arXiv:2305.01905v2 [cs.CV] 7 Sep 2023

使用多焦点空间注意力进行定位的口罩人脸识别

Yooshin Cho[1] Hanbyel Cho[1] Hyeong Gwon Hong[2] Jaesung Ahn[2]
Dongmin Cho[3] JungWoo Chang[3] Junmo Kim[1,2]韩国科学技术院电气工程学院，韩国[2]韩国科学技术院金载哲人工智能研究生院，韩国[3]Alchera公司

摘要——自全球新冠肺炎疫情暴发以来，人们一直建议佩戴口罩以限制疾病传播。然而，口罩会遮挡某些面部属性。因此，现有的人脸识别系统难以对佩戴口罩的人脸进行身份验证。在此背景下，有必要为非接触式生物识别系统开发口罩人脸识别（MFR）技术。为此，本文提出互补注意力学习和多焦点空间注意力，通过训练互补空间注意力分别聚焦于两个不同区域——遮挡区域和背景，从而精确移除遮挡区域。在我们的方法中，标准空间注意力和网络聚焦于未遮挡区域，并提取口罩不变特征，同时尽可能降低对传统人脸识别（FR）性能的影响。对于传统人脸识别，我们在 IJB-C、Age-DB、CALFW 和 CPLFW 数据集上评估性能。我们在 ICCV2021-MFR/Insightface 赛道上评估 MFR 性能，并展示了该方法在 MFR 和人脸识别数据集上的性能均有所提升。此外，我们通过实验证实，所提方法的空间注意力在未遮挡区域的激活更加精准。

I. 引言

随着深度神经网络的出现，在受控环境中，人脸识别（FR）的准确率已超过 99% [6], [23], [1]。因此，采用人脸识别的身份认证系统已广泛应用于我们的日常生活中（例如机场、公司和智能手机）。然而，由于近期全球 COVID-19 大流行，为了保护公共卫生，佩戴口罩已成为强制要求。这些口罩遮挡了鼻子和嘴巴，因此显著降低了以往人脸识别系统的性能。因此，口罩人脸识别（MFR）的必要性日益凸显，并且在疫情期间已有多项口罩人脸识别研究被提出。

以往关于 MFR 的研究主要聚焦于有效提取口罩不变特征 [27], [11], [21], [19], [4]。 [27] 中提出的方法利用口罩位于人脸下半部分区域这一先验知识，舍弃了从图像下半部分区域获取的特征。此外，还采用最大均值差异损失、均方误差损失或结合辅助口罩使用分类分支的对抗损失，以减小从中获得的特征之间的距离

本研究未获得任何机构的资助



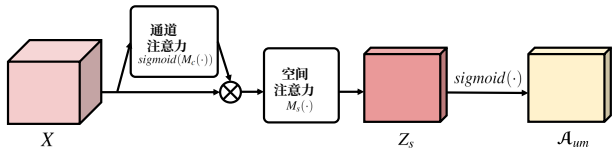
(a) 原始图像 (b) 蓝色 (c) 格林 (d) 黑色 (e) 白色

图 1：合成口罩人脸示例，由 ICCV2021/Insightface 赛道的在线口罩人脸生成函数生成。

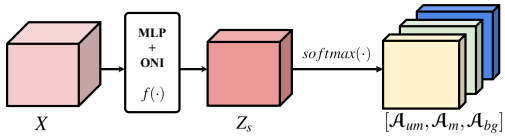
已掩码和未掩码图像 [19], [14]。这些方法通过迫使网络忽略通常被口罩遮挡的面部属性（鼻子和嘴巴），成功提升了 MFR 性能。然而，这与传统人脸识别的目标相冲突，并导致在人脸识别数据集上的性能下降。

为了优化传统人脸识别与MFR性能之间的权衡，我们提出了一种通过采用空间注意力模块 [3]来更精确地区分遮挡区域与未遮挡区域的方法。通常，卷积块注意力模块（CBAM） [25] 等空间注意力模块已被用于使网络聚焦于前景物体，并展示出将前景区域与背景区域分离的定位能力。受近期关于无偏视觉识别的研究启发 [24]，本文提出互补注意力学习（CAL），通过对抗性地利用互补空间注意力来增强CBAM的定位能力。我们训练互补空间注意力以学习不良信息（例如，口罩使用分类），从而防止标准空间注意力聚焦于不良区域（例如，遮挡区域）。具体细节将在第II-B节中介绍。

此外，我们提出了多焦点空间注意力（MFSA），将图像划分为三个区域（即未遮挡区域、遮挡区域和背景区域）。此前，*sigmoid* 函数通常用于对 CBAM 的空间注意力进行归一化，并将元素划分为二分类（例如前景和背景）。因此，如果我们训练互补注意力聚焦于遮挡区域，标准注意力就更可能激活背景区域。为缓解这一问题，我们采用*softmax* 函数将元素分类为 N 类。此外，我们没有采用 CBAM 的通道注意力模块，而是采用卷积



