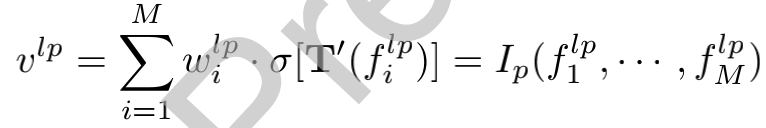
**Semantically-aligned body regions**

在这个子网中，我们利用了human parsing masks，它是通过最新的人类语义解析模型JPPNet [30]获得的，作为额外的监督。 这样做的动机是，我们发现伪影通常与语义区域相关，并且具有区域语义特征（例如，头部变形），而不是整个图像上的随机像素，如图1所示。使用解析mask使我们能够丢弃某些背景并在图像中对齐语义身体区域，并指导我们的模型从信息性语义区域（而不是伪像）中学习更多区分性特征。否则，如果没有语义指导，就很难自动确定整个图像的哪些部分对识别更有帮助。

JPPNet最初获得的mask指示背景和19个身体区域，它们组合成M个均匀区域（M = 3，默认情况下包括头部，上半身和下半身）。注意，所获得的身体区域的适当组合是不平凡的，因为许多原始区域（例如，手臂和鞋子）非常小，以至于它们对最终特征的贡献很小，但增加了计算负担。然后，我们使用mask获取分别包含M个身体区域的最小矩形图像块。具体来说，我们将JPPNet直接应用于HR图像xh并裁剪图像补丁，而对于LR图像xl和xsl，我们将其mask通过因子rN放大，并使用放大的mask裁剪最大的超分辨图像xlgN和xsl gN。在这里，将人类解析模型应用于原始LR而不是SR图像，原因有两个：**首先，SR图像可能有过多的伪像，从而导致解析mask变差。**其次，通过与APSR网络分离的JPPNet模型离线解析LR图像以获取mask，从而节省了计算时间和空间，而在线生成的超分辨率图像则无法实现这一点。

**Body regions automatic composition**

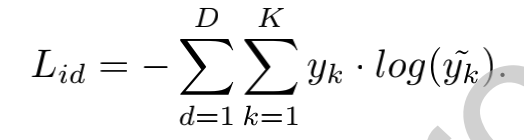
为了**学习最大的超分辨率图像的无噪声特征**，我们再次使用自适应特征集成模块来自动最大化信息区域的使用并最小化伪像。与3.1.2中的超分辨图像的处理过程类似，我们首先通过ResNet50特征提取器提取最大的超分辨图像xlgN和xsl gN或HR图像xh的每个身体区域补丁的特征。以xlgN为例，我们可以得到分别代表M个身体区域的M个2048维特征向量flp i（i = 1，...，M）。然后，将这些特征发送到自适应特征集成模块中，该模块与超分辨率融合主干中的相同，但有细微差别。超分辨图像的特征相似，并且容易在特征转换分支中共享卷积层，但是很难在使用共享层同时处理不同的身体区域。因此，人体区域的特征经过M 个1×1卷积层，分别由M个人体区域负责。类似于公式8，我们可以将上述过程公式表示为：



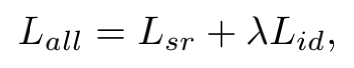
其中wlpi是特征交互分支为特征向量flpi计算的权重向量，而T‘是变换层。 Ip描述了整个特征集成过程，从人体区域flpi（i = 1，...，M）的单独特征向量到集成的无噪声特征向量vlp。 类似地，我们可以分别为xsl gN和xh获取整合的特征向量vslp和vhp。 请注意，身体区域是在线生成，划分和集成的，**整个模型的联合优化使每个部分相互促进**。 SR模块的良好恢复能力有助于有效地提取特征和整合身体区域，**而整合身体区域则推动SR模块恢复更多的区分细节和更少的伪像。** 可视化结果对此进行了验证。

*Overall loss*

获取了整个图像（vl，vsl或vh）或身体区域（vlp，vslp或vhp）的最终特征后，我们将其发送到分类器（即全连接层）中以预测身份。 我们提出了一个两阶段的培训策略。 在第一阶段，将整个图像和身体区域的特征分别发送到两个分类器中，以便对它们进行训练，使其在全局和局部两个级别上具有区分性。考虑到单独训练的全局和局部特征可能不是最优的，我们将它们串联为统一的特征，并在阶段II由另一个分类器对它们进行微调， 在测试时，我们还使用串联的特征。分类器通过交叉熵恒等损失Lid优化，公式为：



K和D分别是训练集中的身份和样本总数。 yk是样本的二进制标签：1或0分别表示真或假标识标签。 〜yk是相应样本属于第k类的预测概率。 整个网络由超分辨率损失Lsr和身份损失Lid的加权和来训练。 总体损失Lall可以写成：

