FANVID：低分辨率视频中面部和车牌识别的基准测试

Kavitha Viswanathan[1] Vrinda Goel[1] Shlesh Gholap[1] Devayan Ghosh[1] Madhav Gupta[1] Dhruvi Ganatra[1] Sanket Potdar[1] Amit Sethi[1]

[1]Department of Electrical Engineering, Indian Institute of Technology Bombay, India , 184070024@iitb.ac.in, 20d070090@iitb.ac.in, shleshgholap@iitb.ac.in, 23m1079@iitb.ac.in, , 21D070043@iitb.ac.in, 21d070027@iitb.ac.in, sanketpotdar.iitb@gmail.com, asethi@iitb.ac.in

May 19, 2025

Abstract

现实世界中的监控通常会使脸部和车牌在单个低分辨率（LR）帧中不可辨认，从而阻碍可靠的识别。为了推进时间识别模型，我们提出了 FANVID，这是一种新颖的视频基准，包含近 1,463 个低分辨率片段（180×320 ，20–60 FPS），其中包含来自三个英语国家的 63 个身份和 49 个车牌。每个视频包括干扰脸部和车牌，增加了任务的难度和真实性。该数据集中含有 31,096 个手动验证的边界框和标签。FANVID 定义了两个任务：（1）脸部匹配——检测低分辨率脸部并将其与高分辨率照片进行匹配，以及（2）车牌识别——从低分辨率车牌中提取文本而无需预定义的数据库。视频是从高分辨率源下采样的，以确保人脸和文本在单帧中是无法辨认的，要求模型利用时间信息。我们引入了从平均精确度均值 (mAP@0.5) > 0.5 适配的评估指标，优先考虑脸部的身份准确性和文本的字符级准确性。通过预训练的视频超分辨率、检测和识别的基线方法实现的性能评分为 0.58（脸部匹配）和 0.42（车牌识别），突出了任务的可行性和挑战性。FANVID 在人脸和车牌的选择上在多样性与识别挑战之间取得了平衡。我们发布了用于数据访问、评估、基线和注释的软件，以支持可重复性和扩展。FANVID 旨在促进低分辨率识别的时间建模创新，应用于监控、法医和自动驾驶汽车等领域。

1 介绍

监控摄像头现在在城市环境中无处不在，通常会产生低分辨率（LR）的画面，在单帧中面孔和车牌无法识别。2025 年 3 月的 BBC 报道指出，许多来自闭路电视摄像机的图像过于模糊，难以进行可靠的人脸识别 [1] 。虽然司法证据必

须依靠高分辨率（HR），但迫切需要稳健的 LR 识别方法，以便缩小车辆、嫌疑人和失踪人员的搜索范围。

大多数现有的识别模型和基准，如 SCFace [10]，都依赖于静态的高分辨率图像，忽视了低分辨率序列中时间上下文的潜力。为了填补这个空白，我们引入了 FANVID（一种用于低分辨率视频中人脸和车牌识别的新基准，320×180，20–60 FPS），其中每个独立的帧对于肉眼来说都是无法辨认的。

FANVID 包含来自三个英语国家的 63 个身份和 49 个车牌的 1,463 个视频，总计 31,096 个经过人工验证的边框和标签。该基准定义了两个任务：（1）人脸匹配：在视频片段中检测低分辨率人脸并将其与高分辨率的面部照片匹配；（2）车牌识别：在没有先验数据库的情况下，检测并转录低分辨率的车牌文本。每个视频都包含干扰人脸和车牌，以模拟现实世界的条件。FANVID 数据集托管在 HuggingFace 上。

我们提出了任务特定的评估指标，这些指标是从平均准确率（mAP@0.5）中适应而来的，包含了面部匹配的身份级别正确性和车牌的字符级别准确性（通过标准化编辑距离）。FANVID 视频从高分辨率来源降采样而来，所有注释在降采样之前都经过了人工验证，以确保高质量的标签。基线流程使用 RCDM [29] 进行视频超分辨率，然后使用 RetinaFace [7] 和 ArcFace [6] 进行人脸识别，使用 EasyOCR [8] 进行文本识别，其准确率分别达到 0.58 和 0.42，这突显了时序建模方法面临的挑战和机会，无论是否采用视频超分辨率。

