利用生成对抗网络在野外寻找小脸

盐城白1 *，* 3张永强1 *，* 2 明利丁2伯纳德·加尼姆1 1

阿卜杜拉国王科技大学视觉计算中心2

哈尔滨工业大学电气工程与自动化学院3

中国科学院软件研究所（CAS）

baiyancheng20@gmail.com { zhangyongqiang，dingml } @ hit.edu.cn bernard.ghanem@kaust.edu.sa



图1。野外微小面孔的检测结果。（a）是原始的低分辨率模糊面孔，（b）是通过双线性核直接调整大小的结果，（c）是通过超分辨率方法生成的图像，而我们的结果（d）同时通过超分辨率（× 4放大）和优化网络获悉。最好以彩色查看并放大。

# 抽象

*面部检测技术已经发展了数十年，并且尚待解决的挑战之一是在不受限制的条件下检测小的面部。原因是细小的面孔通常缺乏详细的信息并且模糊不清。在本文中，我们提出了一种通过采用生成对抗网络（GAN）从模糊的小脸直接生成清晰的高分辨率脸的算法。为此，基本的GAN配方是通过依次超分辨率和精炼（*例如*SR-GAN和cycle-GAN）来实现的。但是，我们设计了一个新颖的网络来共同解决超分辨率和精细化问题。我们还引入了新的训练损失，以指导生成器网络恢复详细信息并促进区分器网络区分真实*与否。*。假冒脸* VS *。同时面对。在具有挑战性的数据集WIDER FACE上进行的大量实验证明了我们提出的方法从模糊的小脸中恢复清晰的高分辨率脸部的有效性，并表明检测性能优于其他最新方法。*

# 简介

人脸检测是计算机视觉中的一个基本且重要的问题，因为它通常是迈向许多后续与人脸相关的应用程序的关键步骤，包括人脸解析，人脸验证，人脸标记和检索等。过去，人脸检测已被广泛研究。几十年来，已经针对大多数受限场景提出了许多准确有效的方法。最近的作品着重于不受控制的环境中的脸部，由于比例，模糊，姿势，表情和照度的显着变化，这更具挑战性。在[ 32 ]中可以找到有关面部检测方法的详尽调查。

现代的脸部检测器在大中型脸部上都取得了令人印象深刻的结果，但是，在小脸部上的性能却远远不能令人满意。对于小面（主要困难*例如*。10×10 像素）的检测是，小面缺乏足够的详细信息，以从所述相似的背景，区分它们。另一个问题是，现代的基于CNN的人脸检测器使用步幅为8、16或32 的降采样卷积（*conv*）特征图来表示人脸，这些人脸失去了大部分空间信息，而且过于粗糙，无法描述小脸。要检测小脸，[ 28]使用双线性操作直接对图像进行升采样，并彻底搜索升采样图像上的脸部。但是，该方法将增加计算成本，并且推理时间也将显着增加。此外，在[ 28 ]中，通常会以较小的放大系数（最多2倍）放大图像，否则会产生伪像。此外，[ 1 ，14 ，25 ，37 ]使用中间*CONV*特征映射在特定尺度，这使计算负担和性能之间的平衡来表示面。但是，浅层但细粒度的中间*转换*特征图缺乏辨别力，这会导致许多假阳性结果。更重要的是，这些方法无需考虑其他挑战，例如模糊和照明。

为了解决人脸检测中的麻烦，我们基于经典的生成对抗网络（GAN）框架，提出了一个统一的端到端卷积神经网络，以实现更好的人脸检测。我们的探测器中有两个子网，一个发生器网络和一个鉴别器网络。在生成器子网络中，超分辨率网络（SRN）用于将小脸部向上采样到合适的比例，以找到这些小脸部。与通过双线性操作调整大小相比，SRN可以减少伪像，并以较大的放大因子（在我们当前的实现中为4倍）提高上采样图像的质量，如图1（b）和（c）所示。但是，即使使用了这种复杂的SRN，由于面临非常低的分辨率，上采样的图像仍不能令人满意（通常是模糊的图像，并且缺少精细的细节）。10×10 像素）。因此，提出了一种细化网络（RN），以恢复上采样图像中的某些缺失细节，并生成清晰的高分辨率图像进行分类。在鉴别器子网中，我们引入了一个新的损失函数，该函数强制鉴别器网络同时区分真实/假人脸和人脸/非人脸。生成的图像和真实图像通过鉴别器网络，以共同区分它们是真实图像还是生成的高分辨率图像，以及它们是人脸还是非人脸。更重要的是，分类损失用于引导生成器生成更清晰的面孔，以便于分类。

