Graphical Abstract

GVCCS：用于可见全空相机序列中航迹识别和追踪的数据集

Gabriel Jarry, Ramon Dalmau, Philippe Very, Franck Ballerini, Stephania-Denisa Bocu

Highlights

GVCCS：用于可见全空相机序列中航迹识别和追踪的数据集

Gabriel Jarry, Ramon Dalmau, Philippe Very, Franck Ballerini, Stephania-Denisa Bocu

- Dataset with instance-level and temporally resolved annotations of contrails from ground-based videos.
- Unified contrail segmentation and tracking model using Mask2Former.
- Robust tracking of individual contrails over time, enabling analysis of their full lifecycle.

GVCCS：用于可见全空相机序列中航迹识别和追踪的数据集

Gabriel Jarry[a], Ramon Dalmau[a], Philippe Very[a], Franck Ballerini[a], Stephania-Denisa Bocu[a]

[a]*EUROCONTROL. Aviation Sustainability Unit (ASU), Aerodrome Centre Bois des Bordes, Brétigny-Sur-Orge, 91220, Essone, France*

Abstract

航空的气候影响不仅包括 CO_2 的排放，还包括显著的非 CO_2 效应，特别是由于凝结尾迹产生的影响。这些冰云可以改变地球的辐射平衡，其潜在的升温效应可能可以与航空 CO_2 的效应相媲美。基于物理的模型提供了对凝结尾迹形成和气候影响的有用估计，但其准确性在很大程度上依赖于大气输入数据的质量以及用于表示复杂过程（如冰粒子形成和湿度驱动的持续存在）的假设。来自遥感设备的观测数据，如卫星和地面摄像机，可以用于验证和校准这些模型。然而，现有的数据集未能探索凝结尾迹动力学和形成的所有方面：它们通常缺乏时间追踪，也没有将凝结尾迹归因为其源航班。为了解决这些限制，我们提出了地面可见摄像机凝结尾迹序列（GVCCS），这是一组通过地面全景摄像机在可见光范围内记录的新的开放数据集。每个凝结尾迹都单独标记并随时间追踪，从而允许对其生命周期进行详细分析。该数据集包含 122 个视频序列（24,228 帧），并为在摄像机上方形成的凝结尾迹提供了航班标识。作为参考，我们还提出了一个统一的深度学习框架，用于利用全景分割模型进行凝结尾迹分析，该模型在单个架构中执行语义分割（凝结尾迹像素识别）、实例分割（个体凝结尾迹分离）和时间追踪。通过提供高质量、时间分辨的注释和模型评估的基准，我们的工作支持更好的凝结尾迹监测，并将促进物理模型更好的校准。这为更准确的气候影响理解和评估奠定了基础。

Keywords: environmental impact, contrails, open data, computer vision

1. 介绍

航空对全球气候变化的贡献不仅来自二氧化碳（CO_2）排放，还包括一系列非 CO_2 效应，其中包括氮氧化物（NO_x）、水蒸气和气溶胶。在这

些影响中，凝结尾迹（航迹云），即飞行器在典型巡航高度形成的冰晶云，由于其潜在巨大但不确定的辐射影响而尤为突出。虽然它们经常在天空中出现为短暂的白色条纹，但持久的航迹云可以扩展成广泛的卷云状的云层，这些云层可以捕获向外辐射的长波辐射，使地球变暖。最近的研究表明，航迹云卷云的气候强迫与航空 CO_2 排放（Lee et al., 2021; Teoh et al., 2023）的量级相当，尽管这取决于所使用的度量标准（Borella et al., 2024）。

然而，准确评估航迹云对气候的影响仍是航空和气候科学家面临的一个重大挑战。航迹云的生命周期取决于复杂而相互关联的过程，如冰核化、晶体生长、风驱动的扩散和与天然云的相互作用，这些过程对环境大气条件非常敏感。温度和湿度（特别是相对于冰的相对湿度）的微小变化可以决定航迹云是迅速消散还是持续存在并扩散。这种敏感性，加上昼夜变化的辐射强迫（白天反射阳光时产生冷却效果；晚上捕获红外辐射时产生加热效果），使得航迹云的净效应既依赖于具体情境又极难可靠地建模。

传统上，航迹云的影响通常使用物理模型进行研究，但遥感和计算机视觉的最新进展现在提供了一种有价值的观察视角。基于物理的模型，如航迹云卷云预测模型（CoCiP）(Schumann, 2012) 或 APCEMM (Fritz et al., 2020)，通过求解描述飞机排放物与大气条件之间相互作用的复杂方程来模拟航迹云的生命周期。这些模型提供了有价值的理论见解，但它们的准确性在很大程度上依赖于输入数据的质量 (Gierens et al., 2020)。关键参数，如大气温度、湿度和飞机引擎特性，通常是不确定的，这些不确定性在计算过程中逐步影响到结果的可靠性。此外，航迹云微物理和辐射效应的详细模拟可能对计算提出很高的要求，特别是当应用于全球尺度的分析时。

使用卫星和地面图像的观测方法提供了一种直接且数据驱动的方式来研究飞机尾迹，与理论模型相辅相成。高分辨率遥感和计算机视觉的进步使这些方法越来越有效 (Meijer et al., 2022; McCloskey et al., 2021; Ng et al., 2023; Chevallier et al., 2023)。除检测外，观测数据未来应在通过提供经验验证和校准上述不确定参数来完善基于物理的模型方面发挥越来越大的作用。

结合观测数据与自动相关监视广播（ADS-B）等空中交通信息和气象数据，对于推进我们对卷云生命周期和气候影响的理解具有重要的前景。将卷云与具体航班关联，其中详细参数（例如发动机类型、飞行高度和大气条件）已知，这将有助于更好地理解这些参数在卷云形成和动态中的作用。然而，实现这种整合需要解决基础性的挑战：在图像和/或视频中准确识别卷云，将其与自然云区分开（语义分割），检测个体实例（实例分割），以及追踪其随时间的演变。本文侧重于这些关键的第一步，开发针对卷云分割和追踪的稳健方法，用于单个地面摄像机图像和视频。虽然特别使用静止卫星 (Chevallier et al., 2023; Riggi-Carrolo et al., 2023; Geraedts et al., 2024; Sarna et al., 2025) 进行大规模属性归因仍然非常具有挑战性，我们的工作

提供了可靠地局部检测和追踪卷云的必要工具，为后续与航班和气象数据的整合奠定了基础。

尽管对观测到的凝结尾迹分析的兴趣日益增长，但公开可用的数据集仍然有限。最普遍使用的资源，谷歌的 OpenContrails，仅在中心的 GOES-16 图像上提供实例级的标注，周围的图像没有进行标注，这阻碍了对凝结尾迹的时间追踪。相比之下，(Sarna et al., 2025) 介绍了 SynthOpenContrails，它将合成的凝结尾迹和注释叠加到真实场景中，提供了完整的逐帧定位、跟踪和飞行归属，展示了即便是合成的凝结尾迹叠加而不是人工注释，也存在详尽标注的数据。理想情况下，将是一个所有帧都有人为标注、每个凝结尾迹都被分配了一个持久标识符的完全标注的视频数据集。

为了推动该领域的研究，本文介绍了地面可见相机凝结尾迹序列 (GVCCS)，这是一个具有实例级别注释的开放数据集 (Jarry et al., 2025)，来源于法国 Brétigny-sur-Orge 的地面视频录制（Réuniwatt CamVision 可见光地面相机）。我们的数据集包括 122 段视频（时长从 20 分钟到 5 小时不等），总帧数约为 24,200，每个帧都注释了实例级别的标签。通过公开提供这个数据集，本文为大气和航空研究界提供了宝贵的基准。

为了支持未来的性能比较，我们在此引入了一种基于深度学习的航迹线分割和跟踪模型。与其依赖于这些任务的独立模型——通常需要复杂的、特定组合的技术——我们采用一个基于 Mask2Former 的统一框架，这是一种先进的计算机视觉模型。Mask2Former 专为泛型分割设计，结合了语义分割（为每个像素分配一个类别标签，例如“航迹线”或“天空”）和实例分割（区分个体对象，例如不同的航迹线）。除了将航迹线与晴空分离之外，它还可以处理复杂的背景，如部分或完全遮挡航迹线的低空云层，通过分配适当的“云”标签，同时仍保持独特的实例识别。例如，在单幅图像中，泛型分割能够识别所有可见的航迹线像素，正确标记间隔的云，并为每条航迹线分配一致的实例掩码，即便它们重叠、交叉、看起来破碎或透过薄云可见。事实上，由于大气条件和自然消散过程，航迹线常常会分裂成多个不连续的部分。一个强大的监测系统不仅必须识别这些碎片，还需要将它们关联到正确的航迹线实例上。

值得注意的是，仅凭图像或视频进行凝结尾迹分析时，碎片化带来了显著挑战：需要在没有外部数据的情况下，将同一航班的视觉上不连续的段落进行分组。而且，低空云层遮挡和阳光眩光可能进一步中断或掩盖凝结尾迹的连续性，即使是同一个物理凝结尾迹也会产生多多边形标注。然而，在实际操作中，可以首先执行单多边形实例分割，然后使用辅助数据（如飞机轨迹和风场）将多个实例与同一航班关联。这个后处理步骤使得能够根据航班身份而非视觉连续性，在时间和空间上进行分组。在这项工作中，我们限制自己仅进行基于图像的分析，并将外部数据源的整合留待未来工作。

Mask2Former 最初是为单个图像设计的，可以很容易地扩展到视频数据，以提高跨帧全景分割的一致性。通过利用时间信息，视频版的 Mask2Former 以集成的方式执行语义分割、实例分割和跟踪。在本文中，我们研究了基于帧和基于视频的两种版本的 Mask2Former，并比较了它们在我们数据集上的性能。

本文的其余部分结构如下。第一部分 2 提供了关于航迹云形成和计算机视觉技术的必要背景，为本文所涉及的挑战奠定了基础。第二部分 3 综述了航迹云数据集和分割模型的相关工作，突出了当前的局限性并激发了我们的研究方法。第三部分 4 介绍了我们新开发的视频数据集，详细说明了其标注方法和独特的实例级结构。第四部分 5 描述了基于 Mask2Former 架构的全景分割框架。第五部分 6 展示并分析了实验结果。最后，第六部分 7 总结了我们的主要贡献并概述了未来的研究方向。

2. 背景

本节介绍了理解本研究中所解决的挑战所必需的关键概念。我们首先概述凝结尾迹形成背后的物理过程及其对气候的影响，重点讨论为什么凝结尾迹特别难以检测和跟踪。然后，我们回顾相关的计算机视觉技术，特别是目标检测和图像分割，并评估它们分析凝结尾迹的适用性。

2.1. 凝结尾迹的科学

凝结尾迹是飞机后方形成的人造云，当热的、湿的发动机废气与巡航高度（通常在 8 到 12 公里之间）的冷、低压空气混合时就会出现。如果大气条件适宜，具体地说，如果温度降至 $-40°$ 摄氏度以下并且空气足够湿润，废气中的水蒸气会凝结并冻结成冰晶。这个过程由施密特-阿普尔曼准则 (Appleman, 1953) 建模和量化，产生了在天空中可见的熟悉的细长白色尾迹。某些凝结尾迹会迅速消散，而另一些则持续存在并扩散，最终形成较大的冰云结构，被称为凝结尾迹卷云。

像自然云一样，凝结尾迹影响地球的辐射预算：它们会捕获向外的长波辐射，导致变暖，同时反射进入的太阳辐射，产生冷却效果。净结果取决于凝结尾迹的高度、光学特性、寿命和一天中的时间。精确的相对影响取决于选择的气候指标（Borella et al., 2024）；然而，人们认为凝结尾迹对气候的变暖作用与航空的 CO_2 排放在同一数量级上（Lee et al., 2021; Teoh et al., 2023）。这使得监测和描述凝结尾迹成为了解航空的完整环境足迹（Teoh et al., 2023) 和制定缓解策略（Teoh et al., 2020) 的一个重要部分。

