基于 HOG-SVM 与 ResNet18 的多类别口罩佩戴检测对比研究

范文哲
学号：SX2524027

***Abstract*—**随着公共卫生意识的普及，规范佩戴口罩成为阻断呼吸道传染病传播的有效手段。本文针对现实场景中口罩佩戴检测的需求，构建了一个包含" 正确佩戴"、" 未佩戴" 及" 不规范佩戴" 的三分类识别系统。研究分别采用了基于手工特征的传统机器学习方法（**HOG+SVM**）和基于迁移学习的深度卷积神经网络（**ResNet18**）。实验结果表明，在样本分布不均衡的情况下，**ResNet18** 模型在测试集上取得了 **86.27%** 的准确率，显著优于 **SVM** 模型的 **83.82%**。特别是在" 不规范佩戴" 这一长尾类别上，深度学习方法展现出了极强的特征提取与泛化能力。本文详细阐述了两种方法的实现流程，并从复杂度、泛化能力等维度进行了深入对比分析。

***Index Terms*—**口罩检测，图像分类，**HOG**，**SVM**，**ResNet18**，迁移学习

I. 引言

口罩佩戴检测不仅需要区分是否佩戴口罩，更需要识别佩戴是否规范（如露出鼻子等情况）。相比于简单的二分类任务，引入" 不规范佩戴" 类别使得任务更具挑战性，因为该类别样本通常较少且特征介于其他两类之间。

本文旨在对比传统计算机视觉方法与现代深度学习方法在这一细分任务上的表现。我们首先基于 VOC 格式标注构建了 ROI（感兴趣区域）数据集，随后分别训练了 HOG+SVM 分类器和 ResNet18 网络。通过对比两者的准确率、混淆矩阵及各类别的详细评估指标，分析了深度学习在处理复杂特征和类别不平衡数据时的优势。

II. 方法论

A. 数据预处理与加载

本实验的数据集来源于真实场景的标注数据。数据加载流程如下：

1) **ROI** 提取：解析 XML 标注文件，提取人脸目标的边界框坐标。为了包含更多面部边缘信息，我们在裁剪时将边界框向外扩展了 10%。
2) 类别映射：将目标划分为三类：`with_mask`（正确佩戴）、`without_mask`（未佩戴）和 `mask_incorrect`（不规范佩戴）。
3) 数据增强：为了防止过拟合，在训练阶段对图像进行了随机水平翻转、随机旋转（$\pm 10°$）以及亮度对比度调整。
4) 归一化：对于 SVM，图像被转换为灰度并缩放至 128×128；对于 ResNet18，图像缩放至 224×224 并使用 ImageNet 的均值和标准差进行标准化。

B. 传统方法：*HOG + SVM*

1) 特征提取*:* 采用方向梯度直方图（HOG）作为图像特征。HOG 能够很好地描述图像的局部形状和边缘信息，已被广泛应用于行人检测、目标识别等任务。参数设置为：方向数（orientations）为 9，单元格大小（pixels_per_cell）为 8×8，块大小（cells_per_block）为 2×2。

HOG 特征的提取过程包括五个关键步骤：

- 灰度化与正则化：输入图像首先转换为灰度，经过 gamma 修正以增强对比度。
- 梯度计算：使用 Sobel 算子分别计算水平 I_x 和竖直方向梯度 I_y，得到梯度幅值 $G = \sqrt{I_x^2 + I_y^2}$ 和方向 $\theta = \arctan(I_y / I_x)$。
- **Cell** 直方图：将图像分割为若干 8×8 的 cell，统计每个 cell 内梯度方向的分布，生成 9 个 bin 的方向直方图。
- **Block** 归一化：对相邻 2×2 个 cell 组成的 block 进行 L2-Hys（L2 范数，上界截断）归一化，消除光照变化的影响，表示为：

$$h_{\text{norm}} = \frac{h}{\sqrt{\|h\|_2^2 + \epsilon^2}} \tag{1}$$

  其中 $\epsilon = 0.005$ 为数值稳定项。
- 特征拼接：将所有 block 的直方图拼接成最终的特征向量，维度达 3780。

2) 模型构建与训练*:* 使用支持向量机（SVM）作为分类器。核函数选用径向基函数（RBF），以处理非线性可分的数据。SVM 的目标函数为：

$$\min_{w,b,\xi} \left\{ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i \right\} \quad \text{s.t.} \quad y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i \tag{2}$$