贡献。综上所述，本文做出以下三个主要贡献。提出了一种新颖的统一的端至端的卷积神经网络的用于面部检测的体系结构（1），其中，超分辨率和细化网络被用于产生实部和尖锐的高分辨率图像和鉴别器网络被引入来分类面*VS*。非面孔。（2）引入新的损失以促进鉴别器网络同时区分真实/伪造图像和面部/非面部。更重要的是，分类损失用于指导生成网络生成更清晰的面孔，以便于分类。（3）最后，我们证明了我们提出的方法可从模糊的小脸部中恢复清晰的高分辨率脸部的有效性，并表明检测性能优于WIDER FACE数据集上的其他最新方法。在最具挑战性的Hard子集上。

# 相关工作

## 人脸检测

作为经典话题，在过去几十年左右的时间里已经提出了许多面部检测系统。现有的面部检测方法，大致可分类为手工特征基础的方法[ 24 ， 29 ，30 ]和CNN-基础的方法[ 34 ，2 ，14 ，25 ，37 ，1]。但是，大多数基于手工特征的检测系统仅训练一个单一比例模型，该模型应用于特征金字塔的每个级别，因此大大增加了计算成本，尤其是对于复杂的特征。此外，手工特征的有限表示限制了探测器的性能，特别是在不受控制的设置中。

通过大获成功的鼓舞更快RCNN，近期的一些工作[ 14 ，25 ，37 ]利用这个框架，以检测在FDDB基准[面孔和表现出骄人的业绩13 ]。但是，在更具挑战性的WIDER FACE数据集[ 31 ] 上，性能会急剧下降，该数据集包含大量分辨率较低的面部。这种差异的主要原因是与较低的空间分辨率深CONV特征地图被用于表示，这不足以用于处理小面[ 34 ，2 ]。为了克服这个问题，检测器[ 14 ，25 ，37）必须在训练和测试期间通过将输入图像调整为不同的比例来进行升采样，这不可避免地增加了内存和计算成本。此外，调整大小方法通常会生成结构失真较大的图像，如图1（b）所示。与这些方法相比，我们的方法利用超分辨率和细化网络来生成具有高分辨率（4 × 放大）的清晰，精细的人脸，如图1（d）所示，然后训练鉴别器来区分人脸和人脸。输入图像。

## 超分辨率和细化网络

随着深度学习的发展，很大的改进已经在超高分辨率[取得5 ，6 ，15 ，26 ]。但是，当获得这些有希望的结果时，存在一个前提，即已知下采样内核，并且这些基于CNN的超分辨率方法中的大多数不能应用于不受控制的设置（*即*在野外）。

针对不同任务有不同的优化网络，与我们的优化网络最相似的优化方法是去模糊方法。大多数现有的去模糊方法在很大程度上依赖于以前的型号解决illposed问题，并事先假设自然图像的梯度有重尾分布[ 21 ，7 ]。近来，传统的神经网络也被用于去模糊的图像盲[ 36 ，23 ，3 ]。然而，这些去模糊方法仍然涉及显式的核估计，并且如果所估计的核不准确，则所恢复的图像通常具有明显的振铃伪影。

尽管现有的超分辨率方法和细化方法分别在上采样和细化图像方面有效，但是要联合超分辨和细化低分辨率图像并不容易。[ 28 ]提出了一种从模糊的低分辨率输入中同时重建清晰的高分辨率图像的方法。但是，它们的模糊低分辨率图像是通过使用双三次插值下采样和高分辨率图像（*即*合成图像）中已知的模糊核获得的。在本文中，我们设计了一种新颖的网络，可以从野外收集的小的模糊面孔中生成清晰的超分辨率面孔。我们想指出的是，我们的工作是试图共同超分辨率和改善野外小模糊面孔的第一项工作。