FANVID 在种族、性别、文化和车牌格式的多样性与识别混淆之间实现了平衡。为了促进负责任的研究，它使用公共领域视频的链接，并包含数据集创建和注释的工具。我们的贡献包括：

1. 一个经过精心整理的基准，由 1,400 多个视频中超过 30,000 个注释和身份/文本标签组成。
2. 使用 RCDM、RetinaFace、ArcFace 和 EasyOCR 的任务适配度量和基线。
3. 用于可重复性和扩展性的评估脚本及数据生成工具（使用 SAM2）。

FANVID 主要用于评估，但包括开源工具以支持未来的数据集增长，以应对额外的挑战和扩展训练。它的规模支持轻量级和深度模型，其对真实数据的关注使其成为推进监控、取证和自动驾驶研究的理想选择。我们相信，FANVID 将推动 LR 识别的时间建模进步，同时促进包容性、可重复和应用驱动的研究。

2 相关工作

一些数据集在面部和车牌检测、识别和验证以及超分辨率方面推动了技术的进步。然而，现有的数据集要么缺乏时间信息、真实的低分辨率条件、干扰项（其他人脸和车牌），要么缺乏反映现实世界监控和取证需求的任务结构。

2.1 人脸和车牌识别数据集

人脸和车牌识别在计算机视觉中已被广泛研究。像 LFW [12]、MegaFace [16] 和 IJB-C [21] 这样的数据集针对无约束的人脸识别，但依赖于高分辨率（HR）的静止图片或具有清晰正面视图的视频帧。面向监控的数据集如 SCFace [10] 和 QMUL-SurvFace [5] 引入了来自监控风格相机的低分辨率（LR）人脸图像，

但仅限于静态帧，缺乏有助于识别的时间信息。类似地，像 UFPR-ALPR [26] 、OpenALPR 和 VIVA 挑战 [14] 的数据集提供了真实世界的车牌图像，但这些通常是高分辨率或仅轻度退化，且通常不包含支持时序建模的序列。

2.2 视频中的识别：人脸、人物和文本

基于视频的识别通过聚合时间上的信息提供了一个克服单帧中低分辨率限制的机会。基准如 Celebrity-1000 [20] 、IJB-S [21] 和 MARS [35] 促进了基于视频的人脸或人物识别，但它们的视频片段质量通常更高。这些数据集没有呈现跨分辨率、跨模态（视频和图像）以及跨场景（不同摄像机、服装、背景）身份匹配的联合挑战。

对于车牌，大多数基准测试关注于孤立图像或帧中的字符检测。在数据集如 COCO-Text [28] 和 ICDAR 视频挑战 [15] 中已经探索了视频中文字的识别，但这些数据集关注于场景文字而非车牌，并且假设每帧具有足够的分辨率以进行准确的 OCR。支持跨帧文字识别的少数数据集（例如，TextVQA [27] ）展示的是高质量的影像。

2.3 视频超分辨率数据集及其适用性

视频超分辨率（VSR）数据集，如 Vid4 [19] 、Vimeo-90K [32] 和 REDS [22] 推动了帧增强技术的发展，但它们的主要任务是提升感知质量。因此，它们不适合评估 VSR 如何提升后续识别的表现。虽然联合 SR-识别管道显示出了一定的前景 [30] ，但当帧内容严重退化或在真实监控视频中物体显得很小或扭曲时，其效果会减弱。

FANVID 弥合了这些差距，开创了一类新的基准，用于从现实的低分辨率和干扰因素丰富的视频中进行身份验证和文本识别。

3 FANVID 数据集

FANVID（用于低分辨率视频中的人脸和车牌识别）是一个全新的基准数据集，旨在支持两个现实任务：(1) 人脸匹配，该任务涉及从低分辨率视频中识别与高分辨率照片中个人的匹配；(2) 车牌识别，该任务需要在没有外部数据库的情况下转录车牌文本。它强调在现实且具有挑战性的条件下使用时间线索，其中单个帧通常难以辨认并且存在干扰因素。数据集包含 1,400 多个面部和车牌视频片段，这些片段来源于公共的高分辨率视频素材，经过手动标注并降采样以模拟监控质量。接下来的小节中，我们将描述数据集的创建和注释过程、其特点，以及为重现性和扩展性提供的工具。

