

Graphical Abstract

GVCCS: A Dataset for Contrail Identification and Tracking on Visible Whole Sky Camera Sequences

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

Highlights

GVCCS: A Dataset for Contrail Identification and Tracking on Visible Whole Sky Camera Sequences

Gabriel Jarry, Ramon Dalmau, Philippe Very, Franck Ballerini, Stefania-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: A Dataset for Contrail Identification and Tracking on Visible Whole Sky Camera Sequences

Gabriel Jarry^a, Ramon Dalmau^a, Philippe Very^a, Franck Ballerini^a,
Stefania-Denisa Bocu^a

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

Abstract

Aviation’s climate impact includes not only CO₂ emissions but also significant non-CO₂ effects, especially from contrails. These ice clouds can alter Earth’s radiative balance, potentially rivaling the warming effect of aviation CO₂. Physics-based models provide useful estimates of contrail formation and climate impact, but their accuracy depends heavily on the quality of atmospheric input data and on assumptions used to represent complex processes like ice particle formation and humidity-driven persistence. Observational data from remote sensors, such as satellites and ground cameras, could be used to validate and calibrate these models. However, existing datasets ~~don’t explore all aspect~~ do not explore all aspects of contrail dynamics and formation: they typically lack temporal tracking, and do not attribute contrails to their source flights. To address these limitations, we present the Ground Visible Camera Contrail Sequences (GVCCS), a new open data set of contrails recorded with a ground-based all-sky camera in the visible range. Each contrail is individually labeled and tracked over time, allowing a detailed analysis of its lifecycle. The dataset contains 122 video sequences (24,228 frames) and includes flight identifiers for contrails that form above the camera. As reference, we also propose a unified deep learning framework for contrail analysis using a panoptic segmentation model that performs semantic segmentation (contrail pixel identification), instance segmentation (individual contrail separation), and temporal tracking in a single architecture. By providing high-quality, temporally resolved annotations and a benchmark for model evaluation, our work supports improved contrail monitoring and will facilitate better calibration of physical models. This sets the groundwork for more accurate climate impact understanding and assessments.

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

1. Introduction

Aviation contributes to global climate change not only through carbon dioxide (CO₂) emissions but also through a variety of non-CO₂ effects, including nitrogen oxides (NO_x), water vapour, and aerosols. Among these, condensation trails (contrails), ice-crystal clouds formed by aircraft at typical cruising altitudes, stand out for their potentially large ~~yet uncertain~~ radiative impact. Though they often appear as ephemeral white streaks in the sky, persistent contrails can spread into extensive cirrus-like cloud formations that ~~trap~~ outgoing long-wave radiation, warming the planet. Recent studies suggest that the climate forcing ~~due to contrail cirrus clouds from contrail cirrus~~ is of the same order of magnitude as ~~that from~~ aviation CO₂ emissions (Lee et al., 2021; Teoh et al., 2023), although this ~~comparison~~ depends on the metric used (Borella et al., 2024).

~~Yet, accurately~~ Assessing the climate impact of contrails remains a significant challenge for both aviation and climate scientists. ~~The lifecycle of the contrails depends~~ Contrail lifecycles depend on complex interrelated processes, ~~such as including~~ ice nucleation, crystal growth, wind-driven dispersion, and interaction with natural clouds, that are sensitive to ambient atmospheric conditions. Small variations in temperature and humidity, particularly relative humidity with respect to ice, can determine whether a contrail dissipates quickly or persists and spreads. This sensitivity, combined with ~~the~~ diurnal variability in radiative forcing (~~cooling when reflecting sunlight during the day; warming when trapping infrared radiation at night~~ daytime cooling from reflected sunlight versus nighttime warming from trapped infrared radiation), makes the net ~~climate~~ effect of contrails ~~both context-dependent and extremely difficult to model reliably~~ highly variable and challenging to model.

While contrail impacts have traditionally been studied using physical models, recent advances in remote sensing and computer vision now offer a valuable observational perspective. Physics-based models, such as the Contrail Cirrus Prediction model (CoCiP) (Schumann, 2012) ~~or APCEMM (Fritz et al., 2020) simulate the lifecycle of a contrail and the Aircraft Plume Chemistry, Emissions, and Microphysics Model (APCEMM) (Fritz et al., 2020), simulate contrail lifecycles~~ by solving complex equations that describe ~~the interaction~~ interactions

between aircraft emissions and atmospheric conditions. These models provide valuable theoretical insights, but their accuracy ~~is heavily dependent~~ depends heavily on the quality of ~~the~~ input data (Gierens et al., 2020). Key parameters, such as atmospheric temperature, humidity, and aircraft engine characteristics, are often uncertain, and these uncertainties propagate through ~~the~~ calculations, affecting ~~the reliability of the results~~ result reliability. Moreover, detailed simulations of contrail microphysics and radiative effects can be computationally demanding, particularly when applied to global-scale ~~analyzes~~ analyses.