## 生成对抗网络

在开创性的工作中[ 8 ]，引入了生成对抗网络（GAN）以从随机噪声中生成逼真的图像。GAN在图像生成[ 4 ]，图像编辑[ 38 ]，表示学习[ 18 ]，图像注释[ 27 ]，图像超分辨率[ 17 ]和字符转移[ 12 ]中取得了令人印象深刻的结果。最近，GAN已应用于超分辨率（SRGAN）[ 17]，并取得了可喜的结果。与自然图像上的超分辨率相比，野外的人脸图像具有任意姿势，照明和模糊，因此人脸图像的超分辨率更具挑战性。更重要的是，由SRGAN生成的高分辨率图像模糊且缺少精细的细节，尤其是对于低分辨率的脸部而言，这对于脸部分类器而言是不友好的。为此，我们设计了一个精炼子网以恢复一些详细信息。在鉴别器网络中，基本GAN [ 17 ，12 ，8 ]被训练来区分真假高分辨率图像。为了对人脸或非人脸进行分类，我们扩展了鉴别器网络以对假冒*vs*进行分类。真实与面对*vs*。同时面对。此外，分类损失会传播回生成器网络，并引导生成器网络重建更清晰的超分辨率图像，以便于分类。

# 提出的方法

在本节中，我们将详细介绍我们提出的方法。首先，我们对经典GAN网络进行简要描述。然后，展示了我们方法的整体架构，如图2 所示。最后，我们详细介绍了网络的各个部分，并定义了分别训练生成器网络和鉴别器网络的损失函数。

## GAN

GAN [ 8 ] 通过对抗过程学习生成模型*G。*它同时训练生成器网络*ģ*和鉴别器网络*d*。训练过程交替优化生成器和鉴别器，它们相互竞争。训练生成器*G*以生成样本来欺骗鉴别器*D*，并且鉴别器*D*被训练以区分真实图像和伪图像与生成器。目标函数可以定义如下：

大号*GAN*（*G，d*）= E *X*〜*p 的数据*（*X*）[日志*d θ*（*X*）] +

（1）电子*Ž* 〜*p Ž*（*ż*） [日志（1 - *d θ*（*G ^ ω*（*ż*）））] *，*

其中*z*是随机噪声，*x*表示真实数据，*θ*和*ω*分别表示*G*和*D*的参数。在这里，*G*尝试最小化目标函数，而鉴别器*D*尝试将其最大化，如下所示Eq（2 ）：

argminmax L *GAN*（*G，D*）*。* （2）

*g ^ d*

类似于[ 8 ，17 ]，我们进一步设计的生成器网络G*ģ 瓦特ģ，*其在一种替代方法与鉴别器网络沿着优化*d θ d*来解决小面超分辨率和分类问题，其被定义如下：精氨酸

è （*我LR ，Y*）〜*p ģ*（*我LR ，Y*）[数（1 - *d θ d*（*g ^ 瓦特ģ*（*我LR，Y*）））] *，*

（3）

其中 *I LR* 表示具有低分辨率的人脸候选者，*I HR*表示具有高分辨率的人脸候选者，而*y*是标签（*即人*脸或非人脸）。与[ 8 ] 不同，发生器的输入是低分辨率图像，而不是随机噪声。与[ 17 ] 不同，我们在生成器网络中同时对输入图像进行升采样和优化。在鉴别网络中，我们区分产生*VS*。真正的高清晰度图像和面*VS*。共同面对。

|  |
| --- |
| 图2.拟议的人脸检测器系统的管线。（A）将图像馈入网络；（B）MB-FCN检测器是我们的基线，它从输入图像中裁剪出正数据（*即人*脸）和负数据（*即*非人脸），以训练生成器网络和鉴别器网络，或生成测试的兴趣（ROI）。（C）MB-FCN检测器生成正数据和负数据（或ROI）。（D）训练生成器网络以从低分辨率输入图像重建清晰的超分辨率图像（4倍放大），该图像包括上采样子网络和细化子网络。（D）鉴别器网络是具有两个并行的vgg19架构*fc*层，第一个*fc*层用于区分自然真实图像或生成的超分辨率图像，第二个*fc*层用于分类人脸或非人脸。 |