(a) 卷积块注意力模块（CBMA）



(b) 多焦点空间注意力 (MFSA)

图 2：**卷积块注意力模块（CBAM）和多焦点空间注意力（MFSA）的示意图**。我们分别用卷积层和*sigmoid* 函数以及*softmax* 函数替换 CBAM 的通道注意力和最后一个函数。此外，我们对卷积层的权重采用牛顿迭代正交化（ONI），以解耦各表征。

后续层采用批归一化和 ReLU。为了解耦 N 路注意力的表示，我们采用牛顿迭代正交化（ONI） [12],，对卷积层的权重进行正交化。权重正交化使每个输出通道依赖于不同的输入通道，从而降低输出通道之间的因果关系和相关性。通过这一改进，我们提升了模型的表示能力，并通过实验证明其在 MFR 和人脸识别数据集上均取得了更好的性能。

我们遵循 ICCV2021-MFR 挑战赛 Insightface 赛道中所介绍的基线模型实验设置 [4]。我们在 CASIA-Webface [26]和 MS1MV3 数据集 [6], [5], [9] 上训练网络，并通过口罩人脸增强，以给定概率合成带口罩的人脸。为实现口罩人脸增强，我们使用 Insightface 提供的在线遮挡人脸生成功能。图 1 展示了合成口罩人脸的示例。我们使用随机合成的口罩人脸训练网络，并在 ICCV2021-MFR/Insightface 赛道上进行评估；该赛道采用的是真实口罩人脸数据集。为分析 MFR 与人脸识别性能之间的权衡，我们在 IJB-C、Age-DB、CALFW 和 CPLFW 数据集上评估人脸识别性能 [28], [17], [29],[16]。从结果来看，我们通过实验验证了所提出的方法在普通人脸识别数据集和 MFR 数据集上均取得了优异性能。

由于强烈建议佩戴口罩，许多需要识别并与人类互动的学科都需要开发口罩人脸识别算法（例如网络安全、交通运输、公共卫生、人机交互和智能技术）。尤其是，我们的方法通过空间注意力图展示遮挡区域，因此对人类而言具有可解释性，这种方法是这是人机交互的一项基本行为。此外，如果存在合成图像生成器，我们的方法也可以应用于其他类型的面部遮挡（例如眼镜和帽子）。因此，我们期望该算法能够广泛应用于其他学科领域。

II. 使用多焦点空间注意力进行定位

我们提出了一种能够精确定位人脸未遮挡区域的方法。首先，我们提出了互补注意力学习（CAL），以防止空间注意力在不期望的区域（例如人脸的口罩区域）被激活。其次，我们提出了多焦点空间注意力（MFSA），它不仅将区域划分为二分类（例如前景和背景），还将其划分为三类（例如口罩区域、未遮挡区域和背景区域）。所提方法的详细内容将在以下各节中介绍。

A. 预备知识：*CBAM*

在介绍我们的方法之前，我们简要描述卷积块注意力模块（CBAM） [25]。CBAM 是一种具有代表性的注意力模块，它依次应用通道注意力模块和空间注意力模块。其可通过以下公式表示:

$$\mathcal{A}_c = sigmoid(M_c(X)), \qquad X_c = \mathcal{A}_c \otimes X, \tag{1}$$

$$\mathcal{A}_s = sigmoid(M_s(X_c)), \qquad Y = \mathcal{A}_s \otimes X_c, \tag{2}$$

其中，$X, Y \in \mathbb{R}^{C\times H\times W}$ 分别表示 CBAM 的输入特征和输出特征。 $\otimes$表示逐元素乘法，而$M_c(\cdot)$和$M_s(\cdot)$分别表示通道注意力模块和空间注意力模块。$X_c \in \mathbb{R}^{C\times H\times W}$ 是经过通道注意力优化的中间特征。通道注意力模块$M_c(\cdot)$由沿空间维度进行的最大池化、平均池化以及多层感知机（MLPs）组成。空间注意力模块$M_s(\cdot)$由沿通道维度进行的最大池化、平均池化以及卷积层组成。$M_c(X) \in \mathbb{R}^{C\times 1\times 1}$和$M_s(X_c) \in \mathbb{R}^{1\times H\times W}$ 分别通过 sigmoid 函数进行归一化，以计算通道注意力和空间注意力。通道注意力 $\mathcal{A}_c \in \mathbb{R}^{C\times 1\times 1}$ 和空间注意力 $\mathcal{A}_s \in \mathbb{R}^{1\times H\times W}$ 分别表示通道和空间位置的重要性。经验表明，CBAM 具有定位能力，并且 $\mathcal{A}_s$ 能够聚焦于前景物体。在本研究中，我们增强了定位能力，使网络能够精确地聚焦于人脸的未遮挡区域。