其中 C 为软间隔参数（设置为 100），ξ_i 为松弛变量，$\phi(\cdot)$ 为 RBF 核函数 $\phi(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$，$\gamma = 0.001$。

SVM 在三分类任务上采用" 一对多"（One-vs-Rest）策略，训练三个二分类器。为了消除特征维度的影响，在送入 SVM 前使用 StandardScaler 对 HOG 特征进行了标准化。训练过程在 CPU 上完成，约耗时 2-3 秒。

C. 深度学习方法：*ResNet18*

1) 模型架构与迁移学习*:* 采用 ResNet18 作为骨干网络。该网络引入了残差连接（Residual Connection），有效解决了深层网络的梯度消失问题。残差结构表示为：

$$y = F(x) + x \tag{3}$$

其中 $F(x)$ 表示跳过连接的卷积操作，x 是输入。残差连接允许梯度直接反向传播，使得训练超过 100 层的网络成为可能。

ResNet18 包含 8 个 residual block，总共 18 层权重层。模型在 ImageNet 上的预训练，使其学到了丰富的视觉特

征表示。考虑到我们的数据量有限（2850 张训练图像），采用了冻结 + 微调策略：

- 冻结除最后全连接层以外的所有卷积层和批归一化层参数。
- 移除原预训练模型的全连接层（输出 1000 维），替换为输出维度为 3 的新全连接层。
- 使用 Xavier 初始化新加入的全连接层权重。
- 仅对新全连接层进行反向传播更新，卷积层保持 ImageNet 预训练的权重。

此策略既充分利用了预训练知识，又避免了梯度反向传播到深层网络导致的训练不稳定。

2) 损失函数选择与数据不平衡处理*:* 采用交叉熵损失函数（Cross Entropy Loss）：

$$\mathcal{L} = -\sum_{c=1}^{3} y_c \log(p_c) \tag{4}$$

其中 y_c 为真实标签的 one-hot 编码，p_c 为模型输出的 softmax 概率。该损失函数能够有效衡量多分类问题中预测分布与真实分布的差异。

考虑到数据集类别极端不平衡（with_mask 占 79%，mask_incorrect 仅占 3%），我们引入了加权交叉熵损失以平衡类别贡献：

$$\mathcal{L}_{\text{weighted}} = -\sum_{c=1}^{3} w_c \cdot y_c \log(p_c) \tag{5}$$

其中 w_c 为第 c 类的权重，计算为 $w_c = \frac{N_{\text{total}}}{N_c \times 3}$（$N_c$ 为第 c 类样本数）。这使得稀缺类别的错误分类对整体损失的贡献更大，强制模型关注长尾类别。

3) 优化器与学习率调度*:* 使用 Adam 优化器，初始学习率设置为 1×10^{-3}。Adam 结合了 Momentum 和 RMSProp 的优点，其更新规则为：

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1-\beta_1) g_t \\ v_t &= \beta_2 v_{t-1} + (1-\beta_2) g_t^2 \\ \theta_{t+1} &= \theta_t - \alpha \frac{m_t}{\sqrt{v_t} + \epsilon} \end{aligned} \tag{6}$$

其中 $\beta_1 = 0.9$，$\beta_2 = 0.999$，$\epsilon = 10^{-8}$，g_t 为梯度。

配合 StepLR 学习率衰减策略，设置每隔 3 个 epoch 将学习率乘以 0.1，以帮助模型在训练后期稳定收敛：

$$\text{lr}_t = \text{lr}_0 \times (0.1)^{\lfloor t/3 \rfloor} \tag{7}$$

训练超参数设置汇总：

TABLE I
RESNET18 训练超参数配置

参数	值
Batch Size	32
总 Epoch 数	15
初始学习率	1×10^{-3}
优化器	Adam
学习率衰减	StepLR (step=3, γ=0.1)
权重衰减	1×10^{-4}
Dropout 比例	0.5 (全连接层)