## 网络架构

我们的生成器网络包括两个组件（*即上*采样子网络和优化子网络），第一个子网络将低分辨率图像作为输入，而输出则是超分辨率图像。由于模糊的小脸缺少精细的细节，并且由于MSE损失Eq（4 ）的影响，生成的超分辨率脸通常会变得模糊。因此，我们设计了第二个子网，以优化来自第一个子网的超分辨率图像。此外，我们将分类分支添加到鉴别器网络中以进行检测，这意味着我们的鉴别器可以对人脸和非人脸进行分类，并区分伪造图像和真实图像。

**生成器网络**。如表1 和图2 所示，我们采用了一种深层的CNN架构，该架构在[ 17 ]中显示了对图像超分辨率的有效性。网络中有两个反卷积层[ 20 ]（*即*反卷积层），这些内核将低分辨率图像上采样到2 × 超分辨率图像。与他们的网络相反，我们的生成器网络包括精炼子网，它也是一种深层的CNN架构。与[ 20 ] 相似，我们使用批处理归一化[ 11]和除最后一层之外的每个卷积层之后的整流线性单位（ReLU）激活。

上采样子网首先对低分辨率图像进行上采样，然后输出4 倍超分辨率图像，当小脸远离相机或处于快速运动状态时，该超分辨率图像会变得模糊。然后，细化子网处理模糊图像，并输出清晰的超分辨率图像，这对于判别器更容易对人脸进行分类。（鉴别非面孔）。

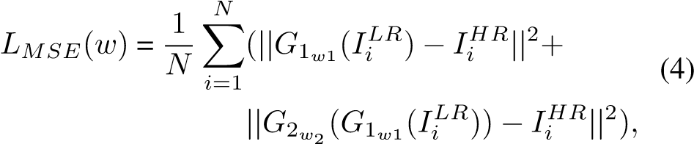
**鉴别器网络**。如表1 所示，我们将VGG19 [ 22 ]用作鉴别器的骨干网。为了避免对小的模糊面孔进行过多的下采样操作，作者从“ conv5”层中删除了最大池化。此外，我们用两个平行全连接层*FC GAN*和*FC CLC*全部更换完全连接层（*即*，*FC* 6 ，*FC* 7 ，*FC* 8 ）。输入是超分辨率图像，*fc GAN*分支的输出是输入为真实图像的概率，*fc*的输出是*clc*是输入*为人*脸的概率。

## 损失函数

我们从一些最先进的方法引入逐像素损失和对抗性损失[ 17 ，12 ]，以优化我们的生成器网络。与[ 17 ]相比，我们消除了由于计算成本引起的VGG特征匹配损失，并引入了分类损失来驱动生成器网络以从模糊的小脸中恢复出精细的细节。

**逐像素损失**。

我们的生成器网络的输入是小的模糊图像，而不是随机噪声[ 8 ]。使生成器的输出接近超分辨率地面实况的自然方法是通过像素级MSE损失，其计算公式为Eq（4 ）：



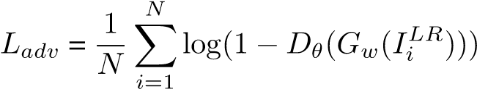
其中*I LR*和*I HR*分别表示小的模糊图像和超分辨率图像，*G* 1表示上采样子网，*G* 2表示细化子网，*w*是生成器网络的参数。

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | |  |  | 根 | erator | 鉴别器 | | 上采样子网 |  | 精炼子网 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 层 | 转换 | 转换  x8 | 转换 | 反转换 | 反转换 | 转换 | 转换 | 转换  x8 | 转换 | 转换 | 转换 | 转换 | 转换 | 转换 | 转换 | 转换 | 转换 | F | | 内核编号。 | 64 | 64 | 64 | 256 | 256 | 3 | 64 | 64 | 64 | 256 | 256 | 3 | 64 | 128 | 256 | 512 | 512 | 2 | | 内核大小 | 3 | 3 | 3 | 3 | 3 | 1个 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | -- | | 大步走 | 1个 | 1个 | 1个 | 2 | 3 | 1个 | 1个 | 1个 | 1个 | 1个 | 1个 | 1个 | 2 | 2 | 2 | 2 | 1个 | -- |   C  表1.生成器和鉴别器网络的体系结构。“ conv”表示卷积层，“ x8”表示具有8个卷积层的残差块，“ de-conv”表示步幅较小的卷积层，“ 2x”表示上采样倍数为2， |