3.1 数据集创建

FANVID 是使用从 YouTube ® 获取的高分辨率 (HR) 视频构建的，专门关注类似于监控录像的内容——例如市民拍摄的街景、监控摄像头记录的画面，或名人在公共场合的抓拍。所有源视频均限制为公开可用内容，以确保可追溯性并遵循平台指南。为了建立身份的真实信息，从不同场景（如新闻活动或采访）中获取了每个主体（如名人）的面部照片或正面高分辨率图像。

最初利用 RetinaFace [7] 检测器生成了

面部标注 面部边界框，并使用 ArcFace [6] 特征嵌入与标准证件照库进行身份匹配。此自动步骤在视频帧中生成了临时的边界框和主体身份标签。这些标注随后被逐帧手动验证和校正，以确保在遮挡、极端姿势或能见度低的条件下实现高精度定位和身份标记。场景中出现的非目标面部（干扰物）未被标注。

针对车牌，一个高质量的单帧图像被手动标注出包围车牌区域的边界框。这个注释然后通过 [25] 方法传播到后续帧。如果需要，边界框会被手动检查和修正。车牌文本是从最清晰的高分辨率帧中手动转录的，即使在部分遮挡或运动模糊的情况下，也要确保字符正确。

经过验证和校正后，所有视频使用 OpenCV（双三次插值）统一降采样至空间分辨率为 320×180 。剪辑选择和降采样参数均经过校准，以确保在任何单独的帧中，人脸身份和车牌文本都不可被人类识别，从而强制依赖时间上下文进行识别。

3.2 数据集描述

FANVID 包含超过 31,096 个带注释的帧，涵盖 1,463 个视频。它代表了 63 个人，选择这些人是为了在种族（包括南亚、东亚、黑人、白人、毛利人、拉丁裔和混血）和性别多样性与外貌相似性之间达到平衡。每个人可以在多个片段中出现，有时是在不同的条件或摄像机角度下，以测试泛化能力。

在车牌方面，数据集包括 49 个来自多个国家的独特车牌，这些车牌在结构、颜色和可见性上有所不同。尽管我们限制使用英文，车牌在颜色（白色或黄色背景）、长度（五到九个字符）和内容（字母数字与数字）上有所不同。车牌可能在多个帧中被分割或遮挡。

该数据集已被分为预定义的 `train` （包括验证）和 `test` 集合。没有个体或板在多个分割中出现，以确保泛化。分割根据难度和身份类别或板特征来平衡。

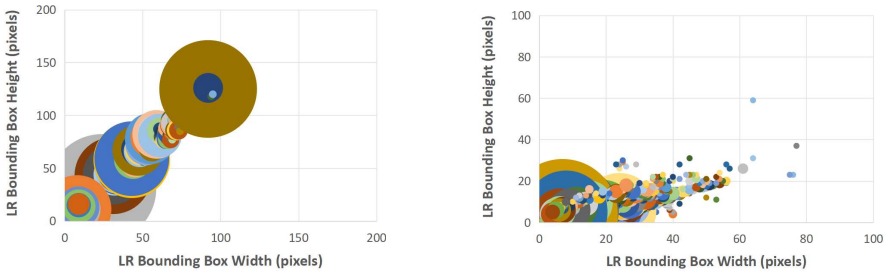
表 1 总结了数据特征，图 1 显示了 LR 边界框的分布。附加的数据特征在补充材料中给出。目前的面部版本是 FANVID-Faces-1，文本版本是 FANVID-Text-1，计划在社区参与下发布更大更具挑战性的未来版本。图 2 展示了一些原始 HR 帧、它们的 LR 版本以及带注释的边界框的例子。

Table 1: FANVID 数据集特征总结

Characteristic	FANVID-Faces-1	FANVID-Text-1	Notes
Frames	19,281	11,815	
LR frame size	320×180	320×180	Fixed size.
Bounding boxes	19,344	11,932	For statistical validity.
Video clips	997	466	Diverse backgrounds.
Entities	63	39	
Test entities	40	25	Challenging disambiguation.
Categories	6 races, 2 genders	3 countries	Challenging diversity.
Normalized Gini coeff.	0.192 (race), 0.308 (gender)	0.436 (country)	Plates from US dominate.