B. 互补注意力学习

为提取口罩不变特征，同时尽量减少人脸识别性能的损失，我们提出了一种方法，通过利用并增强 CBAM 的定位能力来精确分离遮挡区域。首先，我们使用以下公式生成两张注意力图， $\mathcal{A}_{um}$， $\mathcal{A}_m$ 以及两个特征，X_{um}，X_m:

$$Z_s = M_s(X_c), \quad \mathcal{A}_{um} = sigmoid(Z_s), \quad \mathcal{A}_m = 1 - \mathcal{A}_{um}, \tag{3}$$

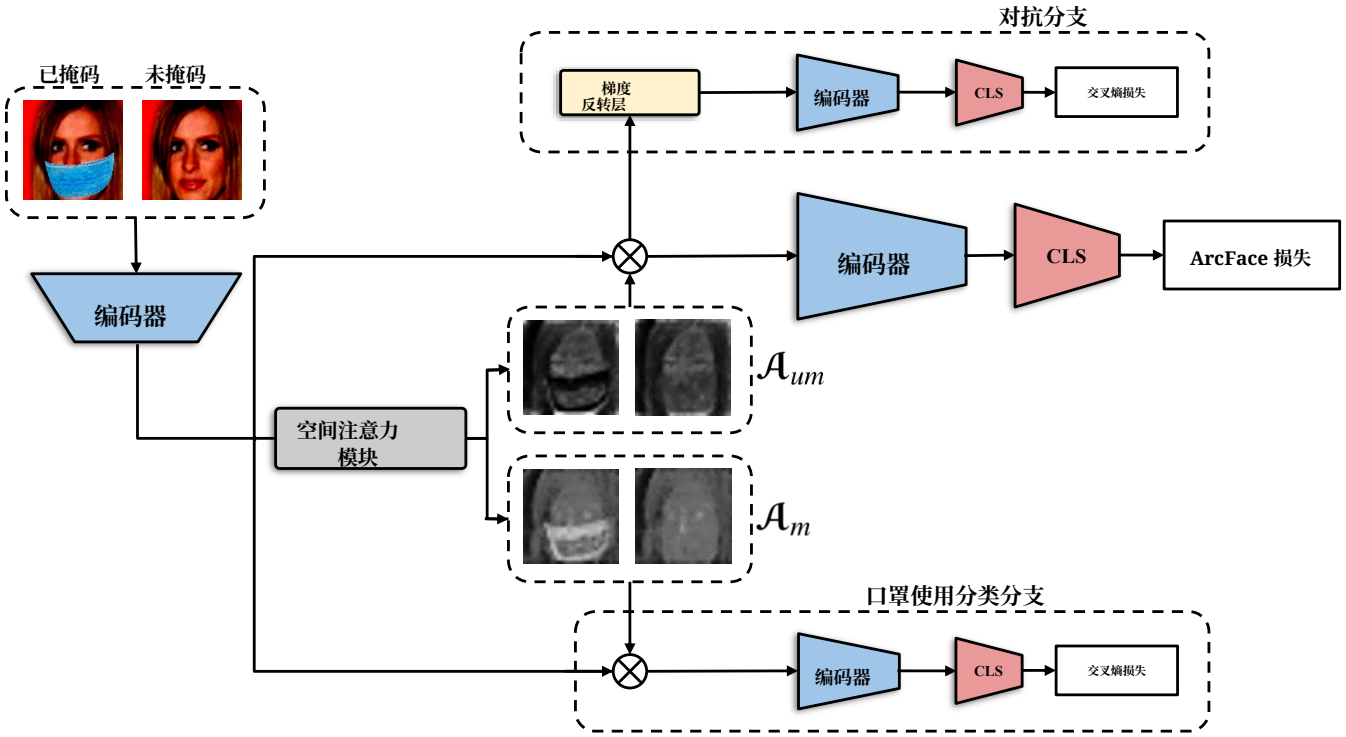


图3：**所提出的定位方法概览**。空间注意力模块根据输入图像生成注意力图，以分离已掩码和未掩码区域。A_m，A_{um} 分别为已掩码注意力和未掩码注意力。为在无标注的情况下定位已掩码和未掩码区域，使用A_m 通过 ArcFace 损失学习人脸识别，并使用A_{um} 通过交叉熵损失学习口罩佩戴分类。我们通过实验验证，已掩码和未掩码区域均得到了成功定位，并且采用 CAL 后，MFR 和人脸识别的性能均有所提升。

$$X_{um} = \mathcal{A}_{um} \otimes X_c, \qquad X_m = \mathcal{A}_m \otimes X_c, \tag{4}$$