“ fc”表示完全连接的层。

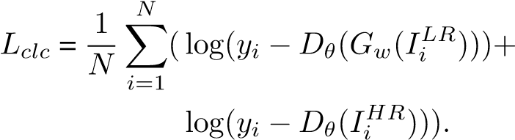
但是，虽然在像素级别上生成的高分辨率图像与自然高分辨率图像之间实现的损失较小，但MSE优化问题的解决方案通常缺少高频内容，这会导致具有过平滑纹理的感知上不令人满意的图像。同样，这是生成的图像模糊的原因之一。

**对抗损失**。为了获得更现实的结果，我们将对抗损失[ 17 ] 引入到客观损失中，定义为Eq（5 ）：

*。* （5）

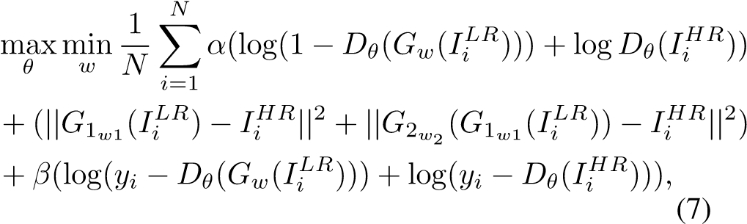
在这里，对抗性损失鼓励网络生成更清晰的高频细节，以试图欺骗鉴别器网络。在公式（5 ）时，*d θ*（*ģ 瓦特*（*我我LR*））的概率是重建图像*ģ 瓦特*（*我我LR*） 是一种天然的超分辨率图像。

**分类损失**。为了使通过生成器网络重建的图像更容易被分类，我们也将分类损失引入到客观损失中。让{ *我我LR，我* = 1 *，* 2 *，...，N* } 和{ *我我HR，我* = 1 *，* 2 *，...，N* } 分别表示小模糊图像和高分辨率的真实自然的图像，和{ *ÿ 我，我* = 1 *，* 2 *，...，N* } 表示对应的标签，其中*ÿn* =1或 *y n* =0分别表示图像是人脸或非人脸。分类损失的公式类似于等式（6）：

 （6）

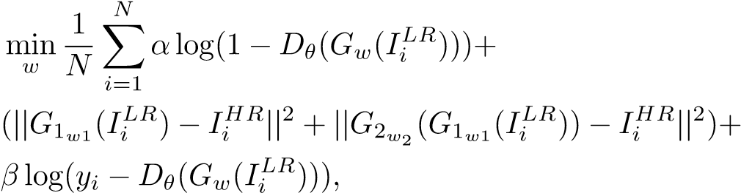
我们的分类损失起着两个作用，其中第一个是区分包括生成的图像和自然的真实高分辨率图像在内的高分辨率图像是鉴别器网络中的人脸还是非人脸。另一个作用是促进生成器网络重建更清晰的图像。

**目标函数**。基于以上分析，我们将对抗损失Eq（5 ）和分类损失Eq（6 ）合并到像素级MSE损失Eq（4 ）中。GAN网络可以通过目标函数Eq（7 ）进行训练：



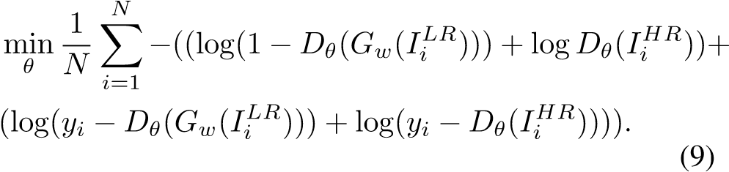
其中 *α* 和*β*是权衡权重。

为了获得更好的梯度行为，我们优化目标函数以另一种方式如在[ 17 ，12 ，10 ]和修改生成器的损失函数*G ^*和鉴别*d*为：



（8）

和



Eq（8 ）中生成器*G*的损失函数由对抗性损失Eq（5 ），MSE损失Eq（4 ）和分类损失Eq（6 ）组成，它们使重建的图像类似于真实的自然高分辨率图像在高频细节，像素和语义级别上。式（9 ）中的判别器*D*的损失函数引入分类损失以对高分辨率图像是人脸还是非人脸进行分类，这与GAN的基本公式相似[ 8]]来区分高分辨率图像是伪造的还是真实的。通过引入分类损失，从生成器中恢复的图像比通过对抗损失和MSE损失优化的结果更加真实。关于每个损失函数影响的进一步消融分析在第4.3 节中介绍。