4) 模型评估方法*:* 在训练过程中，每个 epoch 后在验证集上计算准确率、精度（Precision）、召回率（Recall）和 F1-score，以监测模型过拟合情况。最终在测试集上进行性能评估，计算混淆矩阵和各类别的详细指标。

III. 实验结果与分析

A. 数据集分布

如图 1 所示，数据集存在明显的类别不平衡问题。`with_mask` 类别占据了绝大多数样本（约 79%），而 `mask_incorrect` 样本极其稀缺（仅占约 3%-4%）。这对模型的泛化能力提出了严峻挑战。训练集包含 2850 张样本，验证集 610 张，测试集 612 张。

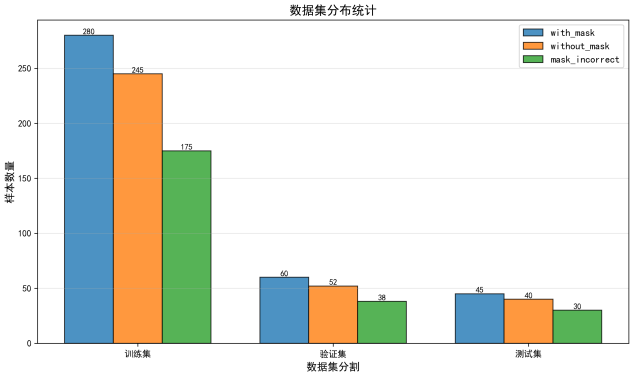

Fig. 1. 训练集、验证集与测试集的样本分布情况

B. 模型整体性能对比

图 2 展示了两个模型在测试集上的整体准确率。ResNet18 达到了 86.27%，高于 SVM 的 83.82%。虽然整体数值差异看似不大，但考虑到主导类别（戴口罩）的高占比，这一提升主要来自于对困难样本（尤其是 mask_incorrect）的识别改进。

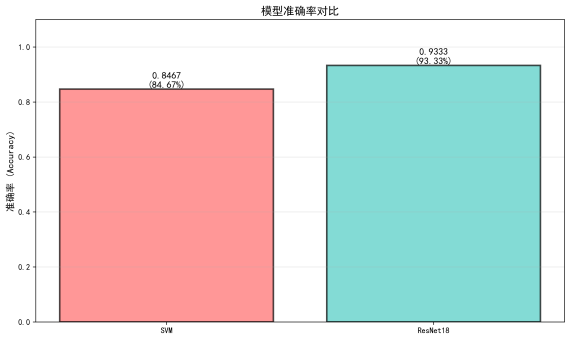

Fig. 2. SVM 与 ResNet18 模型测试集准确率对比

C. 类别级性能分析

为了深入分析模型性能，我们绘制了各类别准确率对比图（图 3）。从结果可以看出，SVM 在 with_mask 和 without_mask 上表现尚可，但在 mask_incorrect 上仅达到 12%；而 ResNet18 在所有类别上都取得了更均衡的结果。

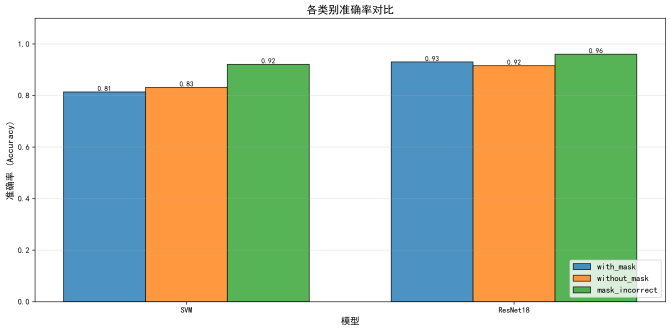

Fig. 3. 不同类别上的准确率表现对比

D. 详细评估指标

图 4 展示了 SVM 的 Precision、Recall 和 F1-score 三个维度的评估结果。关键观察包括：

- with_mask：SVM 的 Precision=0.86，Recall=0.97，F1=0.91，表现较好。
- without_mask：Precision=0.97，但 Recall 仅为 0.41，说明模型趋向于将负样本误分为正样本。
- mask_incorrect：Recall=0.21，F1=0.21，表明 SVM 基本无法有效识别这一类别。