如上所述，观测视角提供了一种替代观点，其重点是直接在大气影像中检测和分析凝结尾迹。然而，检测和跟踪凝结尾迹存在若干技术挑战，这也解释了为什么这一主题的研究兴趣日益增长。卫星影像通常缺乏在凝结

尾迹早期阶段进行检测所需的空间和时间分辨率 (Ng et al., 2023) 。静止卫星的名义空间分辨率约为 0.5 到 2 公里，时间分辨率为 5 到 15 分钟，这往往不足以捕捉新形成的凝结尾迹狭窄、微弱且短暂的特性，除非它们得以持续并增长。即便凝结尾迹确实扩展成可检测到的云结构，也很难将其与自然卷云区分开来，尤其是在具有复杂云层的场景中。此外，当凝结尾迹在卫星图像中可见时，它通常已经漂移并变形，使得将其归因于产生它的航班变得复杂 (Chevallier et al., 2023; Sarna et al., 2025) 。这种联系至关重要，因为识别出起源的航班能让研究人员检索诸如飞机型号和发动机型号等关键细节，这些是通过与实证观测比较来评估凝结尾迹环境影响和改进物理模型的关键输入。

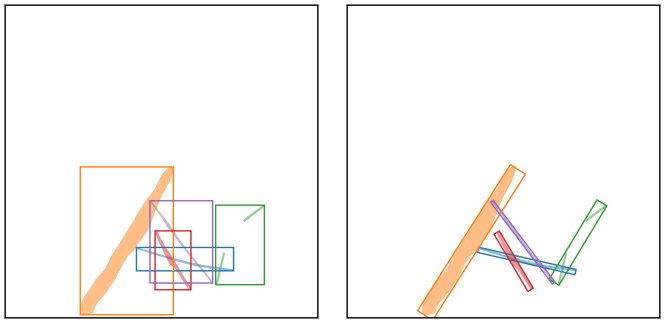
地面摄像机 (Schumann et al., 2013; Low et al., 2025) 提供了一个具有关键优势的补充视角。这些系统位于航线下方，可以比卫星捕获到更高分辨率的图像和视频，具有更高的空间和时间保真度。更重要的是，它们可以在凝结尾迹刚形成时立即检测到，这时凝结尾迹仍然很薄、线性且视觉上很明显。这种早期可见性简化了将观察到的凝结尾迹与具体负责的航班关联的任务，尤其是在结合精确的轨迹数据时。主要的缺点当然是覆盖范围有限，这妨碍了从凝结尾迹形成到消散的监控能力。

虽然这不是本文的重点，但一个有前景的方向是将地面和卫星观测结合到一个统一的监测框架中。在这样的系统中，首先在高分辨率的地面图像中检测到航迹云，并利用航迹和气象数据将其归因于特定的航班，从而获得重要的飞机和发动机参数。关键是，为了实现超出地面摄像机有限视野的连续跟踪，需要将这些航迹云可靠地与卫星图像中随着漂移、扩展和老化而演变的对应物联系起来。成功地将地面和卫星这两种模式下的航迹云关联起来，将允许在保留有关具体飞机和航班信息的同时，监测其从形成到消散的整个生命周期。

2.2. 用于航迹云监测的计算机视觉技术

由于凝结尾迹的形状细长、曲率多变，并且容易随时间碎裂或消退，因此对计算机视觉来说是具有挑战性的视觉目标。这些特征使它们在本质上与常规目标检测基准中处理的对象根本不同，例如 Common Objects in Context（COCO）数据集中车辆和动物等明确的、独立的对象。

传统上，目标检测方法使用边界框来定位目标，通常是轴对齐的矩形。这种方法对像汽车或动物这样紧凑且大致为矩形的物体效果很好，但对航迹云表现不佳。一个轴对齐的边界框可能会不小心包括多个航迹云段或大量背景天空，同时遗漏弯曲或碎片状轨迹的部分。旋转边界框通过允许旋转可以在一定程度上改进，能够更好地适应延长的航迹云的几何形状。然而，它们在捕捉细粒度的形状、间隙或消退的段落方面仍显不足。图 1 显示了轴对齐和旋转边界框用于航迹云目标检测的局限性。



(a) 轴对齐边界框　　(b) 有向边界框

Figure 1: 航迹线的边界框检测示意图。每个检测到的航迹线都用不同的颜色突出显示。注意，拉长或碎裂的航迹对边界框的对齐和分离构成了挑战。

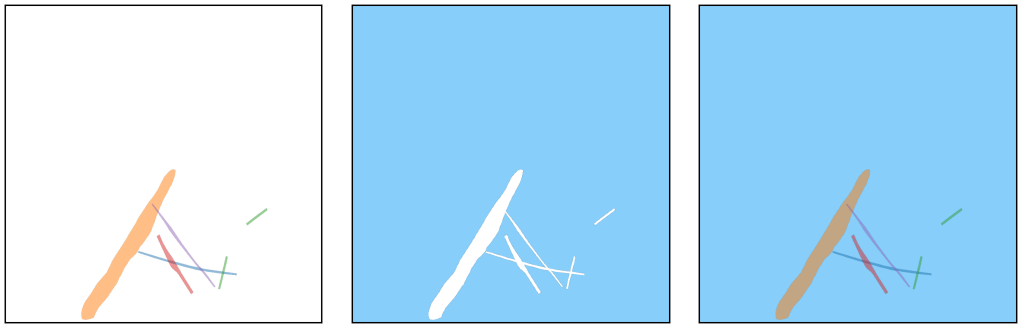
实例分割通过为每个单独的对象预测像素级的掩膜，提供了一种更精确的解决方案。对于凝结尾迹来说，这种方法尤其有益，因为它即使在尾迹相交、重叠或不均匀消散时，也可以准确地描绘出每个尾迹。例如，即使两个重叠的凝结尾迹以不同的速度消退，它们仍然可以分配给不同的实例。

语义分割则是通过类别对每个像素进行标注，例如，“航迹云”或“天空”，但不会区分单个航迹云。在研究具体航迹云的时间演变或相互作用时，这种方法是不够的，因为它将所有航迹云视为一个无法区分的类别。

全景分割结合了两种方法的优势：它为每个像素分配一个类别标签（语义分割）并在适当的地方分配一个实例标识符（实例分割）。在这个框架中，”things”（例如个别的飞行尾迹）被分配唯一的实例标签，而”stuff”（如背景天空或自然云）仅按类别进行标记。这种统一视角非常适合于尾迹监测，使得可以在更广泛的大气上下文中进行细粒度的个别尾迹分析。此外，该框架可以轻松扩展到其他类别（例如，卷层云，积云）以实现更全面的场景理解。不过，前提是这些类别在数据集创建期间已经被有效和一致地标记，这为标注工作带来了额外的复杂性。图 2 展示了实例分割、语义分割以及全景分割方法。

文献中一个重要但常被忽视的问题是如何几何地表示航迹云。实际上，一个单一的航迹云可能由几个不相连的段组成，例如，由于消退或遮挡，使其成为一个多多边形形状。例如，参见图 2a 和 2c 中的绿色航迹云。然而，最自然的方法是通过将每个段作为一个独立的多边形来简化处理，实际上假设每个片段属于不同的航迹云。

虽然这种简化避免了直接处理多多边形的复杂性，但它引入了一个重要的挑战：要重构完整的凝结尾迹，必须找到将分散的部分连接在一起的方法。这需要即兴的连接策略，它们在复杂性和准确性上各有不同。有些方



(a) 实例 (b) 语义 (c) 全景

Figure 2: 对用于示例性航空尾迹的分割方法的比较。不同的颜色表示不同的航空尾迹实例或类别，具体取决于所用的方法。

法纯粹依赖于碎片的几何特性，例如它们的接近性或对齐，而另一些则结合了外部数据，如飞机航线或气象信息，以进行更明智的联系。

在这项工作中，我们采用全景分割作为分割和跟踪凝结尾迹的基础。这一选择的动机是其能够同时实现实例级的精确度并维持对周围场景的上下文感知。此外，通过明确解决碎片化凝结尾迹的问题，我们的方法可以在不需要外部信息来源（如飞行或天气数据）的情况下，实现对凝结尾迹的实例级识别。这在此类数据可能不可用或不完整的情况下尤其有价值。然而，我们还探索了模型的另一种版本，该版本将每个凝结尾迹碎片视为一个独立的实例，假设后续算法将利用外部交通和气象数据，之后会将这些片段与其对应的航班相关联。这两种策略的比较评估，即自包含的实例识别和外部支持的后关联，将在未来的出版物中展示。在本文中，我们仅专注于展示凝结尾迹分割模型本身。

3. 最新技术

本节概述了航迹线分割和分析的先前工作，首先关注于支持这项研究所开发的数据集，然后是用于航迹线分割和飞行归因的计算模型。概述了现有数据集的范围和关键特征，特别注意到时间标注和飞行归因真实数据的有限可用性。随后，我们考察了最先进的分割和跟踪方法，特别是基于深度学习的方法，评估它们在航迹线分析中的适用性和性能。该综述突出了当前研究中的差距，并激发了本文所提出的贡献。

3.1. 数据集

最近，航迹云检测的进展得益于注释数据集的发展，这些数据集主要基于卫星图像。这些数据集促进了计算机视觉技术在航迹云识别中的应用，

尽管大多数情况下，时间连续性和与飞行元数据的整合等方面仍然有限。在本节中，我们回顾了最相关的公开可用数据集，并将我们的贡献置于此背景中。

据我们所知，(Kulik, 2019) 和 (Meijer et al., 2022) 是首次利用现代、数据驱动的深度学习框架进行大规模飞机尾迹分割的研究。作者开发并应用了卷积神经网络，并使用一个由 100 多张手动标注并具实例分割的地球静止 GOES 卫星图像组成的手动整理数据集进行训练。

Google Research 领导了第一个大规模航迹云检测标注工作之一，首先开发了一个基于高分辨率 Sentinel 卫星图像的航迹云数据集 (McCloskey et al., 2021) 。人类专家使用结构化指南手动标注了这些图像，为每个可见的航迹云段生成多边形掩码。多位标注人员对每张图像进行了独立标注，数据集包含的所有个体注释，可选择按多数共识筛选结果。这种方法提高了标签的空间精度和整体质量。

基于这项工作，谷歌发布了 OpenContrails 数据集 (Ng et al., 2023) ，该数据集基于 GOES-16 先进基线成像仪（ABI）拍摄的图像。得益于 GOES-16 静止轨道提供的 10 分钟时间分辨率，该数据集非常适合研究大规模的飞行航迹云。OpenContrails 通过包括围绕每个标注帧的未标注图像的短序列来提供时间上下文，为标注者提供了更准确标注的宝贵信息。每个序列中只有中心帧被标注，因此无法直接将航迹云动力学与物理模型进行比较。值得注意的是，2025 年的更新引入了实例级别的标签，使得该数据集可以用于基于实例的模型并拓展其在更高级航迹云分析中的潜力。

Gourgue et al. (2025) 引入了一个开放获取的语料库，其中包括约 1600 幅在巴黎附近的 SIRTA 大气实验室获取的多边形标注的半球天空图像，提供区分“年轻”、“旧”和“非常旧”航迹云以及几种混淆伪影的类别标签。通过在航迹云形成后几分钟内捕获高分辨率的地面视图，该数据集填补了卫星基准中存在的时空空白。

(Low et al., 2025) 没有创建一个用于在分割任务上训练现代卷积网络的数据集，而是手动标注了从 CoCiP 模型应用和他们的广角地面相机系统观测所获得的航迹航点之间的对应关系。这种方法特别适合直接评估和参数化物理模型。