其中，X_c 是根据公式 1 计算得到的中间特征，M_s 是 CBAM 的空间注意力模块。图 2a 展示了这一过程。

$\mathcal{A}_m \in \mathbb{R}^{1\times H\times W}$ 是 $\mathcal{A}_{um} \in \mathbb{R}^{1\times H\times W}$的互补注意力，因此二者会互斥激活。随后，互斥注意力$\mathcal{A}_{um}$，$\mathcal{A}_m$分别生成两个不同的特征X_{um}，X_m 。我们使用X_{um} 训练人脸识别，使其能够定位图像中的人脸区域。为防止 $\mathcal{A}_{um}$ 关注人脸的遮挡区域，我们使用X_m 训练口罩佩戴分类。X_m 经过优化，以获取与口罩使用相关的信息。这使得 $\mathcal{A}_m$关注遮挡区域，并凭借互补特性防止 $\mathcal{A}_{um}$ 在遮挡区域激活。我们将此方法称为“互补注意力学习”（CAL），并将基线方法的性能与带有辅助对抗学习分支的方法进行比较。图 3 展示了 CAL 和对抗分支的整体流程。对抗分支与口罩佩戴分支结构相同，区别在于分支起始处设有梯度反转层 [7] 。为进行比较，我们使用X_{um} 训练对抗分支，使X_{um} 不受口罩使用的影响。

我们使用 ArcFace损失 [6], 训练人脸识别，并使用交叉熵损失 [8],训练口罩佩戴分类分支和对抗分支，如下列公式所示：

$$\mathcal{L}_{arc} = -\frac{1}{N_b}\sum_{i=1}^{N_b} log(\frac{e^{s(cos(\theta_{y_i}+m))}}{e^{s(cos(\theta_{y_i}+m))}+\sum_{j=1,j\neq y_i}^{N_c} e^{s(cos(\theta_j))}}), \tag{5}$$

$$\theta_j = cos^{-1}(\frac{W_j^{\top}\cdot z_i}{\|W_j\|\|z_i\|}) \tag{6}$$

$$\mathcal{L}_{CE} = -\frac{1}{N_b}\sum_{i=1}^{N_b} log(\frac{e^{z_{y_i}}}{\sum_{j=1}^{N_c} e^{z_{y_j}}}), \tag{7}$$

其中，N_b 和 N_c 分别表示小批量大小和类别数。y_i，z_i 分别表示x_i的目标标签索引和逻辑值。$W \in \mathbb{R}^{C\times N_c}$ 表示最后一个线性分类器的权重，其中C 是z的维度大小。s 和m分别表示 ArcFace 损失的尺度和间隔， θ_j 表示特征z_i与权重W_j之间的角度。

C. 多焦点空间注意力

在 CBAM 中，空间注意力模块将区域划分为二分类区域，例如前景和背景。随后，我们使用 CAL 训练互补注意力$\mathcal{A}_m$ ，以定位口罩区域，并训练标准注意力 $\mathcal{A}_{um}$，以定位未遮挡区域。在这种情况下，难以对不属于口罩区域或未遮挡区域的背景区域进行分类。因此，我们提出

训练数据集	模型	测试数据集					
		MFR	MR-ALL	IJB-C	Age-DB	CALFW	CPLFW
CASIA	基线	18.49	23.65	69.63	94.07	94.07	88.97
	基线 +口罩人脸增强=0.1	31.23	26.35	**58.54**	93.96	93.27	88.7
	基线 + 对抗 +口罩人脸增强=0.1	33.36	26.64	31.63	93.95	93.15	88.7
	CBAM +CAL + 口罩人脸增强=0.1	35.35	**28.58**	50.36	93.92	**93.35**	**89.2**
	MFSA +CAL + 口罩人脸增强=0.1	**35.76**	27.61	48.21	**94.03**	93.22	88.73
	基线 +口罩人脸增强=0.3	40.37	26.55	**47.82**	93.87	**93.32**	88.82
	基线 +对抗 + 口罩人脸增强=0.3	41.73	24.31	30.62	93.33	93	88.37
	CBAM + CAL +口罩人脸增强=0.3	42.1	26.14	35.13	93.8	92.8	88.53
	MFSA + CAL + 口罩人脸增强=0.3	**43.44**	**28.94**	35.99	**93.95**	93.07	**88.92**
	基线 + 口罩人脸增强=0.5	42.83	21.80	**18.34**	**93.07**	92.7	87.68
	基线 + 对抗 +口罩人脸增强=0.5	43.15	**22.36**	8.92	92.95	92.6	**87.98**
	CBAM +CAL + 口罩人脸增强=0.5	44.4	20.74	11.45	92.68	92.52	87.9
	MFSA + CAL + 口罩人脸增强=0.5	**45.2**	21.87	11.86	92.9	**92.83**	87.97
MS1MV3	基线	65.86	80.53	94.80	98.30	96.17	**92.90**
	基线 + 口罩人脸增强=0.5	78.25	69.41	**93.68**	97.90	96.03	92.50
	基线 + 对抗 +口罩人脸增强=0.5	78.48	68.71	93.54	97.03	95.13	92.30
	CBAM + CAL +口罩人脸增强=0.5	78.45	69.30	93.57	**98.03**	**96.13**	92.72
	MFSA +CAL + 口罩人脸增强=0.5	**78.70**	**69.64**	93.62	97.90	96.08	92.70