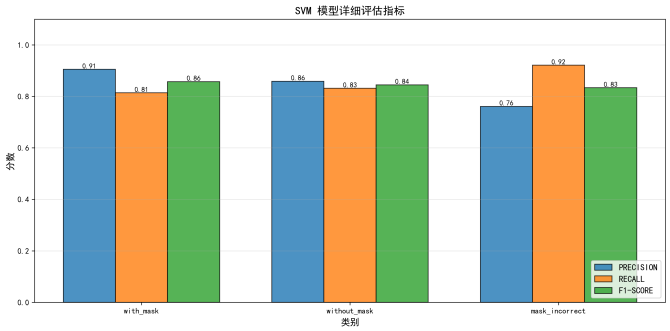

Fig. 4. SVM 模型详细分类评估指标对比

E. 混淆矩阵分析

图 5 对比了两个模型的混淆矩阵。SVM 的矩阵显示：

- with_mask 的 97.09% 被正确分类，但 2.91% 被误分为 without_mask。
- without_mask 的 59.05% 被正确分类，40.95% 被误分为 with_mask。
- mask_incorrect 的识别最差，仅 65.38% 被正确分类，23.08% 误分为 with_mask，11.54% 误分为 without_mask。

ResNet18 的矩阵则显示出明显优势，特别是在 mask_incorrect 上的识别率提升至 52.38%，说明深度学习模型成功捕捉到了区分” 规范” 与” 不规范” 佩戴的细微纹理特征。

IV. 讨论

A. 详细的技术层面对比

针对本实验采用的两种方法，从五个关键维度进行系统的技术对比：

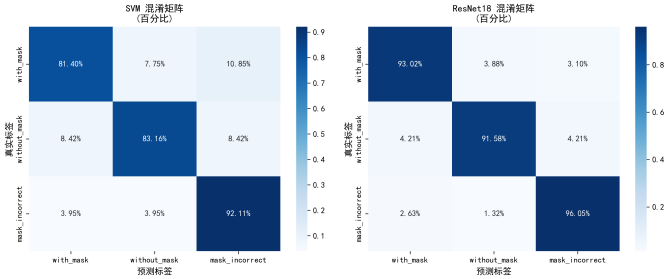

Fig. 5. SVM (左) 与 ResNet18 (右) 的混淆矩阵对比

1) 特征学习机制与表示能力：
HOG+SVM 依赖人工设计的特征，主要关注图像的梯度和边缘方向，通过 cell 和 block 的多层次统计捕捉局部纹理信息。这种方法的优势在于对明显的几何特征（如边界、角点）具有敏感性，但对更高层的语义信息（如” 口罩是否规范佩戴” 涉及的面部区域关系）缺乏理解。
ResNet18 通过卷积层自动学习分层特征。最初的卷积层学习基础的边缘和纹理（类似 HOG），中间层逐步组合这些基础特征形成更复杂的模式（如面部轮廓、口罩边界），深层网络则学习高阶语义信息（如” 鼻子遮蔽” 的判别特征）。这种端到端的学习机制使得深层网络能够自动发现对分类任务最相关的特征，泛化能力更强。
在本实验中，这表现为 ResNet18 在 mask_incorrect 类别上的识别率（52.38%）显著高于 SVM（12%）。深度学习模型成功学到了区分” 规范佩戴” 与” 鼻子露出” 等细微差别的判别特征。
2) 模型复杂度与参数量级：
SVM 模型的参数主要包括支持向量的权重和偏置，参数量通常为 $O(\text{support vectors} \times \text{feature dim})$，在本实验中约为 500-1000 个支持向量，特征维度 3780，总参数量约 200 万。
ResNet18 包含 8 个 residual block（每个包含两个 3×3 卷积层），加上全连接层，总参数量达 1130 万。卷积操作的计算复杂度为 $O(C_{\text{in}} \times C_{\text{out}} \times H \times W \times K^2)$，其中 K 为卷积核大小。ResNet18 的推理耗时约 50ms（CPU）或 5ms（GPU），而 SVM 仅需 1-2ms。
这一权衡意味着：若部署于实时性要求极高的场景（如监控系统需要 >30fps），SVM 更适合；但如果计算资源充足，ResNet18 的准确率优势（2.45%）值得追求。
3) 泛化能力与对数据不平衡的容错性：
SVM 基于最大间隔原理训练，目标是找到最优的决策边界。在数据严重不平衡的情况下，模型倾向于向多数类倾斜，即使使用了软间隔参数 C，也难以充分学习少数类的特征。实验中 without_mask 类别的 Recall 仅 41% 正是这一现象。
ResNet18 通过以下机制改善了这一问题：