3.3 数据和代码可用性

为了促进透明度和负责任的数据使用，FANVID 为每个任务单独分发一个 CSV 文件。这些文件中的每一行对应一个具有以下字段的边界框：



(a) FANVID 人脸的边界框尺寸　　(b) FANVID 板的边界框尺寸

Figure 1: 在 FANVID 的所有 LR 帧中，边界框尺寸分布（频率由气泡面积表示）表明大多数面部和车牌对于人类识别来说太小。

Figure 2: 来自 FANVID 数据集的示例帧，包括用于参考的原始 HR，降低分辨率的 LR（为清晰起见已重新调整大小），以及用于人脸或车牌的标注框。

- 视频网址：访问原始高分辨率的 YouTube ® 视频。
- 元数据：HR 源视频的分辨率（通常为 1020p 或 720p）和帧率。
- 帧号：适用于边界框的原始视频中的帧索引。
- 裁剪坐标（可选）：指定需要裁剪的坐标，如果仅通过简单缩小不足以使面部或文字无法识别。
- 低分辨率帧大小：所有片段的目标低分辨率固定为 320×180 。
- 边界框坐标：注释实体在 LR 帧中的位置。
- 实体标识符：个人的姓名（针对面部）或转录的车牌文本。
- 实体类别：人脸的种族和性别，以及车牌的颜色和国家。
- 划分：用于指示某个实体是否应该用于训练（包括验证）或测试。

为了实现跨场景的人脸匹配，我们还提供了一个额外的文件，其中的行对应于视频数据集中每个人，并附有一个链接，该链接指向从与视频不重叠的场景中获取的单个公开可用的高分辨率头像。

为了尊重原始视频所有者的版权，FANVID 不重新分发视频文件。为了实现可重复性，我们提供了用于帧提取、裁剪、降采样和基于提供的 CSV 文件的注释的代码。FANVID 还遵循“数据集数据表”的最佳实践进行文档记录，概述了收集方法、许可、预期用途和伦理考虑。
[1]

我们还发布了一些工具，用于扩展数据集以进行额外的训练，例如添加新的身份、文本或序列。这包括用于注释流水线自动化部分的完整代码库，包括 VSR、人脸检测和匹配、车牌检测和跟踪，以及结构化注释生成。然而，人工验证仍然是确保注释质量的关键。

我们建议从业人员避免将同一个人或车牌的相关数据（如视频、面部照片）分割成训练和测试部分。这能防止在识别结果中出现特定身份的过拟合。我们还建议每个实验运行多次，比如通过不同的随机权重初始化，以报告标准误差。

4 任务、指标和基线结果

FANVID 定义了两个具有挑战性的识别任务，这些任务要求在存在干扰因素和每帧信息有限的情况下对低分辨率（LR）视频序列进行识别。与以往专注于帧级分类或高分辨率（HR）图像的基准不同，这些任务要求进行时间聚合和稳健的跨域匹配。在本节中，我们将描述识别任务，概述我们的评估指标，并展示基准结果。

4.1 FANVID 任务

FANVID 被设计用于两个现实世界的监控和法医识别任务：

给定一组高分辨率的参考图像（例如，面部肖像）和一个或多个包含面部的低分辨率查询视频，目标是在视频中检测和识别目标个体。每个视频可能包括多个不在图库中的干扰面孔，并且单个帧的分辨率太低，无法在没有时间聚合的情况下进行可靠识别。图库中的每张面部照片必须与所有位置和所有片段进行匹配，以模拟在跨场景中寻找失踪人员的真实搜索。因此，该任务测试模型在跨分辨率和跨域条件下执行多对一面部匹配的能力。

在这个任务中，目标是在每一帧中检测车牌区域，并从低分辨率视频中转录车牌文本，不使用外部数据库。车牌在很多帧中可能部分被遮挡或运动模糊，这使得字符级的时间推理变得必要。车牌必须在不使用它们的真实号码进行匹配的情况下被检测出来。因此，误报不会受到惩罚。与传统的光学字符识别 (OCR) 基准不同，这个任务不假设随时可以获取高质量的静态图像或完整的车牌可见性。

[1] 请注意，本文中描述的数据集目前正在积极生成和完善中。最终版本以及全面的文档和代码将在完成后公开发布于 `https://huggingface.co/datasets/kv1388/FANVID/` 。