# 实验

~~在本部分中，我们将在两个公开的人脸检测基准（~~*~~即）~~*~~上通过实验验证我们提出的方法。~~

WIDER FACE [ 31 ]和FDDB [ 13 ]）。~~首先，我们进行消融实验以证明GAN的有效性。然后，我们对发生器和鉴别器网络中每种损失的重要性进行了详细分析。最后，我们在两个公开基准上对我们提出的人脸检测器进行了评估，同时将其性能与其他最新方法进行了比较。~~

## 培训和验证数据集

~~我们使用了最近发布的大规模人脸检测基准，即WIDER FACE数据集[ 31 ]。它包含32,203张图像，这些图像是从可公开获得的WIDER数据集中选择的。分别随机选择40％/ 10％/ 50％的数据进行训练，验证和测试。WIDER FACE中的图像分为61个社交事件类，它们具有更多的多样性并且更接近于现实世界中的场景。因此，我们使用该数据集来训练和验证提议的生成器和鉴别器网络。~~

~~WIDER FACE数据集根据地面真实面孔的高度分为三个子集：Easy，Medium和Hard。Easy / Medium / Hard子集分别包含高度大于50/30/10像素的面部。与中型子集相比，“硬”子集包含许多高度在10 到30像素之间的面孔。不出所料，要在Hard子集上实现良好的检测性能是非常具有挑战性的。~~

## 实施细节

在生成器网络中，我们设置权衡权重*α* = 0 *。*001 和*β* = 0 *。*01 。在训练期间，我们使用亚当优化[ 16 ]与动量项*β* 1 = 0 *。*9 。从头开始训练生成器网络，并使用零均值高斯分布（标准差为0）初始化每层中的权重*。*02 ，并且偏差使用0 初始化。为了避免不良的局部最优，我们首先训练基于MSE的SR网络来初始化生成器网络。对于鉴别网络中，我们采用了VGG19 [ 22 对ImageNet]模型预先训练作为我们的骨干网络，我们全部更换*FC*有两个平行层*FC*层。的*FC*层通过零均值高斯分布的标准偏差0.1初始化，且所有偏压初始化为0。

我们的基线MB-FCN检测器基于ResNet50网络[ 9 ]，该网络在ImageNet 上进行了预训练。MB-FCN检测器的所有超参数均与[ 1 ] 相同。为了训练我们的生成器和鉴别器网络，我们裁剪了WIDER FACE的面部样本和非面部样本

[ 31 ]使用我们的基线检测器进行训练。通过使用系数为4的双三次插值对高分辨率图像进行下采样来生成相应的低分辨率图像。在测试过程中，裁剪了600个感兴趣区域（ROI），并将这些ROI馈入我们的GAN网络，

Method

Easy

Medium

Hard

|  |  |  |  |
| --- | --- | --- | --- |
| 基线[ 1 ] | 0.932 | 0.922 | 0.858 |
| 没有完善的网络 | 0.940 | 0.929 | 0.863 |
| 无广告损失 | 0.935 | 0.925 | 0.867 |
| 没有CLC损失 | 0.936 | 0.927 | 0.865 |
| 我们的（Baseline + MES + adv + clc） | 0.944 | 0.933 | 0.873 |

表2. WIDER FACE失效集上有无GAN，优化网络，对抗损失和分类损失训练的基准模型的性能。“ adv”表示对抗损失Eq（5 ），“ clc”表示分类损失Eq（6 ），并且“ MES”表示逐像素损失Eq（4 ）。

努力提供最终的检测性能。

所有GAN变体都以前3个时期以10 -4 的学习速率训练，另外3个时期以较低的10 -5的学习速率训练。我们交替更新生成器和鉴别器网络，等效于[ 8 ]中的*k* = 1 。我们的实现基于张量流，所有实验均在NVIDIA TITAN X GPU上完成。