据我们所知，Meijer et al. (2024) 是第一个将两种不同遥感器上的图像进行配准的数据集示例：他们专门为航迹高度估计组建了一个数据集，包括 2018 年至 2022 年期间美国本土的 3000 多个案例。航迹首先通过 GOES-16 ABI 红外图像的自动检测进行定位，然后与 CALIOP 激光雷达剖面精确配准，纠正视差和风偏。团队随后对匹配的图像进行了人工检查，以验证和校准对齐。这个基准数据集将静止轨道航迹特征与高分辨率垂直剖面相链接，从而支持有监督的深度学习方法从 ABI 数据中预测航迹的顶高度。

在航迹云检测方面的一个显著进展是合成标签数据集的开发。(Cheval-

lier et al., 2023) 使用 CoCiP (Schumann, 2012) 生成了一个合成数据集，将航迹云多边形叠加到 GOES-16 影像上，从而开启了首个用于航迹云检测的实例分割流程。航班分配算法的性能通过实际的 GOES 数据进行了验证，但更多是依靠人工检查而非合成参考地面实况。在这一合成基础上，(Sarna et al., 2025) 引入了一个基准数据集，SynthOpenContrails，该数据集包含与已知航班元数据相关联的合成航迹云检测序列，首次提供了定量评估和改进航迹云与航班归因算法的机会。据我们所知，这是唯一能提供具备归因地面实况的定位和跟踪航迹云的数据集，尽管是合成的。虽然使用合成数据集代表了一种现代化和尖端的算法训练技术，但使用人工标注的数据作为测试集仍然在理论上更可取，以客观评估算法性能。然而，在这阶段获得这种地球同步卫星影像数据集，由于其分辨率较低，非常困难，这也激发了作者们采用的这种方法。如在 (Sarna et al., 2025) 中所述，从原则上讲，使用更高分辨率的低轨卫星或地面相机基于人工标注获取用于航班归因的参考数据集是完全可行的，这也是本研究的重点。

总的来说，虽然现有的数据集提供了有价值的资源，但缺乏含有时间上分解的、实例级别的和飞行归属注释的综合性人工标注数据。我们的工作通过引入一个数据集来解决这个问题，该数据集是使用我们的地面摄像头系统收集的，旨在提供这些注释。

3.2. 模型

航迹监测的计算机视觉技术最早在九十年代初期由 (Forkert et al., 1993; Mannstein et al., 1999) 开创，使用的是非数据驱动的图像分析技术。他们的工作应用线性核方法、亮度温差通道的直接阈值化以及早期为线性形状检测优化的霍夫变换算子 (Pratt, 2007) ，以识别 AVHRR 卫星图像中的航迹云。(Vazquez-Navarro et al., 2010) 和 (Duda et al., 2013) 进一步改进了这种方法。

据我们所知，Kulik (2019); Meijer et al. (2022) 代表现代卷积网络在像素级分类和语义分割中的最早应用。基于 OpenContrails 数据集，Ng et al. (2023) 使用语义分割算法，特别是 DeepLabV3 (Chen et al., 2017, 2018) ，通过亮度温度差异在 灰烬-RGB 复合材料 中识别凝结尾迹。他们的研究表明，通过 3D 编码器添加时间上下文，结合时间维度，性能得到了提高。此外，随后的 Kaggle 竞赛 表明，采用现代 transformer 骨干网的 UNet 模型 (Ronneberger et al., 2015) ，如 MaxViT (Tu et al., 2022) 和 CoatNet (Dai et al., 2021) ，取得了更强的结果 (Jarry et al.) 。

通过采用集成方法，Ortiz et al. (2025) 结合了六个神经网络，包括 U-Net、DeepLab 和 transformer 架构，并应用基于光流的修正来保持连续卫星帧之间的时间一致性。同时，Sun and Roosenbrand (2025) 引入了一种霍

夫空间线感知损失用于少样本场景，结合了全局对齐项来辅助 Dice 损失，以促进预测与线性结构对齐。

从像素级的掩膜转向实例级的航迹云分割并利用合成数据，(Chevallier et al., 2023) 引入了第一个专注于航迹云检测实例分割的算法流程，使用了 Mask R-CNN 算法 (He et al., 2017) 。同样地，Van Huffel et al. (2025) 采用了 Mask R-CNN 来处理由其广角地面摄像系统捕获的图像。

在地球同步卫星影像中将检测到的航迹归因于个别航班（通常使用 ADS-B 信息）这一困难任务一直是近来几项研究的关注焦点。(Chevallier et al., 2023) 引入了一条管道，该管道结合了航迹检测、跟踪，并使用几何标准和风修正后的轨迹与飞机匹配。Riggi-Carrolo et al. (2023) 提出了考虑航班数据和大气条件不确定性的概率匹配方法，同时利用基于霍夫变换的线检测。(Geraedts et al., 2024) 提出了一种可扩展系统，旨在大规模地将航迹分配给航班，从而实现航迹形成的常规监测并支持气候评估。(Sarna et al., 2025) 系统性地对这些归因算法进行了基准测试和完善，强调了常见的挑战并提出改进的关联指标，基于合成生成的 SynthOpenContrails 数据集的发布。

相比之下，我们的工作以地面图像为目标，能够在航迹云形成后立即进行捕捉，并通过 ADS-B 数据实现几乎即时的航班归因。我们利用在高分辨率视频上训练的 Mask2Former 进行全景分割，以提取单个航迹云的像素精确蒙版，并随着时间的推移进行跟踪。这弥补了航迹云早期检测的空白，提供了比现有基于卫星的模型更丰富的空间和时间细节。

4. 数据集

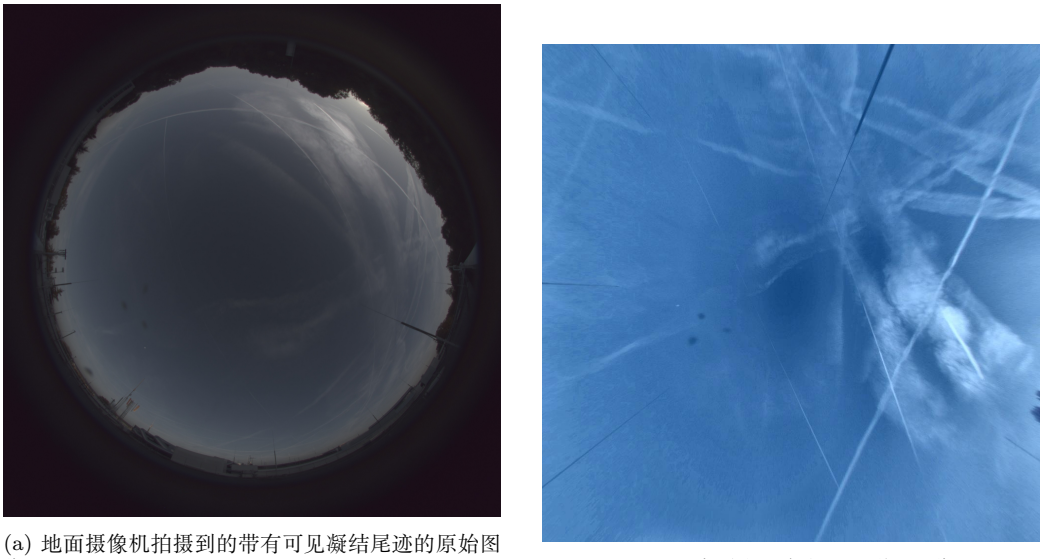
本文的主要贡献是引入一个新的数据集，该数据集旨在支持航迹云的检测、追踪和归因。本节提供了数据集的详细概述。第 4.1 节描述了数据收集和标注活动。第 4.2 节总结了数据集的结构和内容。

4.1. 数据收集和标注活动

为了支持用于航迹云检测的机器学习模型的开发，我们在 ContrailNet 项目中进行了广泛的标注活动。通过安装在 EUROCONTROL 创新中心屋顶的全天空基相机获取了可见光谱图像序列，该相机以 1976×2032 像素的分辨率每 30 秒捕捉一次天空。

我们的摄像机供应商 Reuniwatt 提供了一套双全景天空摄像系统：第一台设备 CamVision 在可见光谱中工作，每 30 秒捕获一次高分辨率鱼眼图像，具有车载处理和自校准功能，确保即使在灰尘或潮湿条件下也能可靠的白天操作。第二台设备 SkyInsight 使用长波红外 (8–13 μm) 成像，通过镀铬半球镜进行成像，将在未来的研究中使用。

首先将原始的全天空图像几何投影到一个方形网格上。此投影过程使用特定相机的校准文件，将每个像素与其对应的方位角和天顶角相关联，有效地去除了镜头畸变，并将天空重新映射到一个统一的笛卡尔表示上。一个在固定云层高度（10 公里）处的地理参考点网格被计算出来，并使用线性插值方案将原始像素值分配到投影帧中。输出的是一个大小为 1024×1024 像素的方形图像，保留了相机上方天空的空间几何。



(a) 地面摄像机拍摄到的带有可见凝结尾迹的原始图像。

(b) 用于标注的几何投影和增强图像。

Figure 3: 并排比较原始地面相机图像和用于标注的几何投影版本。投影修正透视畸变并增强对比度，以便更好地进行分割。

为了提高序列的视觉清晰度和一致性，每个投影图像都会经历一个三步增强过程。首先，使用线性缩放操作增加亮度。其次，通过 CLAHE（对比度限制自适应直方图均衡）增强局部对比度，从而提升像航迹云这样的特征，而不过度曝光图像。最后，通过重新平衡蓝色和红色通道减少色彩的暖色调，从而改善在强光条件下的航迹云可见性。这一预处理流程被证明在突出细微的航迹云结构上至关重要，尤其是在复杂的大气场景中。图 3 显示了原始和增强后的投影图像。

视频序列的标记过程已被应用。每个序列包含 60 到 480 帧图像，对应的持续时间从 30 分钟到 4 个小时，从而能够在整个形成和消散阶段对航迹云进行时间跟踪。

标注过程是使用由 Encord 开发的专用注释工具进行的，该公司还提供了一支专业的标注团队。我们通过定期的协调会议与该团队保持密切合作，在此期间开发并反复完善了标注指南。该标注平台被专门配置为在摄像机

视野上方叠加飞行轨迹数据，帮助标注员识别“新”凝结尾迹，这些尾迹是在摄像机视野内形成的，并且可以显著地与已知的飞机轨迹相关联。相反，“旧”凝结尾迹被定义为那些在序列开始时已经存在或可能是在摄像机视野外形成的，这使得飞行关联变得不可能。

每条航迹线都使用高精度多边形进行了标注，这些多边形跟踪其整个可见演变过程中的空间范围，从早期的线性阶段到扩展的晚期阶段。当航迹线被云层遮挡或部分阻断时，使用多个多边形并通过关系属性（`Fragmented contrail` 和 `Cloud obstruction` ）进行链接以保持时间连续性。

为了确保最高的标注质量，该活动引入了一个多阶段的审核协议。首先进行了一个初步校准阶段，使用一个样本数据集来统一解释并识别边界情况。每个标注序列随后经历了一个两步质量控制过程：先由标注团队进行技术审查，然后由 EUROCONTROL 进行专家审查以确保最终质量。总共进行了 4,536 小时的标注和 431 小时的审查工作。

4.2. 数据集描述

GVCCS 数据集 Jarry et al. (2025) 是首个开放访问的、实例级别注释的视频数据集，专为从可见地面天空相机图像中检测、分割和跟踪航迹而设计。它由法国 Brétigny-sur-Orge 的 EUROCONTROL 创新中心使用 Réuniwatt 的 CamVision 传感器捕获的 122 段高清视频序列（共计 24,228 张图像）组成。每个序列都经过精心注释，具有时间一致的多边形蒙版，用于可见航迹，包括多实例跟踪，并在可能的情况下，使用飞机轨迹数据归因于特定的航班。