4.2 评估指标

我们为人脸识别和文本识别各提出了一种指标。该指标源于交并比大于 0.5 的平均精度（mAP@0.5），这种做法对合理的边界框偏差比较宽容，但对识别错误给予惩罚。虽然识别任务假定类的数量是固定的，但我们的人脸匹配指标（也称为验证）需要适应数据库中参考的标准照数量和低分辨率帧中人脸数量的变化。同样，我们的文本识别指标必须结合字符级精准度，并根据车牌文本长度进行标准化。

基于这些动机，我们提出如下度量。设真实框由 $i \in \{1, ..., n\}$ 索引，检测到的框由 $j \in \{1, ..., m\}$ 索引。设 $I_{i,j} \in [0, 1]$ 为它们的 IoU ，$F_{i,j} \in \{0, 1\}$ 为它们的人脸身份匹配指标，而 $T_{i,j} \in [0, 1]$ 为基于注释的文本匹配分数。二进制人脸匹配指标和连续文本匹配分数之间的差异允许在没有车牌文本或图像数据库的情况下容忍小的文本识别错误，这与可以访问身份数据库相对照。我们实现了 $T_{i,j}$ 作为逆归一化编辑距离，即 1 减去编辑距离（从检测到的文本到真实文本）与真实文本长度的比率（在大写并仅删除字母数字字符之后）。对于每一个真实框，我们简单地将与彼此重叠最多的检测框（随机打破平局）的匹配指标设置为 $M_{i,j} = \mathbb{1}_{I_{i,j} \geq 0.5} \times \mathbb{1}_{j = argmax_k(I_{i,k})}$ ，前提是重叠至少为 0.5（非重叠的真实框，如 FANVID 中的，确保每个检测框最多可以有一个对应的真实框，其 IoU>0.5），其中 $\mathbb{1}_{condition}$ 是当条件为真时取值为 1 的指示函数。这给出了假阴性指示符为 $N_i = \mathbb{1}_{max_j(M_{i,j})=0}$ ，假阳性指示符为 $P_j = \mathbb{1}_{max_i(M_{i,j})=0}$ 。

提供了一个评估脚本，用于以一致的方式将检测表中的框与真实值表中的框进行比较，以实现所提出的度量标准。

因此，所提出的用于人脸检测和匹配的度量是：

$$FaceRecBox = \frac{\sum_{i,j} M_{i,j} F_{i,j}}{\sum_{i,j} M_{i,j} + \sum_i N_i + \sum_j P_j}, \tag{1}$$

由于缺乏可能的车牌号码访问，所提出的文本检测和识别评价标准的分母排除了误报计数，如下所示：

$$TextRecBox = \frac{\sum_{i,j} M_{i,j} T_{i,j}}{\sum_{i,j} M_{i,j} + \sum_i N_i}. \tag{2}$$

根据 NeurIPS 数据集指南，我们发布：

- 一个元数据丰富的 CSV 文件，链接视频来源、注释记录和图库引用，
- 帧提取和注释验证脚本，
- 低分辨率退化生成管道，
- 兼容所提出的 IoU 容忍度指标的评估脚本，

FANVID 的所有组件均设计为轻量级、可扩展和可重现，并且整个推理基线可以在不到半天的时间内在单个 A100 80GB GPU 上运行。研究人员可以通过提供的注释代码和降级模块使用其他片段来扩充数据集。完整数据集和工具将在发表时以 CC-BY-NC 许可证发布。

4.3 基线识别流程和结果

为了建立强大且模块化的基线，我们设计了一条管道，将用于视频超分辨率、物体检测和识别的模型结合在一起——每个模型都选自体现资源受限部署的最新实践。

对于所有识别任务，低分辨率视频输入首先通过记忆高效的视频超分辨率模型 RCDM [29] 进行处理。RCDM 使用可变形卷积进行帧间对齐，使用二维小波分解实现结构稀疏性，并使用时间记忆单元进行一致的帧增强，如图 3 所示。它生成一个超分辨率的帧序列，作为下游模块的输入。