## 消融研究

我们首先将我们提出的方法与基线检测器进行比较，以证明GAN的有效性。此外，我们通过移除精炼网络来执行消融研究，以验证精炼网络的有效性。最后，为了验证每种损失的贡献，包括对抗性损失和分类损失在生成器网络损失函数中的作用，我们还通过将每种损失累加到逐像素损失中来进行消融研究。

GAN的影响。表2 （在1 *日*和5 *个*行）示出了基线检测器的检测性能（AP）和我们对WIDER FACE验证集方法。我们的基线检测器是具有特征图的跳过连接的多分支RPN面部检测器，有关更多详细信息，请参考[ 1 ]。从表2中可以看出，在Hard子集上，我们的检测器性能大大优于基线检测器（在AP中为1.5％）。原因是基线检测器执行下采样操作（*即*。在小面上以步幅2）进行卷积。小脸本身包含的信息有限，经过多次卷积运算后，大部分详细信息将丢失。例如，输入为16 × 16面，在C4特征图上结果为1 × 1，而在C5特征图上未保留任何内容。基于这些有限的功能，获得较差的检测性能是正常的。相比之下，我们的方法首先学习了超分辨率图像，然后对其进行了细化，从而解决了原始的小模糊面孔缺少详细信息并且同时模糊的问题。基于具有精细细节的超分辨率图像，检测性能的提升是不可避免的。

|  |
| --- |
| 图3.在WIDER FACE验证集上，我们将我们的方法与几种最新方法进行了比较：MSCNN [ 31 ]，MTTCNN [ 33 ]，CMS-RCNN [ 37 ]，HR [ 10 ]，SSH [ 19 ]，SFD [ 35 ]。图例中报告了平均精度（AP）。彩色效果最佳。 |

完善网络的影响。 从表2 （的2 *次*和5 *次*行），我们看到，AP性能提高通过在硬盘子集1％通过将细化子网络到生成器网络。有趣的是，“易”和“中等”子集的性能也有所提高（0.4％）。我们将生成器网络中的重构人脸可视化，发现我们的优化网络可以减少照明和模糊的影响，如图4 所示。。在某些情况下，如果基线严重模糊或被照亮，基线检测器将无法检测到这些面部。但是，我们的方法减少了这种归因的影响，并且可以成功找到这些面孔。在这里，我们要注意我们的框架不是特定的，任何现成的面部检测器都可以用作我们的基线。

对抗损失的影响。从表2 （在3 *次*和 5 *次* 行），我们看到，AP性能约1％下降而不对抗损失。原因是由逐像素损失和分类损失导出的生成图像过于平滑。仔细检查生成的图像后，我们发现眼睛周围的细节质量低下。由于这些细节对于鉴别器而言并不是重要的特征，因此在该区域中犯错时，生成器仍然可以欺骗鉴别器。为了鼓励生成器恢复高质量的图像，我们在生成器损失函数中包括对抗损失。

分类损失的影响。从表2 （在4 *个*和5 *个*行），我们看到，通过约1％的AP的性能提高与分类损失。这是因为分类损失会促使生成器恢复精细的细节以便于分类。我们发现，当添加分类损失时，生成的面部具有更清晰的轮廓。我们认为轮廓信息可能是辨别器在人脸太小且模糊不清时对人脸/非人脸进行分类的最重要证据。

## 与最新技术的比较

我们在两个公开的人脸检测基准（*即）*上将我们提出的方法与最新方法进行了比较。

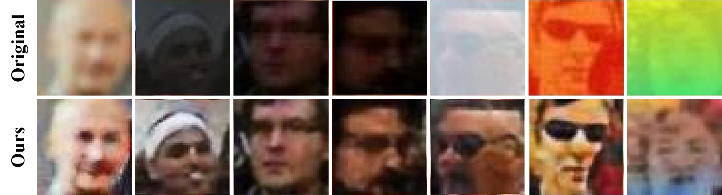
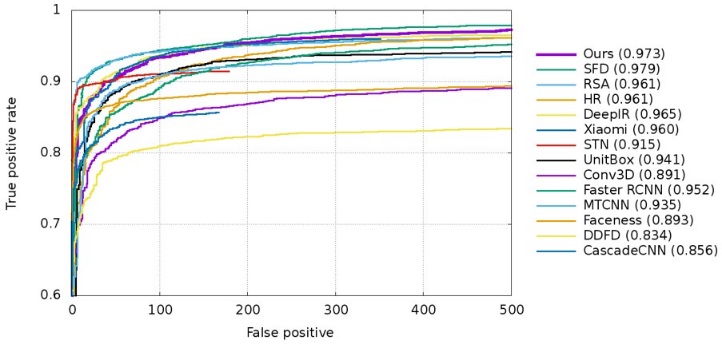