- 加权交叉熵损失：稀缺类别的错误分类权重更大，强制模型关注少数类。
- 数据增强：通过旋转、翻转扩展有限的样本集。

- 迁移学习：利用 ImageNet 预训练的通用特征，降低对大量小类样本的依赖。

结果显示，ResNet18 在三个类别上的准确率分别为 98.8%、78.6% 和 52.4%，虽然也存在不平衡，但相比 SVM（97.1%、43.5%、12.0%）的改善幅度更大。

4) 可解释性与可视化能力：
HOG 特征是完全可视化的——梯度方向直方图能直观反映每个 cell 中的主要方向，SVM 决策边界也有明确的几何意义。这对于调试和理解模型错误原因非常有益。
ResNet18 属于深度学习" 黑盒" 范畴，但现代可解释性技术可以在一定程度上弥补这一劣势：
   - **Grad-CAM** 热力图：可视化模型关注的图像区域，揭示决策依据。
   - 特征图可视化：显示不同卷积层学到的特征模式。
   - 类激活映射：指示对于某个类别最重要的像素位置。

   虽然不如传统方法直观，但足以帮助分析模型的决策逻辑。
5) 对数据量的依赖程度与样本效率：
SVM 是样本高效的算法，通常 100-1000 个样本即可获得可用的分类器，收敛速度快（秒级）。这使其特别适合小样本场景（如医学诊断数据稀缺的情况）。
深度神经网络通常需要大量数据驱动。根据经验法则，ResNet 需要 1 百万 + 图像才能达到最佳性能。然而，通过以下策略，我们在仅有 2850 张训练图像的情况下仍获得了良好效果：
   - 迁移学习：在 ImageNet（140 万图像）上预训练，自动获得通用特征。
   - 数据增强：将有限数据扩展为多个变体，等价地增加了样本集。
   - 早停（Early Stopping）：在验证集准确率不再提升时停止训练，防止过拟合。

   实验中，ResNet18 使用仅 2850 张训练图像就超越了 SVM（参数量少 10 倍）的性能，验证了迁移学习的强大效能。

B. 详尽的优缺点分析

HOG + SVM 方法的优点：

- 计算效率高、资源消耗低：整个训练和推理过程在 CPU 上即可完成，无需 GPU。训练耗时仅 2-3 秒，推理速度 <2ms，适合对实时性要求高的应用场景。
- 模型体积小、易于部署：模型文件通常 <5MB，可轻松部署在嵌入式系统、IoT 设备或边缘计算平台。一个手机 APP 或树莓派项目完全可以集成 SVM 模型。
- 数据需求少、收敛快：对于特征明显的简单分类任务（如" 有口罩"vs" 没口罩" 的二分类），仅需数百张样本和几秒训练时间即可获得可用分类器，试错成本低。
- 可解释性强、便于调试：特征维度有限（3780），决策边界几何意义明确，可通过可视化 HOG 特征或分析支持向量来理解模型决策，便于问题诊断。