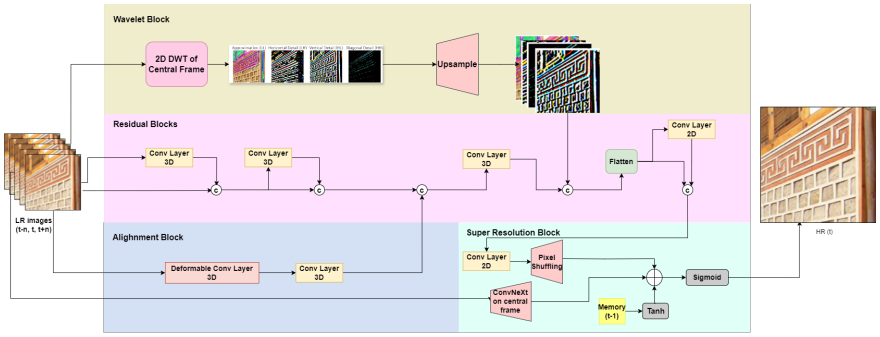


Figure 3: 残差 ConvNeXt 可变形卷积与记忆（RCDM）VSR 模型的架构 [29]，它集成了 ConvNeXt 模块、可变形对准模块、多级小波特征提取和时间记忆传播，用于低资源视频增强（经许可转载）。

对于人脸识别，我们使用 RetinaFace [7] 进行逐帧检测，然后使用 ArcFace [6] 进行身份嵌入。通过对帧间特征向量进行平均，并使用余弦相似度将得到的视频级嵌入与嫌疑犯照片库进行匹配。该方法在运动模糊、遮挡和尺度变化方面表现出鲁棒性，并且只需进行最小的时间平滑处理。

对于车牌识别，我们使用了 YOLOv10 [31] 来进行边界框检测，利用其低延迟架构和改进的空间精度。使用 SAM2 [25] 追踪后续帧中的车牌区域，以克服模糊、部分遮挡和干扰车牌的错配问题。分割后的区域被传递给 EasyOCR [8]，这是一种多语言文本识别引擎，用于提取字符级别的文字记录。对跨多个帧追踪的车牌的文本记录进行了时间平滑，以获得更好的识别结果。

流水线的模块化使得可以对组件和特性（例如时间融合）进行消融实验。提供了基准代码和设置，以确保可重复性。

基线实验的结果如表 2 所示

如上表所示，即便在低分辨率和运动模糊的情况下，测试方法也能提供一个强有力的基线，但仍显然需要更好的方法来识别低分辨率视频中的人物和文本。失败的情况通常涉及严重的遮挡或极端的光照情况。车牌识别错误与帧级别的字符模糊和车牌弯曲有关。某些成功和失败的例子在补充材料中展示。

5 局限性

FANVID 虽然提供了一个用于识别低分辨率视频的新数据集，但存在几个局限性。应用的降解通过双三次下采样合成生成，并未捕捉额外的挑战，如压缩伪

Table 2: 在 FANVID 上的基线结果显示，没有使用 VSR（RCDM）时，结果受到显著影响

Dataset	Split	Metric	Value	Technique
FANVID-Faces-1	Test	FaceRecBox 1	0.58	Pre-trained RCDM +RentinaFace+ArcFace
FANVID-Text-1	Test	TextRecBox 2	0.42	Pre-trained RCDM +EasyOCR
FANVID-Faces-1	Test	FaceRecBox 1	0.19	Pre-trained RentinaFace+ArcFace
FANVID-Text-1	Test	TextRecBox 2	0.15	Pre-trained EasyOCR

影、恶劣天气以及低光条件下的颗粒感。该数据集来自英语国家，对全球剧本和场景背景的代表性有限。尽管每帧都提供了标注，但并不包括遮挡标签。高分辨率参考是从不同领域精选的头像照片，这可能无法反映业余或警察摄影的现实查询条件。此外，虽然我们提供脚本而不是重新分发原始视频，原始内容仍在所有者的自由裁量范围内公开访问，需要负责使用以符合许可和隐私指南。

6 结论与未来工作

FANVID 引入了一个新的基准，用于低分辨率视频识别，反映了现实世界的监控条件——在这种情况下，单个帧是极其不足的，身份线索仅通过时间整合才会显现。在这方面，FANVID 不同于之前的数据集，这些数据集专注于没有其他面孔或文本干扰的静态，高分辨率图像。我们定义了两个任务——从低分辨率视频中进行人脸匹配和车牌识别，并使用手动验证的注释和高分辨率参考。为了确保可重复性和扩展性，我们发布了用于数据提取、注释、基准方法（结合视频超分辨率、检测和识别）和指标评估的组织良好且有文档的代码。我们还提供了与这些代码配合使用的电子表格，其中包含视频链接、其降尺度和裁剪参数、带有实体标签的边框坐标、实体多样性类别和测试划分指示符。我们提供了有鼓励性的基线结果，突出多帧识别方法的优势。我们鼓励探索可能甚至不需要超级分辨率作为中介步骤的端到端时间模型。