图4.我们的生成器网络从模糊的面孔生成的清晰面孔的一些示例。上排显示受模糊和照明影响的小脸，下排显示通过我们的方法生成的较清晰的脸。调整第一行中的低分辨率图像的大小以进行可视化。

WIDER FACE [ 31 ]和FDDB [ 13 ]）。

WIDER FACE的评估。我们比较我们的与国家的最先进的面部检测器[方法31 ，33 ，37 ，10 ，19 ，35 ]。图3 显示了WIDER FACE验证集的性能。从图3中，我们可以看到我们的方法实现了最高的性能（*即*。Hard子集上达到87.3％），其性能比最新的人脸检测器高出2％以上。与这些基于CNN的方法相比，我们性能的提升主要来自以下三个方面：（1）我们在生成器中的向上采样子网学习了超分辨率图像，从而减少了由于向下采样而导致的过多信息丢失在小面上执行卷积运算时；（2）生成器中的细化子网学习更精细的细节并重建更清晰的图像。基于清晰的超分辨率图像，与基于低分辨率模糊图像相比，区分者更容易对面部或非面部进行分类；（3）分类损失Eq（6）促进生成器学习更清晰的脸部轮廓，以便于分类。此外，我们在Easy / Medium子集上也获得了最高的性能（94.4％/ 93.3％），分别比最新的面部检测器高出0.7％和0.9％。这是因为某些大脸受照明和模糊的严重影响，如图4 所示。结果是，



|  |
| --- |
| 图5.我们提出的方法的定性检测结果。绿色边界框是地面真相注释，红色边界框是我们方法的结果。最好在计算机上以彩色和放大显示。 |

图6.在FDDB数据集上，我们将我们的方法与许多最新方法进行了比较。报告的误报率为500次。最好以彩色查看并放大。

基于CNN的方法无法检测到这些面孔。但是，我们的方法减轻了这些归因的影响，并成功找到了这些面孔。

FDDB的评估。我们遵循FDDB的标准指标（*即*特定误报率下的精度）[ 13 ]，并使用此指标与其他方法进行比较。FDDB中有许多未标记的面孔，因此在较小的误报率下，精度不准确。因此，我们报告了500次误报时的准确率。如图6 所示，我们的面部检测器比除SFD [ 35 ]检测器之外的所有其他现有技术的面部检测器均具有出色的性能（0.973）。我们要注意的是，SFD [ 35在测试集上手动添加238个未标记的面孔后，即可达到[]。但是，我们在原始标记的测试集上测试模型。在这种不公平的条件下，我们的方法仍然可以获得可比的性能，这进一步证明了我们方法的有效性。

## 定性结果

在图5中，我们显示了通过我们提出的方法生成的一些检测结果。可以发现，即使有些人脸很小且模糊，我们的人脸检测器也能成功找到几乎所有的人脸。但是，图5 还显示了一些失败案例，包括一些假阳性结果。这些结果表明，需要进一步的进步来进一步改善小脸检测性能。未来的工作将通过添加上下文来检测这些更具挑战性的小面孔来解决此问题。

# 结论

在本文中，我们提出了一种使用GAN在野外寻找小脸的新方法。在生成器网络中，我们设计了一种新颖的网络，可以从模糊的小图像中直接生成清晰的超分辨率图像，并且以端到端的方式训练我们的上采样子网和优化子网。此外，我们在鉴别器网络中引入了一个额外的分类分支，可以同时区分伪造/真实和面部/非面部。此外，分类损失被带到生成器网络以恢复更清晰的超分辨率图像。与以前的最新人脸检测器相比，在WIDER FACE和FDDB上进行的大量实验表明，我们的方法在Hard子集以及Easy / Medium子集中均得到了重大改进。

1. 原文

提供更好的翻译建议