总体而言，标注团队标记了 4651 条单独的飞机尾迹，总共 176,194 个多边形。序列涵盖的持续时间范围广泛（每条飞机尾迹从 0.5 到 142.5 分钟不等），每个飞机尾迹包含 1 到 589 个多边形（平均：37.8）。平均而言，每段视频序列跨度为 96.6 分钟，包含大约 193 张标注的图像。大约 3346 条飞机尾迹与从筛选超过 15,000 英尺的同步飞行轨迹数据中获取的唯一飞行标识符相关联。

GVCCS 数据集被分为 `train/` 和 `test/` 文件夹，每个文件夹包含 `images` 、`annotations.json` （COCO 格式）以及 `parquet` 格式的相关飞行数据。该数据集支持多种研究任务，包括语义和全景分割、时间跟踪、生命周期分析以及航迹云与飞行的归因，并在 CC BY 4.0 许可下发布。

5. 分割模型

本节回顾了用于识别（以及某些情况下跟踪）航迹云的分割模型。我们关注两个模型系列：Mask2Former，一种基于变压器的最先进的分割模型；

Table 1: 标注的飞机尾迹数据集的描述性统计

Metric	Value
Total sequences (labelled)	122
Total images	24,228
Average sequence duration in minutes	96.6
Images per sequence (min / max / mean)	41 / 600 / 198.6
Total annotated contrail instances	4651
Total unique flight IDs assigned	3354
Total polygons annotated	176,234
Contrail duration in minutes (min / max / mean)	0.5 / 142.5 /14.6
Polygons per contrail (min / max / mean)	1 / 589 / 37.8
Polygons per frame per contrail (min / max / mean)	1 / 4.5 / 1.2

以及使用判别嵌入损失的 U-Net。两者都在单个图像上进行了评估，而只有 Mask2Former 还在视频上进行了额外评估。

我们还探讨了两种问题表述：在单多边形的情况下，每个可见的凝结尾迹碎片被视为一个独立的实例；在多多边形的情况下，给定凝结尾迹的所有碎片都被标记为一个单一实例，即使它们在空间上是不连接的。单多边形的设定假定一个后续的链接算法（未在此工作中实现）可以将碎片组合成完整的凝结尾迹。相比之下，多多边形的表述期望模型能够隐式地推断出这种分组。

5.1. Mask2Former

Mask2Former 是一种通用分割架构，在单一模型中统一了语义、实例和全景分割。它围绕一个层级编码器-解码器结构构建，由三个主要组件组成：用于多尺度特征提取的卷积主干、生成密集空间嵌入的像素解码器，以及带有可学习掩码查询的变换器解码器，该解码器迭代地优化分割预测。

Mask2Former 的一个核心创新在于其在 transformer 解码器中使用所谓的掩码注意力。与考虑整个图像的标准交叉注意力不同，掩码注意力将注意力限制在当前预测掩码周围的区域。这样的局部聚焦能够更精确地优化物体边界，这对于像飞机尾迹这样的细长、高纵横比结构尤为有利。模型的可学习查询作为物体提议，并通过多个解码层进行优化，以端到端的方式生成最终的实例掩码和类别标签。

Mask2Former 效果的重要方面在于其损失函数（即训练目标），该函数引导模型学习准确的分割掩码及其对应的类别。Mask2Former 使用的损失

函数结合了多个部分。首先，它使用分类损失来帮助模型为每个预测的掩码分配正确的类别（例如，航迹云与天空）。其次，它包含一个掩码损失，该损失衡量预测掩码与该对象的真实掩码的相似程度，通常使用逐像素的二元交叉熵或 Dice 损失。最后，Mask2Former 结合了基于匈牙利算法的匹配步骤，以最佳一对一的方式将预测与真实值对齐。这确保每个预测掩码与最合适的参考对象进行评估，避免重复分配。

本文不包括对该模型的详细技术描述，因为我们的重点是将 Mask2Former 应用于航迹云分割；对于该架构的全面概述及其在流行数据集上的表现，我们建议读者参考 Cheng et al. (2022) 的原始工作。

为了捕捉航迹云演变中固有的时间动态，我们扩展了 Mask2Former 以处理短视频序列。尽管该模型设计用于处理单张图像，但它可以通过将时间视为空间维度旁的一个附加轴，处理多个连续帧作为一个三维时空体积，这遵循了 Cheng et al. (2021a) 引入的扩展。

与传统的分割模型相比，Mask2Former 在架构上具有显著优势。虽然 Mask R-CNN (He et al., 2017) 的效果不错，但其将检测和分割作为独立的阶段来执行，这可能导致空间的不对齐和效率低下，尤其是在分割长而不连续的物体时。尽管 DETR（DEtection TRansformer） (Carion et al., 2020) 是端到端的且基于 transformer，但主要关注于目标检测，缺乏精确掩码预测所需的细粒度空间建模。MaskFormer (Cheng et al., 2021b) 引入了基于 transformer 的分割解码，但依赖于全局注意力，这可能导致空间精度下降。Mask2Former 通过掩码注意力和迭代精细化改进了这种方法，从而提高了准确性，尤其是在物体常常细、淡以及视觉上模糊的挑战性任务中。

5.2. 具有判别损失的 U-Net

作为基线，我们实现了一个两步实例分割模型。首先，我们使用经典的 U-net 架构 Jarry et al. 进行分割。U-Net 是专门为图像分割任务设计的，具有对称的编码器-解码器结构。网络的编码器部分逐渐减少输入图像的空间尺寸，提取捕捉整体上下文的高级特征。然后，解码器通过对这些特征进行上采样逐渐恢复空间分辨率，以生成与原始图像尺寸匹配的分割图。重要的是，U-Net 使用了快速连接，直接连接编码器和解码器中的对应层。这些连接允许在下采样过程中丢失的细粒度空间细节被恢复，从而提高分割输出的质量和精度。

其次，我们使用了类似的架构，该架构通过使用判别性损失函数为图像中的每个像素学习一个独特的特征表示或嵌入。在这个模型中，U-Net 的最终层并不生成带有类别标签的典型分割图。相反，它为每个像素生成一个嵌入，即一个高维特征空间中的向量。目标是让属于同一对象实例的像素拥有相似的嵌入（意味着它们在这个特征空间中彼此接近），而不同实例

的像素具有相距较远的嵌入。通过这种方式，模型有效地学会根据其学习的特征对像素进行分组。

识别单个实例的过程分为两个独立的步骤。第一步是使用 U-Net 生成这些像素嵌入，而第二步是将这些嵌入分组或聚类成单个实例。对于聚类，我们使用 HDSCAN 算法来寻找聚类，并使用最终的 k-means 将异常值与最近的聚类关联。

用于训练模型的判别损失由三个部分组成。第一部分称为 “拉近项”，促使属于同一实例的像素嵌入彼此靠近，从而使集群紧凑。第二部分称为 “推开项”，迫使不同实例的嵌入彼此足够分离，从而防止集群重叠。第三部分是一个正则化项，防止嵌入的幅度变得过大，这有助于稳定训练过程和嵌入空间。这样的组合使得模型能够学习到有意义且分离良好的像素嵌入，而无需依赖显式的对象边界框或预定义的区域提案。对于感兴趣于判别损失的数学公式和详细原理的读者，我们建议参考原论文 Brabandere et al. (2017) 。

需要指出的是，该模型仅适用于单张图像。与前一节中提到的视频模型如 Mask2Former 不同，它不包含任何时间或顺序信息，也不包括处理视频的递归层或机制。将这一方法扩展到处理视频序列并融入时间一致性，将需要对架构和使用的算法进行重大更改，这超出了本研究的范围。

基于嵌入的方法非常适合于分割那些可能在空间上不连续的物体，例如形状支离破碎的凝结尾迹。由于该模型不需要空间连续性，如果这些分离且不连接的部分具有共同的视觉特征并属于同一标签，它可以学习将同一凝结尾迹的不同部分嵌入到特征空间的相似区域。然而，这种方法也面临一些挑战。如果同一凝结尾迹的不同部分在外观上显著不同，可能是由于照明变化、大气条件变化或背景纹理差异等因素，它们可能会被不同地嵌入，并错误地分配到不同的簇。相反，视觉上相似但无关的凝结尾迹碎片可能会被错误地分组在一起，因为该模型完全依赖于学习到的嵌入进行聚类。

图 5 展示了实例判别分割模型的一个定性结果。左侧显示的是真实标签，突出显示了对卷云航迹实例的像素级分配。右侧展示的是相应的判别嵌入空间，通过主成分分析 (PCA) 降维至二维用于可视化。每个点代表一个像素嵌入，颜色指示其所属的实例。此可视化提供了模型如何通过判别损失训练来学习将同一实例的像素在特征空间中嵌入得更近，同时将不同实例的像素区分开的洞察。嵌入空间中观察到的分离证明了该模型将零散的卷云航迹结构聚类的能力，虽然在视觉上相似但不相关的片段仍可能由于共享的外观特征而在嵌入中局部重叠。

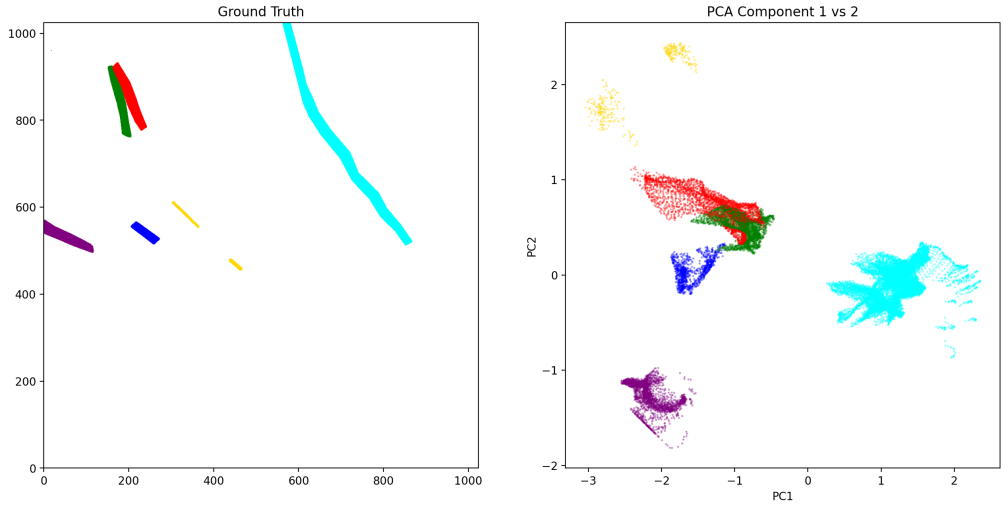


Figure 4: 真实标签显示在左侧，判别嵌入显示在右侧。后者是使用主成分分析（PCA）创建的。颜色表示航迹实例。

6. 结果

本节介绍了在尾迹分割任务中，前面第 5 节中引入的模型的性能。我们的主要目标不是取得最前沿的结果，而是建立明确的应用示例和有意义的基准性能。通过这样做，我们强调了该数据集所提供的独特机会，并为研究界提供了一个基础，鼓励在航空对气候影响这一关键领域的快速进展。

6.1. 训练

所有模型都是从现有的预训练检查点初始化的。我们为单一图像分割任务训练了两个版本的 Mask2Former 架构。这两个模型共享相同的核心架构，但在其 Transformer 主干的大小上有所不同：一个使用 Swin-Base（Swin-B）配置，另一个使用更大的 Swin-Large（Swin-L）。这两个模型之间的主要区别在于模型容量，Swin-L 具有显著更多的参数，使其能够以更高的计算需求为代价学习更丰富的表示。

两个图像模型都是从 Mask2Former 模型库 [1] 中公开可用的预训练检查点进行初始化的。每个模型首先在 ImageNet-21k (IN21k) (Ridnik et al., 2021) 分类数据集上进行预训练，然后在 COCO 全景分割数据集上微调。虽然 COCO (Lin et al., 2014) 不包含航迹云，但它涵盖了广泛的自然（包