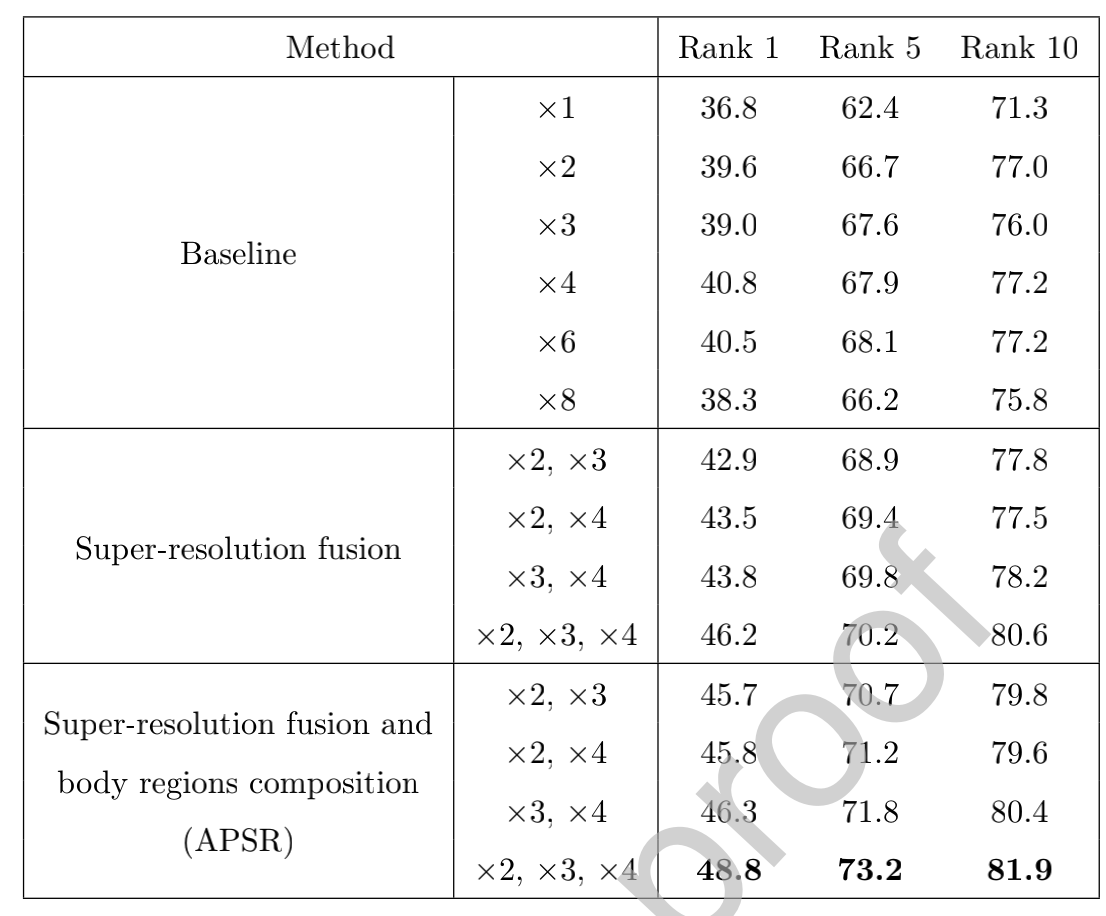


其中λ是平衡因子。 由于我们的研究重点是SR和re-ID之间的关系，而不是SR本身，因此我们在公式5中固定Lsr的内部权重。根据ESRGAN，我们将Lsr的α设置为0.01，将β设置为0.005。 总损失桥接了两个最初独立的SR和re-ID任务，并指导多个SR模块在联合优化过程中恢复以re-ID为导向的辨别性。

**Experiment**

**Effffect analysis of the super-resolution fusion and body regions composition.**

在本节中，我们通过定量和可视化结果验证超分辨率融合和人体区域组成的有效性。我们设置了几个基线，每个基线仅包含一个单独的SR模块（×1，×2，×3，×4，×6或×8），而没有SR模块或身体部位的融合。如表1所示，我们在MLR-VIPeR上评估他们的等级1、5和10。 ×1表示保持LR图像的原始分辨率不变，并且不使用任何SR模块进行超分辨率。 ×1基线的得分明显低于使用SR模块的基线，这表明超分辨率是解决LR re-ID中分辨率不匹配问题的有效方法。我们在最终版本的APSR中使用×2，×3和×4 SR模块的主要原因有两个。首先，这三个比例因子具有较高的性能，分别为×1和×8。尽管×6可以达到与×4相当的结果，尤其是在等级5处，但由于生成更多的像素值，因此会导致较高的计算负担。其次，这与我们利用不同比例因子的互补特性的动机是一致的：较小的比例因子×2可以更好地保留原始信息，而较大的比例因子×4具有恢复更多高频细节的优势。在以下实验中，我们不再讨论×1，×6或×8，而是重点分析×2，×3和×4 SR模块。