仍有几个方向尚未探索。联合学习检测和视频身份识别的统一架构（例如，基于变压器的时空骨干网络）可以提高鲁棒性。数据集中的干扰因素丰富、部分遮挡的序列引发了对耐遮挡和少样本学习方法的研究。此外，可以收集具有其他现实挑战的视频——例如，压缩伪影、恶劣天气、低光噪声——或本身低分辨率的视频，前提是可以可靠地标记原始实体。

通过在现实的、低保真条件下进行识别，并提供一个可重复、可扩展的基准，FANVID 为推进既有效又负责任的时间感知识别系统奠定了基础。

7

更广泛的影响和伦理考量

FANVID 完全由公开可用的视频和图像构建，没有再分发任何受版权保护的材料。相反，我们提供脚本以直接访问这些资源，符合现行规范。所有边界框和标签均经过人工验证以确保标注质量。

我们并不是试图最小化不同身份之间的视觉重叠，而是有意包含具有自然相似性的个体——比如相似的肤色、发型或年龄组——以反映现实世界中的视觉

模糊性。这样的设计鼓励模型学习更为稳健的、时空基础上的身份表征，而不仅仅是过度依赖表面的视觉线索。同时，我们致力于保持种族、性别和年龄的多样性。数据集包括来自不同种族和性别的个体，我们的策展流程确保这些属性上的比例代表。

所有视频都经过降采样以防止单帧识别，从而缓解隐私风险。我们强烈反对未经正式审查和未遵守隐私法规的情况下在实际监控或执法环境中使用 FANVID。FANVID 仅用于在真实世界视觉限制下进行公平、可解释且注重隐私的识别研究。

FANVID 的目标是成为发展负责任的识别技术的基础工具。我们支持与 AI 伦理和研究社区的持续对话，并将根据不断演变的社会标准对数据集、协议和文档进行迭代。

References

[1] BBC News. Custody photos too poor for facial recognition software. *BBC News*, March 2025. URL: `https://www.bbc.com/news/articles/cdrx7ry3dl4o`, Accessed: 2025-05-14.

[2] J. Ross Beveridge, P. J. Phillips, David Bolme, Bruce A. Draper, Geof H. Givens, Yui Man Lui, Mahesh Teli, and Hao Zhang. The challenge of face recognition from digital point-and-shoot cameras. In *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*, pages 1–8. IEEE, 2013.

[3] Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman. VGGFace2: A dataset for recognising faces across pose and age. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 67–74. IEEE, 2018.

[4] Kelvin CK Chan, Shangchen Zhou, Xintao Xie, Tak-Wai Hui, and Chen Change Loy. BasicVSR++: Improving Video Super-Resolution with Enhanced Propagation and Alignment. In *CVPR*, 2022.

[5] Debayan Cheng, Shaogang Gong, et al. Surveillance face recognition challenge. In *ICB*, 2018.

[6] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In *CVPR*, 2019.

[7] Jiankang Deng, Jia Guo, Yuxiang Zhou, Jinke Yu, Irene Kotsia, and Stefanos Zafeiriou. RetinaFace: Single-Shot Multi-Level Face Localization in the Wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[8] Jaided AI. EasyOCR: Ready-to-use OCR with 80+ Languages Supported. `https://github.com/JaidedAI/EasyOCR`, 2020.

[9] Gabriel Gonçalves, Matheus Weber, and Thiago Oliveira-Santos. A Benchmark for License Plate Character Segmentation. In *IJCNN*, 2018.

[10] Mislav Grgic, Kresimir Delac, and Sonja Grgic. SCFace–surveillance cameras face database. *Multimedia Tools and Applications*, 51:863–879, 2011.

[11] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. MS-Celeb-1M: A Dataset and Benchmark for Large-scale Face Recognition. In *ECCV*, 2016.

[12] Gary B Huang, Matt Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. *University of Massachusetts, Amherst, Technical Report*, 2008.

[13] Zhenyao Huang, Chen Wang, Ying Wang, Tieniu Tan, and Junzhou Huang. A Benchmark and Comparative Study of Video-based Face Recognition on COX Face Database. In *IEEE Transactions on Image Processing*, 2015.