[1] https://github.com/facebookresearch/Mask2Former/blob/main/MODEL_ZOO.md

括云和天空）和人为物体，提供了有用的通用分割特征。这种两阶段的预训练，即 IN21k 之后是 COCO，已在文献中被广泛验证，并为在航迹云图像上的微调提供了强有力的初始化。

我们训练了 Swin-B 和 Swin-L 两种变体，使用 200 个可学习的物体查询在单个图像帧上进行训练。鉴于我们的硬件配置，包括两张每张拥有 48 GB 内存的 NVIDIA RTX 6000 GPU，我们能够在图像数据集上训练这两种变体而没有显著的内存限制。

对于视频分割，我们使用了 Mask2Former 的特定视频变体，该变体扩展了原有架构以处理时间序列。与基于图像的模型类似，它也使用 200 个物体查询和 Swin Transformer 骨干网络，并由在 YouTubeVIS 2019 数据集上预训练的检查点初始化。尽管 YouTubeVIS 不包含航迹云，但其强调跨帧学习时间一致的物体掩码，使其非常适合在视频数据中捕捉航迹云的动态。由于 GPU 内存限制，我们将训练和推理限制在由少量连续帧组成的短视频片段中。虽然这个限制是为了适应可用的硬件资源，特别是对于内存密集型架构是必要的，但它也塑造了我们的训练策略。在训练过程中，这些片段是从较长的视频序列中随机抽样的，以引入时间多样性。通过改变抽样片段的起点，模型能够见到航迹云生命周期中不同阶段的情况，如形成、延伸、消散，以及在各种大气环境中的表现。这种随机抽样鼓励模型学习更具泛化能力的时间表示。

为了支持这一设置，我们使用 Swin-Base (Swin-B) 和 Swin-Large (Swin-L) 骨干网络训练了视频 Mask2Former 模型。然而，每个片段的帧数必须根据模型容量和内存可用性进行调整。对于较轻量的 Swin-B 变体，我们能够在 5 帧的片段上进行训练，而由于 Swin-L 模型显著更大的内存占用，我们只能在 3 帧的片段上进行训练。这反映了时间上下文与模型表达能力之间的权衡：较长的片段可能更好地捕捉航迹的动态演变，而像 Swin-L 这样较大的模型则提供了更丰富的每帧表示。训练这两种配置使我们可以探索这两个维度，即时间深度和模型容量，在航迹分割背景下的相互作用。

对于 U-Net 模型，我们使用了基于 MaxViT-B 的主干，这是一个混合视觉转换器架构，将卷积层与自注意力机制相结合，以实现高效且可扩展的视觉表示学习。该主干在 ImageNet-21k 上进行了预训练，随后在 ImageNet-1k 上进行了微调，提供了强大的特征表示，以支持在航迹云分割训练期间采用的判别性损失函数。

每个模型的训练过程涉及几次周期的监督学习，其中基于验证集的表现进行了早停。在视频级别上，数据集被随机分割为 70-10-20 的训练、验证和测试集。这意味着来自给定视频的所有帧被专门分配到三个集之一，以避免任何潜在的数据泄漏。为了确保公平和无偏倚的评估，我们还在三个子集中平衡了空序列（不含飞机尾迹的视频）的数量。

我们没有对任何模型进行详尽的超参数调整。相反，我们设计这一实验

的目标是建立基线结果，并在实际计算和数据限制下定性和定量分析模型性能。所有模型均使用其原始文献中报告的默认超参数进行训练。表格 2 和 3 总结了每个模型最重要的训练参数。请注意，模型在其架构和训练设置相关的具体超参数上有所不同。未来的工作将集中在探索更复杂的建模策略、系统的超参数优化以及额外的训练改进上。

Table 2: Mask2Former 模型的默认超参数。

Hyper-parameter	Default value	Notes / Differences
Training iterations	20K	Same for image and video
Learning Rate	—	3.75e-5 (Image), 1.25e-5 (Video)
Batch Size	—	6 (Image), 2 (Video)
Image Size	1024 × 1024	Same for image and video
Class Weight	2.0	Same for image and video
Mask Weight	5.0	Same for image and video
Dice Weight	5.0	Same for image and video
Importance Sample Ratio	0.75	Same for image and video
Oversample Ratio	3.0	Same for image and video
Augmentations	Rotation (90°), vertical flip, horizontal flip	Applied at image level (Image); applied at clip level (Video)

请记住，每个模型均在两种不同的实例分割任务中进行了训练和评估。第一种形式将一条航迹云视为一个单一对象，即使它由多个不连接的区域或碎片段构成。在此设置中，模型必须学习将视觉和空间上分离的区域分为同一个物理航迹云。第二种任务通过将每个可见多边形视为一个独立实例来简化问题。在这种形式中，模型不需要将属于同一航迹云的不相连段进行分组；相反，只需检测和分割每个不同的区域。这种方法对应于一种模块化的处理流程，其中实例合并和飞行归属将在后续阶段进行讨论，并将在未来的工作中讨论。

6.2. 评价

我们结合标准和任务自适应的指标来评估语义和实例级的分割性能。对于语义分割，我们报告像素级别的分数，如平均交并比和 Dice 系数。对于实例分割，我们采用 COCO 评估协议，并进行了修改，以更好地反映航迹云薄而细长的结构。所有指标都是在完整的测试集上全球计算的。在接下来的部分中，我们描述了我们的评估程序、视频模型的滑动窗口推理策略，以及我们选择指标背后的理由。结果的呈现和解释将在最后提供。

Table 3: 使用辨别损失训练的 U-Net 模型的默认超参数。

Hyper-parameter	Default value
Architecture	U-Net
Backbone	tu-maxvit_base_tf_512.in1k
Input image size	1024×1024
Precision	16-mixed
Epochs	100
Batch size	1
Gradient accumulation steps	32
Learning rate	5×10^{-6}
Optimizer	AdamW (weight decay = 10^{-4})
Scheduler	Cosine with warm-up
Augmentations	Rotation (90°), vertical flip, horizontal flip

6.2.1.

时间评估策略

对于基于视频的模型，推理是使用滑动窗口方法进行的，其中每个视频被划分为固定长度的重叠短片段，与训练期间使用的片段长度相匹配（例如，Swin-L 模型使用 3 帧，Swin-B 模型使用 5 帧）。这些片段每次前进一个帧（步幅为一），允许模型在推理期间有效利用时间上下文，同时考虑到内存限制。重要的是，分割准确率仅在每个短片段的中心帧上计算。该设计确保视频中的每一帧仅在其作为片段的中心帧出现时，才对评估指标有贡献，从而避免重复评估并实现与基于图像的模型的公平比较，后者是对单帧独立进行预测的。例如，如果一个 5 帧片段用于帧编号为 1 至 10 的视频，第一个评估片段跨越帧 1–5，并在帧 3 上进行评估；下一个片段覆盖帧 2–6（在帧 4 上评估），以此类推。这保证了对帧 3 至 8 的唯一评估，每个帧正好评估一次。

需要注意的是，基于视频的 Mask2Former 模型在每个剪辑中保持时间一致的实例标识符。也就是说，如果某一条尾迹在一个剪辑的一个帧中被标记为实例 # 3，那么在同一个剪辑的所有帧中，它都会保持这个标识符。然而，由于剪辑是独立处理的，这些标识符在连续的剪辑之间不一定保持一致。给定的尾迹可能在相邻的剪辑中获得不同的标识符。为了在整个视频中实现尾迹的连续跟踪，我们引入了一种简单的后处理方法，将这些实例标识符链接和协调，以生成连贯的、连续的轨迹；这种方法在 Appendix A 中详细描述。

6.2.2.

分割指标

模型性能通过语义和实例级别的分割指标进行评估。所有指标都是通过在整个测试集上聚合预测和真实值后进行的全局计算。这种全局计算防止了在每个观测（即帧）上独立计算指标后取平均值所带来的偏差，这在数据不平衡或稀疏（如航迹线分割）等情况下尤其重要。

对于语义分割，我们报告平均交并比（mIoU）和 Dice 系数。mIoU 通过计算预测和真实值二进制掩码的交集面积与并集面积的比率来衡量它们之间的重叠，因此对假阳性和假阴性进行惩罚。Dice 系数定义为重叠面积的两倍除以预测和真实值掩码的总面积，强调了正确的重叠，特别对细薄或碎片状结构敏感，使其成为评估航迹云的合适指标。

使用全局计算的 COCO 风格的指标来评估实例分割性能。为了适应尾迹所带来的特定挑战，我们调整了 IoU 阈值范围，并用以下符号表示指标：AP@[IoU 范围 | 尺寸类别 | 最大检测数]，其中 IoU 范围指明计算平均精度（AP）或平均召回率（AR）的 IoU 阈值范围，尺寸类别表示考虑的对象大小子集，最大检测数为每张图像考虑的最大检测数。例如，AP@[0.25:0.75 | all | 100] 表示在 IoU 阈值从 0.25 到 0.75 之间计算的平均精度，考虑所有对象大小并最多检测每张图像中的 100 个目标。对象尺寸类别遵循 COCO 风格指标中使用的标准定义：小对象的面积小于 $32^2 = 1,024$ 像素；中等对象的面积范围在 32^2 到 $96^2 = 9,216$ 像素之间；大对象的面积超过 96^2 像素。像 AP@[0.25:0.75 | small | 100] 这样的指标则反映的是在指定的 IoU 和检测限制下，对小尺寸对象的性能。

我们将 IoU 阈值范围限制在 [0.25, 0.75]，而不是标准的 COCO 范围 [0.50, 0.95]，以更好地适应飞机尾迹的细长和纤细几何形状，其中非常高的 IoU 阈值过于严格。飞机尾迹是细长且不规则的，可能延伸到图像的大部分区域，因此使得精确的掩模重叠具有挑战性。在典型的 COCO 指标下，一个预测与部分但语义上正确的重叠可能会被不公平地惩罚。例如，一个预测的掩模仅与 30 % 的飞机尾迹重叠的情况下，在 COCO 默认的最小 IoU 为 0.5 时会被忽略，但在我们更宽松的阈值下会被视为真正的正例。

通过调整 IoU 范围，这些指标更好地反映了航迹云的实际分割质量，在对空间精度的敏感性与对该领域固有的轻微错位和碎片化的容忍度之间取得了平衡。需要注意的是，这些调整后的指标不能直接与标准 COCO 评分进行比较，而是专门定制来在航迹云分割的背景下提供有意义的评估。

该评估框架结合了全局计算的语义和实例分割指标，通过适当的 IoU 阈值和尺寸类别，提供了一种全面且可解释的方式来评估模型性能。它有助于跨模型的公平比较，并支持我们航迹云数据集的未来基准测试。

表格 4 和 5 分别总结了语义和实例分割任务的结果。所有结果都针对单

图像和基于视频的模型进行了报告。实例分割结果根据注释风格进一步细分：M 代表多多边形注释，而 S 代表单多边形注释。对于 Mask2Former 模型，没有括号的值对应于 Swin-B 骨干网络，而括号内的值对应于 Swin-L。

Table 4: 语义分割指标。对于 Mask2Former 的变体，没有括号的值指的是 Swin-B；括号中的值指的是 Swin-L。

	Single Images		Videos
Metric	Mask2Former	U-Net	Mask2Former
Dice	0.56 (0.60)	0.59	0.57 (0.59)
mIoU	0.38 (0.43)	0.42	0.40 (0.42)

在语义分割任务中，各个模型和变体的性能保持一致，Dice 和 mIoU 分数几乎没有变化。这种稳定性是预期的，因为语义分割只需将每个像素分类为尾迹或天空，而不需区分独立的尾迹实例。U-Net 模型的结果与更先进的 Mask2Former 模型相当，这表明每像素的尾迹检测主要依赖于局部视觉特征，如形状、亮度和纹理，而 U-Net 能够有效捕捉这些特征。