表I：**开源人脸识别数据集上的验证性能**。我们报告 IJB-C 数据集上的 1:1验证 TAR（@假接受率=1e- 5），以及 Age-DB、CALFW 和 CPLFW 上的验证性能（%）。“MFR”和“MR-ALL”分别表示 ICCV2021-MFR/Insightface 赛道中口罩测试集上的 TAR（@假接受率=1e-4）和多种族测试集上的 TAR（@假接受率=1e-6）。“MA”表示遮挡人脸增强概率。最佳结果以粗体标出，次佳结果加下划线。

多焦点空间注意力（MFSA），用于对区域进行 N 分类。MFR 有三类区域（未遮挡区域、口罩区域和背景区域），因此我们使用采用 3 分类的 MFSA。随后，MFSA 可由以下公式表示：

$$Z_s = f(X), \qquad [\mathcal{A}_{um}, \mathcal{A}_m, \mathcal{A}_{bg}] = softmax(Z_s) \quad (8)$$

$$X_{um} = \mathcal{A}_{um} \otimes X, \quad X_m = \mathcal{A}_m \otimes X, \quad X_{bg} = \mathcal{A}_{bg} \otimes X, \quad (9)$$

其中，$f(\cdot)$ 是由逐点卷积层、批归一化层和 ReLU [15], [18] 组成的网络。为增强 $f(\cdot)$ 的判别能力，我们采用牛顿迭代正交化（ONI） [12]。该方法对逐点卷积层的权重进行正交化，并解耦注意力表示。正交化是一种常用技术，能够改善网络训练行为的条件性 [2], [13]。此外，权重正交化强制 Z_s 的每个通道依赖于不同的输入通道，从而降低注意力之间的因果性和相关性。为计算注意力，$Z_s \in \mathbb{R}^{3\times H\times W}$ 沿通道维度通过 $softmax$ 函数进行归一化。最后，$\mathcal{A}_{um}$，$\mathcal{A}_m$，$\mathcal{A}_{bg} \in \mathbb{R}^{1\times H\times W}$ 分别与输入特征 X 进行逐元素相乘，以生成三个特征 X_{um}，X_m，X_{bg}。我们在图 2b 中对 MFSA 的过程进行了可视化。我们遵循第 II-B 节所述的 CAL，利用 X_{um}，X_m 。X_{bg} 未在训练过程中显式使用，但能够缓解背景区域的歧义。

III. 实验

A. 训练详情

我们采用 ResNet-50 [10] 作为骨干网络架构。我们使用标准数据增强（即翻转、平移、裁剪）以及使用 ICCV2021-MFR/Insightface 赛道中介绍的工具进行口罩增强 [4] 来训练网络。我们使用 SGD，以批大小为 512 的小批量训练 CASIA-Webface 和 MS1MV3 数据集 [26], [5], [6], [9]。动量和权重衰减分别设为 0.9 和 5e-4。我们将初始学习率设为 0.2，并采用多项式学习率衰减调度器 [20],[22] ，进行 2 个周期的热重启。对于 MS1MV3 和 CASIA-Webface 数据集，我们分别在 25 个周期和 34 个周期后完成训练。遵循 [6], 的设置，我们将尺度 s 设为 64，将间隔 m 设为 0.5，用于 ArcFace 损失。

B. 评估细节

我们在四个基准人脸识别数据集 IJB-C、Age-DB、CALFW、CPLFW [28], [17], [29], [16] 上评估常规人脸识别性能。此外，我们还报告了 ICCV2021-MFR/Insightface 赛道 [4] 提供的多种族人脸数据集（MR-ALL）和已掩码人脸数据集（MFR）上的性能。

ICCV2021-MFR/Insightface 赛道（MFR）的口罩测试集： 其中包含 6,964 张真实场景中的已戴口罩人脸图像和 13,928 张未戴口罩人脸图像。共有 20,892 张图像，涵盖 6,964 个身份。我们评估 1:1 人脸验证的真实接受率（@假接受率=1e-4）。