我们首先通过从APSR中删除人体区域组成子网络来验证超分辨率融合的个体效果。 表1中的实验结果表明，两次SR融合总是优于两个相应的一次SR基线。 例如，当我们融合×2和×3 SR模块时，排名1的分数从39.6％或39.0％上升至42.9％。 此外，额外融合×4 SR模块使等级1达到46.2％。 超分辨率融合的优势在于，它可以使不同SR模块在恢复方面具有更好的互补性，在re-ID和多个SR模块在识别方面具有更好的兼容性。 通过自适应功能集成进行的端到端培训使每个部分相互影响并相互适应，以实现更强大的恢复和识别。 我们的模型还显示了可扩展性，因为增加SR模块的数量会在一定程度上带来更好的re-ID性能。

然后，我们添加身体区域组成子网络（这意味着我们评估整个APSR模型）并证明其有效性。 从表1中，我们可以很容易地观察到，对于SR模块的所有组合，在各个等级上，身体部位的组成都进一步明显改善了结果。 例如，通过将人体区域组成添加到×2，×3和×4 SR模块的融合中，等级1/5/10得分分别增加了2.6％/ 3.0％/ 1.3％。 这表明身体部位的组成可以在语义信息的引导下自动整合可辨别的局部特征，从而学会从文物中区分信息丰富的细节。

我们还直观地比较了图4中两个MLR-VIPeR查询的排名前8位的图库图像。比较三个基线的结果，我们发现×4基线使查询A的地面真理达到了最佳结果（第四个） ），而查询B的基本真值最多以×2基线排在第三位。 我们可以看到不同的LR查询具有不同的SR模块可实现更好的单SR性能。 很难预测哪个SR模块更适合查询，但是我们的超分辨率融合可以利用不同SR模块的互补性来学习更强大的功能，并获得比每个单个SR模块更好的性能。 当我们进一步将子网添加到超分辨率融合主干中时，两个查询的两个基本事实都上升到了第一位，这表明主体区域组成进一步提高了模型的鲁棒性。



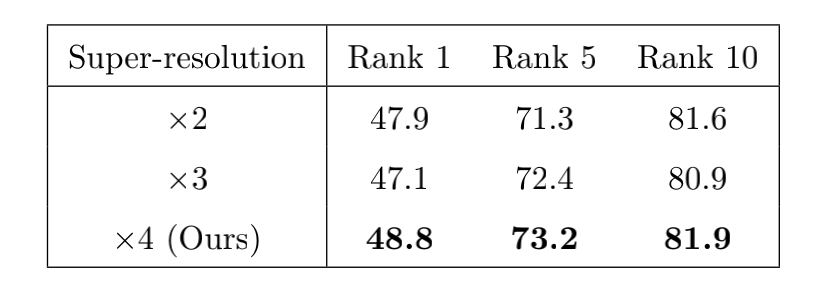
**Visualization of super-resolved images**

通过图5中的不同方法恢复的图像还显示了超分辨率融合和人体区域组成的影响。尽管超分辨率融合的排序结果比表1中的基线更好，但它可以恢复带有更多伪像的细节，例如带有黄色圆圈的嘈杂的头部，手臂和背景。这可能是因为超分辨率融合过于关注SR模块之间的适应性，而忽略了恢复可识别外观的re-ID要求。相比之下，APSR生成视觉上更好的面向re-ID的图像：外观更自然，伪像更少，背景更平滑，并且不会严重降低性能。**原因在于，通过端到端的培训，身体区域的在线集成引导生成者将更多的精力集中在身体区域细节的恢复上，从而增强了身份信息恢复与识别之间的一致性。**



**Super-resolution for body regions composition**

为了研究执行我们的身体区域合成的哪种超分辨图像更有效，我们分别使用从×2，×3和×4超分辨图像捕获的身体区域对MLR-VIPeR进行了实验。 表2所示的实验结果表明，×4实现了最佳性能。 这是因为较大的超分辨图像在恢复丢失的高频方面具有优势



**Balance factor**

我们评估了平衡因子λ的影响，该影响决定了等式11中超分辨率损失Lsr和同一性损失Lid之间的关系。图6绘制了MLR-VIPeR和MLR-SYSU上随λ变化的等级1分数。 我们观察到，当λ设置为1时，秩1达到峰值，我们的模型对λ∈（0，1]更为稳健。将λ设置为0（即去除Lid）会大大降低性能，因为Lid是必不可少的 我们还尝试删除Lsr，仅保留Lid（将λ设置为1），这会使Lid在训练过程中达到无穷大，这表明Lsr对于从LR图像中恢复有效的视觉内容是必不可少的。 当λ> 1时，这意味着Lid的权重比Lsr大，两个数据集的得分随λ的增加而急剧下降，这表明Lsr相对更为重要，因为超分辨率的性能直接决定了Ld的回收质量。 重新ID。