这些结果也反映了我们数据集的质量和一致性：尽管基于地面图像，分割性能与先前使用卫星数据的研究报告的结果一致 (Jarry et al.; Ortiz et al., 2025) 。虽然成像方式和场景几何的差异阻碍了直接比较，但结果的一致性表明语义航迹云分割对于现代架构来说是一个良好定义的任务，可以在各种数据源中实现强大的性能。

实例分割结果揭示了模型架构之间的明显差异。这些差异比在语义分割任务中观察到的还要显著，突出了由实例级推理带来的额外复杂性。专为全景分割设计的 Mask2Former，通过物体级查询和全局空间推理，在所有实例指标上稳定地优于 U-Net。在多多边形设置中，性能差距尤为明显，因为航迹云看起来是碎片状的，必须正确地分组成连贯的实例。这些结果强调了专为实例感知任务构建的架构的价值：Mask2Former 能够进行全局推理并关联不连续的片段，使其更适合检测和跟踪单个航迹云。

在评估基于图像和基于视频的 Mask2Former 模型时，会出现更为细致的比较。对于 Swin-B 主干网络，基于图像的模型在实例分割性能上表现更好，而基于视频的模型在语义分割指标上略胜一筹。这表明，尽管视频模型得益于时间一致性和运动线索，但由于要强制跨帧一致性所增加的复杂性可能带来挑战，从而略微影响到实例级别的预测准确性，特别是在使用类似 Swin-B 这样容量较小的主干网络时。

在 Swin-L 设置中，基于图像的模型整体表现最佳。它不仅实现了最高的实例分割得分，而且在语义分割性能上也略有优势。这些结果表明，时

Table 5: 实例分割指标。“M” 表示多多边形，而 “S” 表示单多边形。对于 Mask2Former 变体，未括号的值指的是 Swin-B；括号中的值指的是 Swin-L。

		Single Images		Videos
Type	Metric	Mask2Former	U-Net	Mask2Former
M	AP@[0.25:0.75 | all | 100]	0.34 (0.34)	0.05	0.31 (0.33)
	AP@[0.25:0.75 | small | 100]	0.21 (0.21)	0.01	0.14 (0.17)
	AP@[0.25:0.75 | medium | 100]	0.39 (0.40)	0.13	0.37 (0.38)
	AP@[0.25:0.75 | large | 100]	0.44 (0.47)	0.12	0.46 (0.47)
	AR@[0.25:0.75 | all | 1]	0.10 (0.10)	0.03	0.09 (0.09)
	AR@[0.25:0.75 | all | 10]	0.41 (0.41)	0.18	0.38 (0.40)
	AR@[0.25:0.75 | all | 100]	0.44 (0.44)	0.22	0.43 (0.44)
	AR@[0.25:0.75 | small | 100]	0.30 (0.30)	0.14	0.26 (0.29)
	AR@[0.25:0.75 | medium | 100]	0.50 (0.50)	0.25	0.49 (0.50)
	AR@[0.25:0.75 | large | 100]	0.55 (0.55)	0.22	0.57 (0.56)
S	AP@[0.25:0.75 | all | 100]	0.35 (0.37)	0.06	0.31 (0.34)
	AP@[0.25:0.75 | small | 100]	0.24 (0.26)	0.03	0.17 (0.21)
	AP@[0.25:0.75 | medium | 100]	0.44 (0.45)	0.14	0.41 (0.43)
	AP@[0.25:0.75 | large | 100]	0.37 (0.43)	0.11	0.46 (0.47)
	AR@[0.25:0.75 | all | 1]	0.08 (0.08)	0.03	0.07 (0.08)
	AR@[0.25:0.75 | all | 10]	0.37 (0.38)	0.18	0.35 (0.37)
	AR@[0.25:0.75 | all | 100]	0.44 (0.45)	0.21	0.42 (0.45)
	AR@[0.25:0.75 | small | 100]	0.33 (0.34)	0.15	0.28 (0.32)
	AR@[0.25:0.75 | medium | 100]	0.53 (0.53)	0.26	0.52 (0.55)
	AR@[0.25:0.75 | large | 100]	0.54 (0.56)	0.25	0.58 (0.60)

间建模并不总能带来性能提升，特别是在时间上下文有限（例如，3 帧片段）或模型的空间表示能力已经很高的情况下。基于图像的模型受益于在 COCO 上的预训练，这可能有利于精确的空间划分，而基于视频的变体依赖于在 YouTubeVIS 上的预训练，更注重时间一致性。然而，需要注意的

是，基于视频的模型执行了一个额外的任务：跟踪。通过在帧间保持一致的实例身份，它实现了基于时间的一致分割，这是基于图像的模型无法实现的。总体而言，这里报告的指标是基于每帧计算的，并未考虑时间上的闪烁或实例身份的一致性。这些时间方面在视频应用中特别重要，而这里呈现的传统帧级评估分数没有捕捉到这些方面。

总体而言，Swin-L 在所有设置中都表现优于 Swin-B，这进一步证明了增加模型容量对细粒度空间理解和实例级推理的好处。尽管如此，这也带来了更高的计算需求，特别是在视频设置中，强调了性能与可扩展性之间的权衡。

在评估中观察到的另一个重要趋势是模型性能受航迹云大小和检测上限的强烈影响。一般来说，较大的航迹云由于像素数量较多且歧义性较低，分割得更准确，而允许更多的预测实例（例如，增加检测上限）通过消除对可报告对象数量的限制来提高召回率。这些趋势与目标检测中的一般发现一致，并加强了航迹云分割与更广泛的实例分割任务之间的共同挑战。

比较多边形和单边形形式揭示了任务难度的差异：单边形设置本质上更容易。在所有模型和数据模态中，使用单边形形式时，实例分割指标始终较高。这是因为任务消除了将破碎或空间不连续的尾迹段分组成单独实例的需求。相反，不论分离与否，尾迹的所有部分都被视为一个单一的掩膜，大大简化了模型的目标。模型不再需要学习复杂的分组策略或在空间和时间不连续性上进行推理。请注意，在两种形式之间，语义分割指标几乎没有变化，这表明识别尾迹像素在两种情况下都同样可行。不同之处仅在于这些像素如何被分组成实例。这一区别证实，多边形任务的主要挑战不是像素分类，而是实例关联。

这些结果对不同的航迹云检测场景具有重要的实际意义。对于较老的航迹云，例如那些通常在卫星图像中观察到或在地面图像中观察到的航迹云，当航迹云形成在相机视野之外时，很难将航迹云与其源航班关联起来。在这些情况下，唯一可行的方案是基于视觉信息将可见的片段分组为实例。这使得多多边形实例分割变得至关重要，因为它允许模型在不依赖外部数据的情况下检测和关联不相连的航迹云段。我们的数据集和基于 Mask2Former 的模型专为这种设置而设计，即使在航迹云片段化、被遮挡或空间上断开的情况下，也能够进行有效的实例级别检测。

相比之下，当凝结尾迹直接在相机上方形成，并且有飞行轨迹和风场等附加数据可用时，可以采用不同的方法。在这些情况下，可以进行单多边形实例分割，其中的凝结尾迹片段被基于飞行路径和平流后处理关联为单个实例。从计算机视觉的角度来看，这种表述更为简单，并且在文献中常被使用 (Ortiz et al., 2025; Chevallier et al., 2023; Van Huffel et al., 2025)，主要是因为直到现在还没有可用的多多边形标注数据集。然而，这种方法依赖于外部数据的访问，并且仅适用于在飞机进入视野后于观测窗口期间

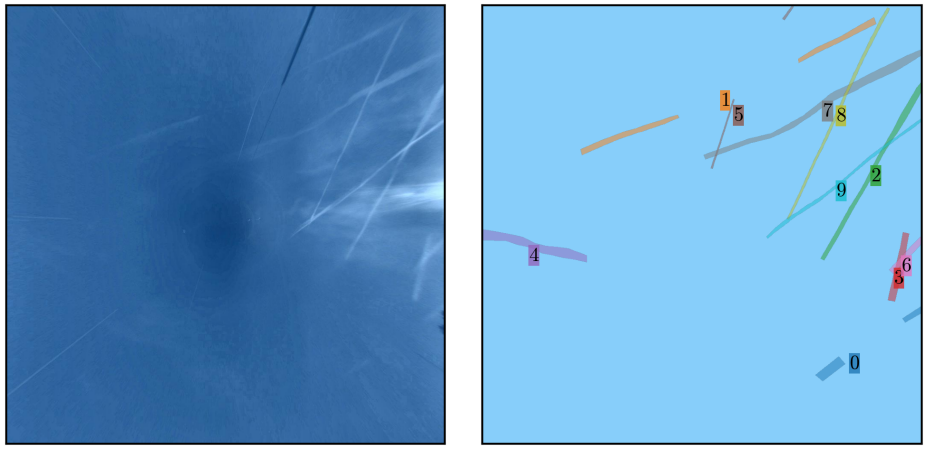
形成的凝结尾迹。

通过支持多多边形和单一多边形的公式，我们的数据集能够在更广泛的操作用例中进行训练和评估。多多边形任务对于仅基于视觉的较旧航迹云或卫星图像中的航迹云检测至关重要，而当附加元数据能够实现航迹云到航班归因时，单一多边形公式可能更为合适。该区别将在未来的工作中进一步探讨，重点是将航迹云与其来源飞机相联系。

6.3. 示例说明

我们展示了两个测试集示例，以说明多多边形航迹云分割任务的挑战。在这两种情况下，我们比较了基于图像和基于视频版本的 Mask2Former 模型的预测结果，这些模型从预训练的 Swin-L 骨干网络中训练得出。这些示例强调了时间上下文如何影响实例预测，并揭示了典型的失败模式，包括航迹云碎片化、被云遮挡以及在航迹云和视觉相似的云结构之间的混淆。

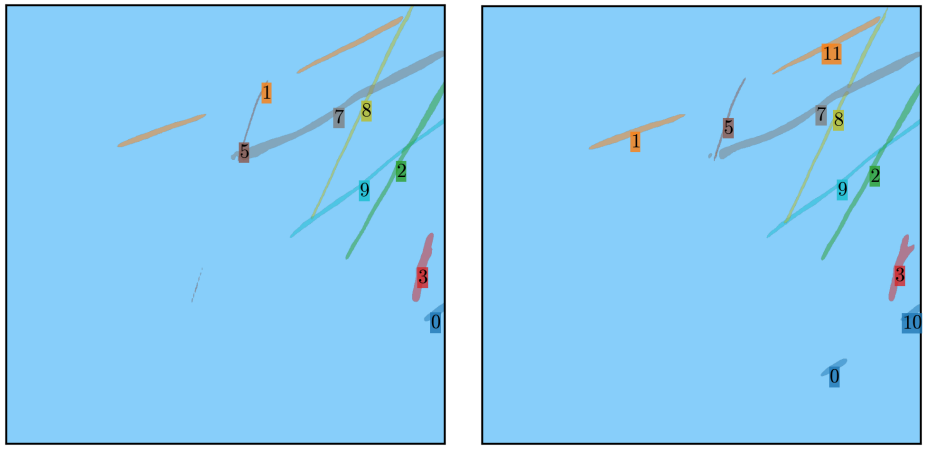
图 5 显示了 2024 年 4 月 25 日 05:51:00 的一个帧，在晴空条件下拍摄。背景是均匀的蓝色，为人类和机器分割提供了良好的条件。对应的真值标注包括若干被标记为断裂的凝结尾迹（例如，标识符 0、1 和 5），这些基于标注过程中可用的已知飞行轨迹。这使得该示例适合用于评估多多边形设置中的实例级理解。



(a) 原始图像。　　(b) 真实标注

Figure 5: 2024 年 4 月 25 日 05:51:00 的原始图像和真实值标注。

尽管有利的背景条件下，两种模型都表现出实例级别的错误。基于图像的模型正确推断出凝结尾 1 是分段的，但仅检测到凝结尾 0 的一个部分，完全遗漏了另一个部分。它完全没有检测到凝结尾 4，并错误地将凝结尾 5