ICCV2021-MFR/Insightface 赛道的多种族测试集（MR-ALL）： 由四个人口群体组成

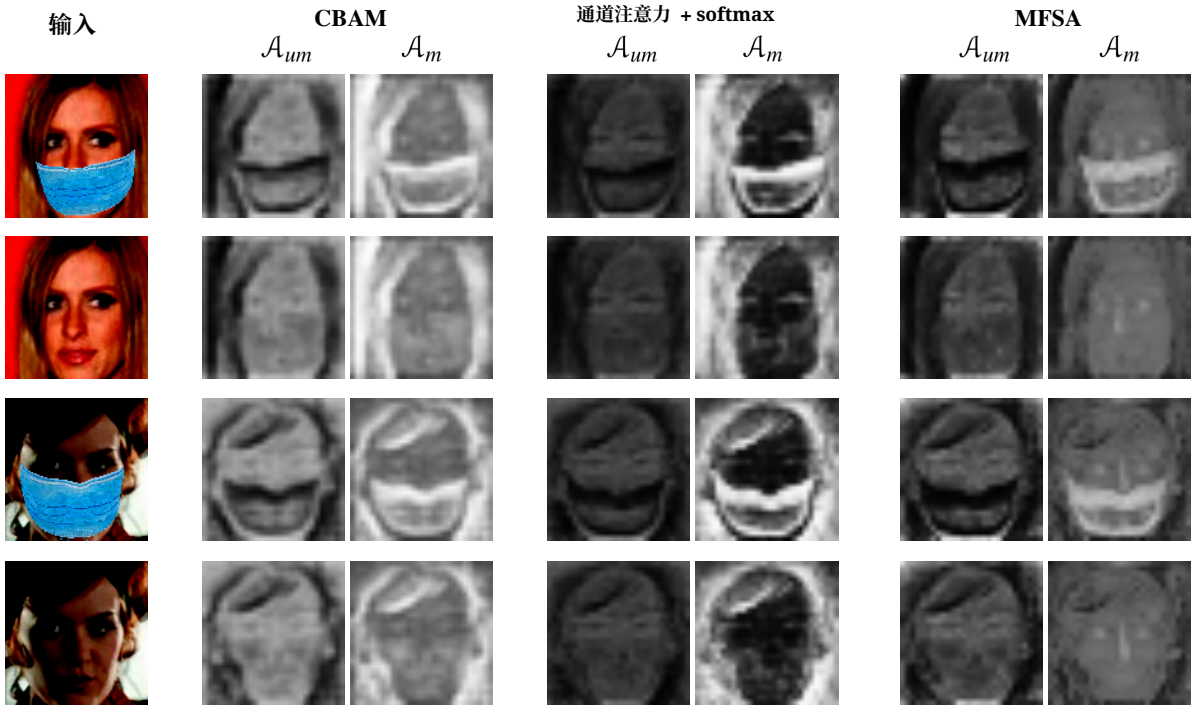


图 4：**注意力图的可视化**。空间注意力模块根据输入图像生成注意力图 $\mathcal{A}_{um}$,$\mathcal{A}_m$。我们通过改变注意力模块来可视化注意力图。“通道注意力 $+softmax$”表示中间注意力模块，该模块用 $softmax$ 函数替换 CBAM 的 $sigmoid$ 函数。

（非洲人、高加索人、南亚和东亚），包含 24.2 万个身份的 160 万张图像。我们评估 1:1 人脸验证的真实接受率（TAR）（@假接受率=1e-6）。

IJB-C：这是一个大规模人脸识别数据集，包含来自 3,531 个身份的 148.8K 张图像。我们评估 1:1 人脸验证的真实接受率（TAR）（@假接受率=1e-5）。

为探究所提方法的有效性，我们通过调整空间注意力模块（CBAM、MFSA）和训练方法（对抗、CAL），评估人脸识别和 MFR 的性能。“基线”表示使用 ArcFace 损失训练的 ResNet-50 的性能。“口罩人脸增强”表示已掩码人脸增强的概率。“对抗”表示采用辅助对抗分支进行对抗学习，使特征不受口罩使用的影响。

从表 I 所示结果可以看出，无论口罩人脸增强（MA）的概率和训练数据集如何，结合互补注意力学习（CAL）的多焦点空间注意力（MFSA）在 MFR 上的表现均最佳。此外，与对抗训练相比，CAL 在 IJB-C 上的性能下降幅度更小。对于 IJB-C，Baseline+MA 的表现最佳，而 CAL+MA 的表现排名第二。这表明 CAL 成功增强了空间注意力模块的定位能力，并能更精确地提取口罩不变特征。因此，传统人脸识别数据集上的性能退化较小。