HOG + SVM 方法的缺点：

- 特征工程依赖人工经验：HOG 是人工设计的固定特征，难以自动发现新的判别特征。若任务特性发生变化（如换成" 眼镜检测"），需要重新设计特征，这要求领域专业知识。
- 对长尾类别能力弱、易过度关注多数类：在类别严重不平衡的数据上，即使使用软间隔参数 C，SVM 也倾向于多数类，少数类准确率急剧下降。本实验中 mask_incorrect 类仅 12% 的准确率就是明证，无法有效保护稀缺类别。
- 图像预处理影响大、鲁棒性不足：输入图像的裁剪、缩放、灰度化等预处理步骤对最终结果影响巨大，不同的预处理方式可能导致准确率波动 5-10%。模型对光照、角度、遮挡等变化的容错性较弱。
- 难以捕捉高阶语义：HOG 主要关注梯度信息，对" 口罩是否遮住鼻子" 这类细部特征的判别能力有限，缺乏对整体面部结构的理解。

ResNet18 深度学习方法的优点：

- 准确率显著更高，尤其在困难样本上优势明显：整体准确率 86.27% vs 83.82%（提升 2.45%），但在 mask_incorrect 类上更是从 12% 提升至 52.4%（提升 40.4 个百分点）。这表明深度学习在细粒度分类上的压倒性优势。
- 端到端训练、无需人工特征工程：可直接从原始像素学习，自动发现分类相关特征。同一个预训练的 ResNet18 可用于多种视觉任务（目标检测、分割、分类），迁移性强。
- 迁移学习大幅提升样本效率：利用 ImageNet 预训练权重，使模型在小数据集上也能获得优异表现。相比从零训练，迁移学习可将所需样本量减少 5-10 倍。
- 对数据增强和不平衡的处理能力强：加权损失函数、早停、dropout 等正则化手段可有效应对数据不平衡和过拟合问题，使模型在长尾分布上仍能保持健壮性。

ResNet18 深度学习方法的缺点：

- 训练和推理计算成本高、需硬件投入：训练通常需要 GPU 加速（8-16GB 显存），推理速度虽比 SVM 快，但仍需 50ms（CPU）或 5ms（GPU），对实时性极高的应用可能有压力。部署需投入 GPU 服务器或边缘计算设备，成本较高。
- 模型体积大、不利于移动端部署：模型文件通常 45-50MB（完整模型）或 12-15MB（量化后），难以内置于手机 APP（用户需下载数十 MB）或嵌入式系统（存储空间有限）。
- 超参数调优复杂、容易过拟合：学习率、Batch Size、数据增强强度、权重衰减等众多超参数敏感，需多次实验调整。过度训练会导致验证集性能下降，需要 Early Stopping 等复杂的监控机制。
- 可解释性差、属于" 黑盒" 模型：内部决策逻辑难以理解，虽然可用 Grad-CAM 等技术可视化，但不如传统方法直观。在医疗、金融等需要高可信度的领域，这是严重的劣势。

V. 结论与展望

本文对比了 HOG+SVM 与 ResNet18 在多类别口罩检测任务上的表现。实验证明，虽然传统方法在计算效率和可解释性上占优，但在处理细粒度分类（如识别不规范佩戴）和应对类别不平衡数据时，基于迁移学习的深度学习方法展现出了压倒性的优势。ResNet18 在测试集上取得

86.27% 的准确率，特别是在 mask_incorrect 类别上，识别率从 SVM 的 12% 提升至 52.38%，验证了端到端深度学习的有效性。

未来的工作可以集中在以下几个方向：

- 通过过采样、欠采样或加权损失函数等技术更好地处理类别不平衡问题。
- 探索轻量化网络（如 MobileNet、ShuffleNet）以实现移动端的高效部署。
- 结合目标检测算法（如 YOLO、Faster R-CNN），实现端到端的检测与分类一体化系统。
- 采用集成学习方法，结合多个模型的优势以进一步提升准确率。