(a) 基于图像的模型预测。 (b) 基于视频的模型预测。

Figure 6: 使用带有图像和视频输入的 Swin-L 模型对图 5 中显示的帧进行预测的实例。

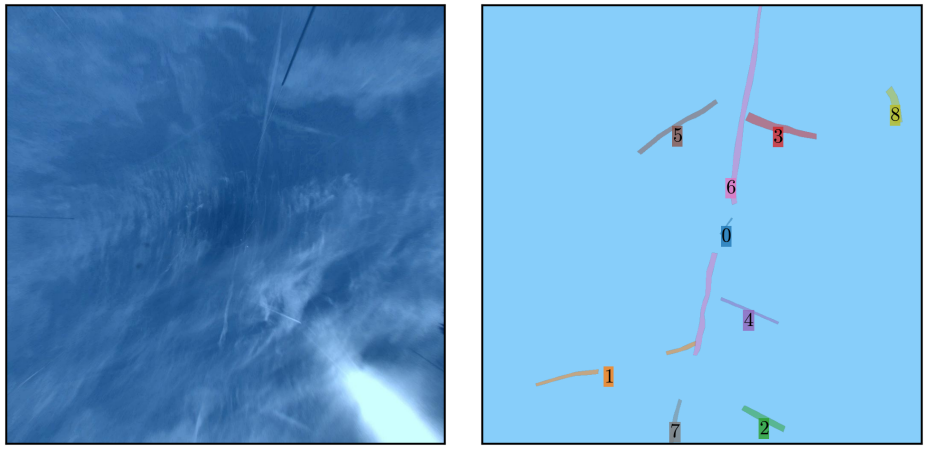
和 6 合并为一个预测。基于视频的模型也犯了类似的错误：它同样将凝结尾 5 和 6 合并，并未能检测到凝结尾 4。此外，它预测了凝结尾 0 的第二个片段，但分配给了一个不同的实例，并且错误地将凝结尾 1 分成两个独立的实例。

从语义分割的角度来看，两种模型的表现都相对较好，符合高对比度场景的预期。基于图像的模型获得了 Dice 分数 0.76 和平均 IoU 为 0.64，而基于视频的模型略胜一筹，Dice 为 0.79 且平均 IoU 为 0.67。然而，由于实例分组错误，图像模型在 AP@[0.25:0.75 | all | 100] 上稍微高于视频模型（图像模型为 0.62，视频模型为 0.55）。

图 7 显示了一个在 2023 年 11 月 19 日 08:49:30 捕捉到的更具挑战性的画面。这里，背景中有几个卷云，这些云结构中的一些类似于航迹云，从而引入了歧义。此场景还包括多个空间排列且碎片化的航迹云，增加了实例分割任务的复杂性。

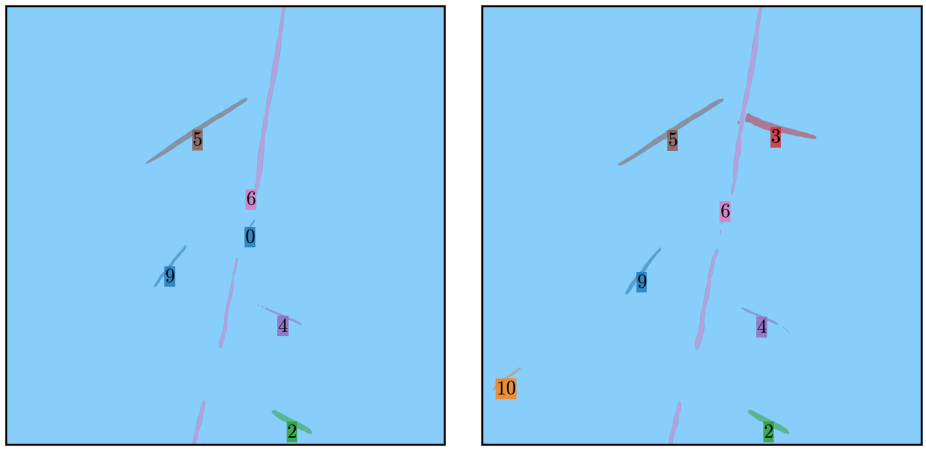
该场景展示了一个常见的失败模式：视觉上对齐但语义上不同的飞机尾迹的碎片化和误分组。尾迹 6 被分成了两个段，中间夹着尾迹 0；虽然它们看起来是共线的，但尾迹 0 是由另一次飞行产生的不同实例。尾迹 7 紧接着出现，并可能在缺乏飞行元数据的情况下与尾迹 6 和 0 错误关联。基于图像的模型正确地将尾迹 0 和 6 分开，但错误地将尾迹 6 和 7 合并。视频模型则将所有三个，6、0 和 7，分为一个单一的预测。有趣的是，这种错误反映了在没有飞行背景下人类可能的解释，凸显了任务的挑战性。

这两个模型都未能检测到卷云条纹 1 和 8，因为它们部分被云遮挡。它



(a) 原始图像。 (b) 真实标注

Figure 7: 2023 年 11 月 19 日 08:49:30 的原始图像和真实值标注。



(a) 基于图像的模型预测。 (b) 基于视频的模型预测。

Figure 8: 使用 Swin-L 模型和图像及视频输入预测图 7 所示帧的实例。

们还产生了一个误报（标记为卷云条纹 9），分割了一个类似卷云条纹的卷层云结构。尽管数据集质量很高，并通过飞行信息进行仔细标注，一些视觉上模糊的情况，如讨论中的例子，仍然很难绝对确定地进行标注。在这个例子中，预测的区域在结构和强度上都类似于卷云条纹，因此不清楚这个误报是源于模型错误还是地面实况标注中可以理解的遗漏。这些罕见的边缘情况强调了在视觉复杂场景中轻微标签噪声的潜在影响。未来的工作

可以从补充策略中获益，例如通过自信学习 (Northcutt et al., 2021) 来进一步完善标注并提高在边界情况下的鲁棒性。

该场景中的语义分割性能比前一场景低，反映了难度的增加。图像模型取得了 Dice 得分 0.61 和 mIoU 得分 0.43，而视频模型分别取得了 0.70 和 0.54。实例级 AP@[0.25:0.75 | all | 100] 分别为 0.35 和 0.37，与平均指标相似，使其成为一个具有代表性的案例。

这些例子展示了多边形航迹云分割中的几个关键挑战：（1）正确分组来自同一航班的碎片航迹云段；（2）由于类似航迹云的云层导致的视觉模糊性；（3）遮挡；以及（4）来自不同航班的航迹云的空间重叠。虽然基于视频的模型受益于时间信息，但它们可能会过度分组不同的实例。基于图像的模型避免了这一点，但往往无法连接碎片段。总体而言，这些例子展示了任务的内在难度和当前模型的局限性。

7. 结论

这项工作引入了一个新的数据集 (Jarry et al., 2025) 和用于从地面相机图像进行航迹云分割的基线模型。我们的实验表明，现代计算机视觉方法，尤其是像 Mask2Former 这样的全景分割模型，可以有效地应用于该任务，特别是在使用大型预训练模型和时间信息时。然而，性能的提升往往伴随着计算和内存需求的增加，凸显了准确性与实用性之间的权衡。

本研究的主要贡献是发布了第一个专门为实例级凝结尾迹分割、跟踪和飞行归因而设计的视频注释数据集。在详细的评估指标的支持下，包括跨多个交集-并集阈值和对象尺寸区间的平均精确度和召回率，该基准为这一新兴领域的进一步研究提供了可复现的基线。

我们当前装置的一个关键限制是可见光相机将观测限制在白天条件。然而，凝结尾迹在夜间通常具有最大的辐射影响，因为它们捕获了向外发射的长波辐射并促进了大气变暖。为了解决这个问题，我们部署了一个同地红外成像系统，使得可以进行昼夜连续监测。这也可能使我们能够开始在实际大气条件下估算单个凝结尾迹的辐射强迫。

同时，我们正在开发一种航迹归因算法，将观测到的航迹与特定飞机联系起来，该算法使用自动相关监视广播（ADS-B）轨迹数据。这一工具以及相关的数据和代码将在未来的出版物中公开发布。归因工作极为重要，因为它使每一条航迹都能连接到详细的飞机和发动机参数，比如飞机类型、发动机型号、燃油消耗率、飞行高度和环境条件。这些输入对于使用像 CoCiP 这样的物理模型重现航迹、评估其预期特性（例如冰晶数量、光学深度、寿命）是必需的，并且最终可以利用实地观测验证或优化这些模型。

我们还在通过对卫星图像中的凝结尾迹进行新的数据集注释，使用实例级和基于序列的标签来扩展这项工作。这个数据集将使我们能够测试和评

估本文提出的完整多尺度跟踪流程：从高分辨率的地面检测开始，然后归因于航班，最后在卫星图像中连接到相同的凝结尾迹随着它们的发展。该方法提供了一个独特的机会来研究凝结尾迹的形成、扩散和消散随时间和规模的变化。我们还计划使用我们基于地面的数据集来评估物理模型的预测，如 CoCiP。观测到的凝结尾迹演变与模拟结果之间的直接比较将有助于评估模型的准确性，并有可能为改进凝结尾迹预测和气候建模提供信息。

理想情况下，航迹云的检测、追踪和归因应该由一个能够联合处理视频、飞行轨迹数据和气象场的单一深度学习架构来解决。可以为此目的调整像 Mask2Former 的变体这样的模型。将这些任务整合到一个架构中将使端到端学习成为可能，并利用输入的互补性质，天气情况和航空交通数据对于检测和追踪航迹云都是非常有信息量的。然而，这种整合并不简单。它需要仔细设计输入数据的表示，以处理时空和多模态输入，创建对齐和一致的注释以适用于所有任务，并开发平衡检测、分割、追踪和归因之间矛盾目标的损失函数。尽管存在这些挑战，我们仍鼓励研究社区探索这种统一的做法。

更广泛地说，我们希望这项工作能够鼓励在其他地区开发类似的地面航迹监测系统。一种协作的、开放科学的方法，分享数据集、模型和观测基础设施，对于构建航迹行为的地理多样性和时间连续性图景至关重要。我们将本文视为朝着航迹研究的数据驱动生态系统迈出的第一步：一个整合物理建模与观测数据、跨越空间和时间尺度，并支持长期努力以更好地理解和减少航空对气候影响的系统。

References

Appleman, H., 1953. The formation of exhaust condensation trails by jet aircraft. Bulletin of the American Meteorological Society 34, 14 – 20. doi:10.1175/1520-0477-34.1.14.

Borella, A., Boucher, O., Shine, K.P., Stettler, M., Tanaka, K., Teoh, R., Bellouin, N., 2024. The importance of an informed choice of co 2-equivalence metrics for contrail avoidance. Atmospheric Chemistry and Physics 24, 9401–9417.

Brabandere, B.D., Neven, D., Gool, L.V., 2017. Semantic instance segmentation with a discriminative loss function. CoRR abs/1708.02551. URL: http://arxiv.org/abs/1708.02551, arXiv:1708.02551.

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko,

S., 2020. End-to-end object detection with transformers, in: European Conference on Computer Vision (ECCV), pp. 213–229.

Chen, L.C., Papandreou, G., Schroff, F., Adam, H., 2017. Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587 .

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European conference on computer vision (ECCV), pp. 801–818.

Cheng, B., Choudhuri, A., Misra, I., Kirillov, A., Girdhar, R., Schwing, A.G., 2021a. Mask2former for video instance segmentation. arXiv preprint arXiv:2112.10764 Extends Mask2Former to video by processing 3D segmentation volumes without modifying architecture or loss.

Cheng, B., Misra, I., Schwing, A.G., Kirillov, A., Girdhar, R., 2022. Masked-attention mask transformer for universal image segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

Cheng, B., Schwing, A.G., Kirillov, A., 2021b. Per-pixel classification is not all you need for semantic segmentation, in: Advances in Neural Information Processing Systems (NeurIPS).