C. 定性结果

图 4 展示了随着空间注意力模块变化而产生的两张注意力图 $\mathcal{A}_{um}$ 和 $\mathcal{A}_m$。正如预期，在 CBAM 中，背景区域难以被分类。因此，$\mathcal{A}_{um}$中的背景区域没有被清晰地去除，这并非理想现象。“Channel Att $+softmax$”表示将 CBAM 的 $sigmoid$ 函数替换为 $softmax$ 函数的注意力模块。在“Channel Att $+softmax$”中，背景区域能够被清晰分类，但未遮挡区域在 $\mathcal{A}_{um}$处没有被激活。此外，背景区域在 $\mathcal{A}_m$处被激活。我们认为应提高表征能力以缓解这一问题。因此，我们提出 MFSA，使用带有 ONI 的卷积层替代通道注意力，从而取得了理想结果。MFSA 的 $\mathcal{A}_{um}$ 对口罩使用相对不变，未遮挡区域、遮挡区域和背景区域也能得到更清晰的区分。

D. 消融实验

模型	MFR	MR-ALL	IJB-C
基线	18.49	23.65	**69.63**
基线 + 口罩人脸增强=0.1	**31.23**	**26.35**	58.54
基线 + 对抗 +口罩人脸增强=0.1	33.36	26.64	31.63
CBAM + 口罩人脸增强=0.1	32.71	27.24	37.20
CBAM +对抗 + 口罩人脸增强=0.1	34.04	27.47	33.66
CBAM + CAL + 对抗 + 口罩人脸增强=0.1	33.89	27.148	40.02
CBAM + CAL + 口罩人脸增强=0.1	**35.35**	**28.5**8	**50.36**

表 II：MFR、MR-ALL 和 IJB-C 上的 1:1 验证性能（%）对比。CAL 在所有测试集上均表现最佳。最佳结果以粗体标出。

为探究 CAL 的有效性，我们通过改变训练方法开展消融研究。如表 II 所示。

我们验证了口罩人脸增强能够提升 MFR 的性能，但会导致 MR-ALL 和 IJB-C 的性能下降。采用对抗方法后，MFR 的性能提升了 2.13%，但 IJB-C 的性能显著下降了 16.91%。相比之下，结合 CAL 的 CBAM 使 MFR 的性能提升了 3.12%，并使人脸识别性能下降了 8.18%。值得注意的是，CAL +对抗方法的性能不如 CAL。这表明，与对抗学习相比，CAL 是提取口罩不变特征的更有效训练方法。

IV. 结论

在本文中，我们提出了一种通过利用并增强 CBAM（代表性的注意力模块）的定位能力来提取口罩不变特征的方法。首先，我们提出互补注意力学习（CAL），以对抗性地利用互补注意力，防止标准注意力在不理想区域被激活。通过消融研究，我们从实验上证明，与简单的对抗学习相比，CAL 能更有效地提取口罩不变特征。其次，我们提出多焦点空间注意力（MFSA），将一幅图像划分为 N 路。它缓解了背景区域分类的歧义。通过 CBAM 和 MFSA 的可视化注意力图，我们验证了 MFSA 能够成功忽略背景区域并提取口罩不变特征。此外，无论训练数据集和 MA 概率如何，结合 CAL 的 MFSA 都能取得最佳的 MFR 性能，同时人脸识别性能退化相对较小。

参考文献

[1] F. Boutros、N. Damer、F. Kirchbuchner 和 A. Kuijper。Elasticface：用于深度人脸识别的弹性边际损失。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第 1578–1587 页，2022 年。[2] Y. Cho、H. Cho、Y. Kim 和 J. Kim。通过卷积单元优化提升批量白化的泛化能力。载于*IEEE/CVF*国际计算机视觉会议论文集，第 5321–5329 页，2021 年。[3] Y. Cho、Y. Kim、H. Cho、J. Ahn、H. G. Hong 和 J. Kim。重新审视 softmax 对轻量级非局部神经网络的效用。载于*2022*年*IEEE*国际图像处理会议（*ICIP*），第 1031–1035 页。IEEE，2022 年。[4] J. Deng、J. Guo、X. An、Z. Zhu 和 S. Zafeiriou。遮挡人脸识别挑战赛：InsightFace 赛道报告。载于*IEEE/CVF*国际计算机视觉会议论文集，第 1437–1444 页，2021 年。[5] J. Deng、J. Guo、E. Ververas、I. Kotsia 和 S. Zafeiriou。Retinaface：野外场景中的单次多层级人脸定位。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第 5203–5212 页，2020 年。[6] J. Deng、J. Guo、N. Xue 和 S. Zafeiriou。Arcface：用于深度人脸识别的加性角度间隔损失。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第 4690–4699 页，2019 年。[7] Y. Ganin 和 V. Lempitsky。通过反向传播进行无监督域适应。载于国际机器学习会议，第 1180–1189 页。PMLR，2015 年。[8] I. J. Good。理性决策。载于统计学突破，第 365–377 页。Springer，1992 年。[9] Y. Guo、L. Zhang、Y. Hu、X. He 和 J. Gao。Ms-celeb-1m：用于大规模人脸识别的数据集和基准。载于欧洲计算机视觉会议，第 87–102 页。Springer，2016 年。[10] K. He、X. Zhang、S. Ren 和 J. Sun。用于图像识别的深度残差学习。载于*IEEE*计算机视觉与模式识别会议论文集，第 770–778 页，2016 年。