[14] Pierre-Luc Jodoin, Guillaume-Alexandre Bilodeau, and Robert Bergevin. Tracking and recognition of cars and people in VIVA challenge. In *CVPR Workshops*, 2016.

[15] Dimosthenis Karatzas, Lluis Gomez-Bigorda, Anguelos Nicolaou, Suman Ghosh, Andrew D Bagdanov, Masakazu Iwamura, Jiri Matas, et al. ICDAR 2015 competition on robust reading. In *ICDAR*, 2015.

[16] Ira Kemelmacher-Shlizerman, Steven M Seitz, Daniel Miller, and Evan Brossard. The MegaFace benchmark: 1 million faces for recognition at scale. In *CVPR*, 2016.

[17] Jingyun Liang, Yuchen Zhang, Guolei Sun, Shiyi Gu, Luc Van Gool, and Radu Timofte. VRT: A Video Restoration Transformer. In *CVPR*, 2022.

[18] Xin Li, Tao Yang, Xiaoguang Hu, and Zheng-Jun Zha. RVRT: Residual Video Restoration Transformer. In *ECCV*, 2022.

[19] Ce Liu, Deqing Sun, and William T Freeman. Bayesian adaptive video super resolution. In *CVPR*, 2011.

[20] Yang Liu, Shuai Yan, Xiao Wang, and Jie Shao. Dependency-aware attention control for video face recognition. In *ECCV*, 2018.

[21] Brianna Maze, James Adams, James Duncan, Nathaniel Kalka, Tim Miller, Charles Otto, Anil Jain, et al. IARPA Janus Benchmark–C: Face dataset and protocol. In *International Conference on Biometrics (ICB)*, 2018.

[22] Seungjun Nah, Seokil Baik, Sungyong Hong, Gyeongsik Moon, Seungjun Son, Radu Timofte, and Kyoung Mu Lee. NTIRE 2019 challenge on video super-resolution: Methods and results. In *CVPR Workshops*, 2019.

[23] OpenALPR. OpenALPR benchmark dataset. `https://github.com/openalpr/openalpr`, 2017. Accessed 2024-05-10.

[24] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *ECCV Workshops*, pages 17–35. Springer, 2016.

[25] Anonymous. SAM2: Segment Anything Model with Temporal Consistency. *arXiv preprint arXiv:2401.xxxxx*, 2024. To appear.

[26] S. Silva and C.R. Jung. License plate detection and recognition in unconstrained scenarios. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.

[27] Amanpreet Singh, Vivek Natarajan, Yash Jiang, Xinlei Chen, Marcus Rohrbach, Dhruv Batra, and Devi Parikh. Towards VQA models that can read. In *CVPR*, 2019.

[28] Andreas Veit, Tomas Matera, Lukas Neumann, Jiri Matas, and Serge Belongie. COCO-Text: Dataset and benchmark for text detection and recognition in natural images. In *CVPR*, 2016.

[29] Kavitha Viswanathan, Pathak, Shahswat and others, et al. Low-Resource Video Super-Resolution using Memory, Wavelets, and Deformable Convolutions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

[30] Xintao Wang, Kelvin C.K. Liang, Chao Dong, Ying Shan, et al. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. In *ICCV Workshops*, 2021.

[31] Ao Wang, Hui Chen, Lihao Liu, Kai Chen, Zijia Lin, Jungong Han, and Guiguang Ding. YOLOv10: Real-Time End-to-End Object Detection. *arXiv preprint arXiv:2405.14458*, 2024.

[32] Tianfan Xue, Baian Chen, Jiajun Wu, Donglai Wei, and William T Freeman. Video enhancement with task-oriented flow. In *IJCV*, 2019.

[33] Zhenbo Xu, Cheng Yang, Lei Sun, et al. Towards End-to-End License Plate Detection and Recognition: A Large Dataset and Baseline. *arXiv preprint arXiv:1802.10130*, 2018.

[34] Sergey Zherzdev and Alexey Gruzdev. Lprnet: License plate recognition via deep neural networks. In *ECCV Workshops*, pages 261–266. Springer, 2018.

[35] Liang Zheng, Zhiwu Bie, Yunchao Sun, Jingdong Wang, Chi Su, Shengjin Wang, and Qi Tian. MARS: A video benchmark for large-scale person re-identification. In *ECCV*, 2016.