Chevallier, R., Shapiro, M., Engberg, Z., Soler, M., Delahaye, D., 2023. Linear contrails detection, tracking and matching with aircraft using geostationary satellite and air traffic data. Aerospace 10, 578.

Dai, Z., Liu, H., Le, Q.V., Tan, M., 2021. Coatnet: Marrying convolution and attention for all data sizes. Advances in neural information processing systems 34, 3965–3977.

Duda, D.P., Minnis, P., Khlopenkov, K., Chee, T.L., Boeke, R., 2013. Estimation of 2006 northern hemisphere contrail coverage using modis data. Geophysical Research Letters 40, 612–617.

Forkert, T., Strauss, B., Wendling, P., 1993. A new algorithm for the automated detection of jet contrails from noaa-avhrr satellite images, in: Proc.

of the 6th AVHRR Data Users' Meeting, EUMETSAT-Joint Research Centre of the Commission of the EC. pp. 513–519.

Fritz, T.M., Eastham, S.D., Speth, R.L., Barrett, S.R., 2020. The role of plume-scale processes in long-term impacts of aircraft emissions. Atmospheric Chemistry and Physics 20, 5697–5727.

Geraedts, S., Brand, E., Dean, T.R., Eastham, S., Elkin, C., Engberg, Z., Hager, U., Langmore, I., McCloskey, K., Ng, J.Y.H., et al., 2024. A scalable system to measure contrail formation on a per-flight basis. Environmental Research Communications 6, 015008.

Gierens, K., Matthes, S., Rohs, S., 2020. How well can persistent contrails be predicted? Aerospace 7. doi:10.3390/aerospace7120169.

Gourgue, N., Boucher, O., Barthès, L., 2025. A dataset of annotated ground-based images for the development of contrail detection algorithms. Data in Brief 59, 111364.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2961–2969.

Jarry, G., Very, P., Ballerini, F., Dalmau, R., 2025. Gvccs : Ground visible camera contrail sequences. URL: https://doi.org/10.5281/zenodo.15743988, doi:10.5281/zenodo.15743988.

Jarry, G., Very, P., Heffar, A., Torjman-Levavasseur, V., . Deep semantic contrails segmentation of goes-16 satellite images: A hyperparameter exploration .

Kuhn, H.W., 1955. The hungarian method for the assignment problem. Naval Research Logistics Quarterly 2, 83–97.

Kulik, L., 2019. Satellite-based detection of contrails using deep learning. Ph.D. thesis. Massachusetts Institute of Technology.

Lee, D.S., Fahey, D.W., Skowron, A., Allen, M.R., Burkhardt, U., Chen, Q., Doherty, S.J., Freeman, S., Forster, P.M., Fuglestvedt, J., et al., 2021. The contribution of global aviation to anthropogenic climate forcing for 2000 to 2018. Atmospheric Environment 244, 117834.

Lin, T., Maire, M., Belongie, S.J., Bourdev, L.D., Girshick, R.B., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft COCO: common objects in context. CoRR abs/1405.0312. arXiv:1405.0312.

Low, J., Teoh, R., Ponsonby, J., Gryspeerdt, E., Shapiro, M., Stettler, M.E., 2025. Ground-based contrail observations: comparisons with reanalysis weather data and contrail model simulations. Atmospheric Measurement Techniques 18, 37–56.

Mannstein, H., Meyer, R., Wendling, P., 1999. Operational detection of contrails from noaa-avhrr-data. International Journal of Remote Sensing 20, 1641–1660.

McCloskey, K.J.F., Geraedts, S.D., Jackman, B.H., Meijer, V.R., Brand, E.W., Fork, D.K., Platt, J.C., Elkin, C., Van Arsdale, C.H., 2021. A human-labeled landsat-8 contrails dataset, in: Proceedings of the ICML Workshop on Tackling Climate Change with Machine Learning.

Meijer, V.R., Eastham, S.D., Waitz, I.A., Barrett, S.R., 2024. Contrail altitude estimation using goes-16 abi data and deep learning. Atmospheric Measurement Techniques 17, 6145–6162.

Meijer, V.R., Kulik, L., Eastham, S.D., Allroggen, F., Speth, R.L., Karaman, S., Barrett, S.R., 2022. Contrail coverage over the united states before and during the covid-19 pandemic. Environmental Research Letters 17, 034039.

Ng, J.Y.H., McCloskey, K., Cui, J., Meijer, V.R., Brand, E., Sarna, A., Goyal, N., Van Arsdale, C., Geraedts, S., 2023. Opencontrails: Benchmarking contrail detection on goes-16 abi. arXiv preprint arXiv:2304.02122 .

Northcutt, C., Jiang, L., Chuang, I., 2021. Confident learning: Estimating uncertainty in dataset labels. J. Artif. Int. Res. 70, 1373–1411. doi:10.1613/jair.1.12125.

Ortiz, I., García-Heras, J., Jafarimoghaddam, A., Soler, M., 2025. Enhancing goes-16 contrail segmentation through ensemble neural network modeling and optical flow corrections. Authorea Preprints .

Pratt, W.K., 2007. Digital image processing: PIKS Scientific inside. volume 4. Wiley Online Library.

Ridnik, T., Baruch, E.B., Noy, A., Zelnik-Manor, L., 2021. Imagenet-21k pretraining for the masses. CoRR abs/2104.10972. arXiv:2104.10972.

Riggi-Carrolo, E., Dubot, T., Sarrat, C., Bedouet, J., 2023. Ai-driven identification of contrail sources: Integrating satellite observation and air traffic data. Journal of Open Aviation Science 1.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer. pp. 234–241.

Sarna, A., Meijer, V., Chevallier, R., Duncan, A., McConnaughay, K., Geraedts, S., McCloskey, K., 2025. Benchmarking and improving algorithms for attributing satellite-observed contrails to flights. EGUsphere 2025, 1–58.

Schumann, U., 2012. A contrail cirrus prediction model. Geoscientific Model Development 5, 543–580. URL: https://gmd.copernicus.org/articles/5/543/2012/, doi:10.5194/gmd-5-543-2012.

Schumann, U., Hempel, R., Flentje, H., Garhammer, M., Graf, K., Kox, S., Lösslein, H., Mayer, B., 2013. Contrail study with ground-based cameras. Atmospheric Measurement Techniques 6, 3597–3612.

Sun, J., Roosenbrand, E., 2025. Few-shot contrail segmentation in remote sensing imagery with loss function in hough space. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing .

Teoh, R., Engberg, Z., Schumann, U., Voigt, C., Shapiro, M., Rohs, S., Stettler, M., 2023. Global aviation contrail climate effects from 2019 to 2021. EGUsphere 2023, 1–32.

Teoh, R., Schumann, U., Stettler, M.E., 2020. Beyond contrail avoidance: Efficacy of flight altitude changes to minimise contrail climate forcing. Aerospace 7, 121.

Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A., Li, Y., 2022. Maxvit: Multi-axis vision transformer, in: European conference on computer vision, Springer. pp. 459–479.

Van Huffel, J., Ehrmanntraut, R., Croes, D., 2025. Contrail detection and classification using computer vision with ground-based cameras, in: 2025 Integrated Communications, Navigation and Surveillance Conference (ICNS), IEEE. pp. 1–6.

Vazquez-Navarro, M., Mannstein, H., Mayer, B., 2010. An automatic contrail tracking algorithm. Atmospheric Measurement Techniques 3, 1089–1101.

Appendix A. 一致性实例跟踪算法

由于内存限制，视频分割模型在固定长度为 N 帧的短时间片上操作，使用步长为 1 的滑动窗口。虽然每个片段内的实例分割在时间上是一致的（即，在片段内的帧中实例标识保持不变），但模型是独立处理每个片段的。因此，实例标识在不同的片段之间不一定是一致的。

为了在整个视频序列中强制实施全局一致的实例标识符，我们实施了一种确定性的后处理方法，该方法对齐重叠片段之间的实例预测。该方法使用掩模重叠相似性，特别是共享帧之间的 IoU 值，并使用匈牙利算法执行最优二分匹配。下面，我们提供该方法的严格描述。

对于给定的帧索引 $t \in \{N, N+1, \ldots, T\}$ ，我们定义：

- 当前的片段作为序列 $F_{t-N+1}, F_{t-N+2}, \ldots, F_t$ 。
- 前一个片段作为序列 $F_{t-N}, F_{t-N+1}, \ldots, F_{t-1}$ 。

这两个片段有 $N-1$ 帧的重叠：$F_{t-N+1}, \ldots, F_{t-1}$ 。只有帧 F_t 是在当前片段中新引入的。在每一步，我们通过匹配重叠帧中的实例来传播一致的实例标识符。设：

- $\mathcal{I}_{\text{prev}} = \{1, \ldots, K\}$ ： 前一个片段中的实例标识符。
- $\mathcal{I}_{\text{curr}} = \{1, \ldots, M\}$ ：当前剪辑中的实例标识符。

我们定义一个代价矩阵 $C \in \mathbb{R}^{M \times K}$ ，其中每个元素 C_{ij} 表示实例 $i \in \mathcal{I}_{\text{curr}}$ 和实例 $j \in \mathcal{I}_{\text{prev}}$ 在重叠帧上的负时间 IoU：

$$C_{ij} = -\frac{1}{N-1} \sum_{f=t-N+1}^{t-1} \text{IoU}\left(\mathcal{M}_{i,f}^{\text{curr}}, \mathcal{M}_{j,f}^{\text{prev}}\right),$$

，其中 $\mathcal{M}_{i,f}^{\text{curr}}$ 和 $\mathcal{M}_{j,f}^{\text{prev}}$ 分别表示在帧 f 上实例 i 和 j 的二进制掩码。如果某个实例在给定帧中不存在（例如，掩码丢失），则其贡献被视为零重叠。

为了消除不太可能或噪声较大的匹配，我们在平均 IoU 上应用一个阈值 $\tau \in [0, 1]$ ：

$$C_{ij} = \begin{cases} C_{ij} & \text{if } -C_{ij} \geq \tau, \\ +\infty & \text{otherwise.} \end{cases}$$

，其中阈值 τ 是通过经验选择来平衡精度和鲁棒性；我们推荐 $\tau = 0.1$ 。

我们移除成本矩阵中仅包含 $+\infty$ 项的行和列。使用修改后的成本矩阵，我们通过匈牙利算法 (Kuhn, 1955) 解决二分匹配问题，从而获得当前实例和先前实例之间的一对一（或部分）映射。用 $\sigma : \mathcal{I}_{\text{curr}} \to \mathcal{I}_{\text{prev}} \cup \{\varnothing\}$ 表示所得的分配。然后，我们更新当前片段中的实例标识符，以匹配先前片段中分配的实例。未匹配的实例被分配新的唯一标识符。算法的伪代码在算法 1 中展示。

Algorithm 1 一致实例跟踪的后处理

Require: Predicted instance masks for video frames $F_1, \ldots, F_T$, threshold τ

Initialize unique identifier counter
Previous clip instances $\leftarrow$ Predicted instances on clip $(F_1, \ldots, F_N)$
Assign unique identifiers to all instances in previous clip instances
for $t = N + 1$ to T **do**
    Current clip instances $\leftarrow$ Predicted instances on clip $(F_{t-N+1}, \ldots, F_t)$
    Compute cost matrix C over frames $F_{t-N+1}, \ldots, F_{t-1}$
    Apply threshold τ and prune rows/columns with all $+\infty$
    $\sigma \leftarrow$ Hungarian Algorithm(C)
    Update instance identifiers in current clip using mapping σ
    Assign new identifiers to unmatched instances
    Previous clip instances $\leftarrow$ Current clip instances
end for

这个过程从帧 $t = N$ 到 T 按顺序应用，确保实例标识符在整个视频中全局一致。