[11] S. Hemathilaka 和 A. Aponso。口罩遮挡条件下遮挡不变人脸识别的全面研究。*arXiv*预印本 *arXiv:2201.09089*，2022 年。[12] L. Huang、L. Liu、F. Zhu、D. Wan、Z. Yuan、B. Li 和 L. Shao。深度神经网络训练中的可控正交化。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第 6429–6438 页，2020 年。[13] L. Huang 、 Y. Zhou 、 F. Zhu 、 L. Liu 和 L. Shao。迭代归一化：超越标准化，实现高效白化。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第 4874–4883 页，2019 年。[14] M. Huber、F. Boutros、F. Kirchbuchner 和 N. Damer。通过模板级知识蒸馏实现口罩不变人脸识别。载于*2021*年第*16*届*IEEE*国际自动人脸与手势识别会议（*FG 2021*），第 1–8 页。IEEE，2021 年。[15] S. Ioffe 和 C. Szegedy。批归一化：通过减少内部协变量偏移加速深度网络训练。载于国际机器学习会议，第 448–456 页。PMLR，2015 年。[16] B. Maze、J. Adams、J. A. Duncan、N. Kalka、T. Miller、C. Otto、A. K. Jain、W. T. Niggel、J. Anderson、J. Cheney 等。IARPA Janus基准-C：人脸数据集与协议。载于*2018*年国际生物特征识别会议（*ICB*），第 158–165 页。IEEE，2018 年。[17] S. Moschoglou、A. Papaioannou、C. Sagonas、J. Deng、I. Kotsia 和 S. Zafeiriou。AgeDB：首个通过人工收集的野外年龄数据库。载于*IEEE*计算机视觉与模式识别会议研讨会论文集，第 2 卷，第 5 页，2017 年。[18] V. Nair 和 G. E. Hinton。修正线性单元提升受限玻尔兹曼机的性能。载于*ICML*，2010 年。[19] P. C. Neto、F. Boutros、J. R. Pinto、N. Darner、A. F. Sequeira 和 J. S. Cardoso。FocusFace：用于口罩人脸识别的多任务对比学习。载于*2021*年第*16*届*IEEE*国际自动人脸与手势识别会议（*FG 2021*），第 01–08 页。IEEE，2021 年。[20] B. T. Polyak 和 A. B. Juditsky。通过平均加速随机逼近。*SIAM*控制与优化杂志，30(4):838–855，1992 年。[21] D. Qi、K. Hu、W. Tan、Q. Yao 和 J. Liu。平衡已掩码与标准人脸识别。载于*IEEE/CVF*国际计算机视觉会议论文集，第 1497–1502 页，2021 年。[22] H. Robbins 和 S. Monro。一种随机逼近方法。数学统计年刊，第 400–407 页，1951 年。[23] H. Wang、Y. Wang、Z. Zhou、X. Ji、D. Gong、J. Zhou、Z. Li 和 W. Liu。CosFace：用于深度人脸识别的大间隔余弦损失。载于*IEEE*计算机视觉与模式识别会议论文集，第 5265–5274 页，2018 年。[24] T. Wang、C. Zhou、Q. Sun 和 H. Zhang。用于无偏视觉识别的因果注意力。载于*IEEE/CVF*国际计算机视觉会议论文集，第 3091–3100 页，2021 年。[25] S. Woo、J. Park、J.-Y. Lee 和 I. S. Kweon。CBAM：卷积块注意力模块。载于欧洲计算机视觉会议（*ECCV*）论文集，第 3–19 页，2018 年。[26] D. Yi、Z. Lei、S. Liao 和 S. Z. Li。从零开始学习人脸表征。*arXiv*预印本 *arXiv:1411.7923*，2014 年。[27] Y. Zhang、X. Wang、M. S. Shakeel、H. Wan 和 W. Kang。通过双分支训练策略学习上部图块注意力，以实现口罩人脸识别。模式识别，126:108522，2022 年。[28] T. Zheng 和 W. Deng。Cross-Pose LFW：用于研究非受限环境下跨姿态人脸识别的数据库。技术报告 18-01，北京邮电大学，2018 年 2 月。[29] T. Zheng、W. Deng 和 J. Hu。Cross-Age LFW：用于研究非受限环境下跨年龄人脸识别的数据库。*CoRR*，abs/1708.08197，2017 年。