Observational methods ~~;~~ using satellite and ground-based imagery ~~;~~ offer a direct ~~and~~ , data-driven way to study contrails, complementing approach to studying contrails that complements theoretical models. Satellite-based contrail detection has a long history, beginning with early automated methods that leveraged brightness temperature differences and Hough transforms in NOAA-AVHRR imagery (Mannstein et al., 1999). Subsequent work extended these techniques to study regional radiative forcing (Meyer et al., 2002), contrail coverage and properties (Minnis et al., 2005; Palikonda et al., 2005; Mannstein and Schumann, 2006), and global contrail distributions (Meyer et al., 2007). Advances in ~~high-resolution remote sensing and computer vision have made these methods increasingly effective (Meijer et al., 2022; McCloskey et al., 2021; Ng et al., 2023; Chevallier et al., 2023).~~ Beyond detection, observational data should play an increasing role in the future in refining physics-based models by providing empirical validation and calibrating the uncertain parameters mentioned above.

Integrating observational data with air traffic information like Automatic Dependent Surveillance-Broadcast (ADS-B) and meteorological data holds significant promise for advancing our understanding of the contrail lifecycle and climate impact. Linking contrails to specific flights, for which detailed parameters (e.g., engine type, altitude, and atmospheric conditions) are known, will allow for a better understanding of the role of these parameters into contrail formation and dynamics. However, achieving this integration requires addressing foundational challenges: accurately identifying contrails in images and sensor technology, particularly with MSG/or videos, distinguishing them from natural clouds (semantic segmentation), detecting individual instances (instance segmentation), and tracking their evolution over time. This paper focuses on these critical first steps, developing robust methods for contrail segmentation and tracking in both individual ground camera images and videos. While attribution remains a very challenging task to perform at scale using in particular geostationary satellites (Chevallier et al., 2023; Riggi-Carolo et al., 2023; Gerae

our work provide the necessary tools to reliably detect and track contrails locally, laying the groundwork for subsequent integration with flight and meteorological data SEVIRI, enabled rapid-scan observations that facilitated automated contrail tracking (Vázquez-Navarro et al., 2010), lifecycle analysis (Vázquez-Navarro et al., 2012), and improved detection algorithms (Ewald et al., 2013; Mannstein et al., 2012). Ground-based validation campaigns (Mannstein et al., 2010; Schumann et al., 2013) provided essential verification of satellite-derived contrail properties. More recently, high-resolution remote sensing combined with modern computer vision and deep learning has further enhanced detection capabilities (Meijer et al., 2022; McCloskey

Despite growing interest in observational contrail analysis, publicly available datasets ~~are still remain~~ limited in scope. ~~The most universally used resource, Google~~ Existing datasets of contrails annotated in observational data, such as Google’s OpenContrails, ~~do not track individual contrails over time or provide information on the flights that formed them.~~ Specifically, OpenContrails offers instance-level masks only on the central GOES-16 GOES-16 frame, with surrounding images left unannotated, ~~hindering contrail tracking across time.~~ In contrast, ~~(Sarna et al., 2025)~~ Sarna et al. (2025) introduced SynthOpenContrails, which overlays synthetic contrails and annotations onto real scenes, providing full per-frame localization, tracking, and flight attribution, ~~demonstrating.~~ This demonstrates that richly annotated data ~~exists~~ can exist, even if confined to synthetic contrail overlays rather than human annotation. An ideal scenario would be a fully annotated video dataset where every frame is ~~humanly labelled~~ labeled and each contrail is assigned a persistent identifier across time.

To advance research in this area, ~~this paper we~~ present the Ground Visible Camera Contrail Sequences (GVCCS), an open dataset (Jarry et al., 2025) with instance-level annotations, derived from ground-based video recordings in Brétigny-sur-Orge, France (Réuniwatt CamVision visible ground-based camera). Our dataset includes 122 videos (of duration between 20 minutes ~~to and~~ 5 hours) with a total ~~frame number of around of approximately~~ 24,200 frames, each annotated with instance-level labels. By making this dataset openly available, ~~this paper provides we provide~~ a valuable benchmark for both the atmospheric and aviation research communities.

To support future performance comparisons, we introduce ~~here~~ a deep learning-based model for contrail segmentation and tracking. Instead of relying on separate models for these tasks, an approach that often requires complex, ad-hoc combinations of techniques, we adopt a unified framework based

on Mask2Former (Cheng et al., 2021b), a state-of-the-art computer vision model. Mask2Former is designed for panoptic segmentation, which combines semantic segmentation (labeling each pixel with a class, e.g., ~~“contrail”~~ or ~~“contrail”~~ or “sky”) and instance segmentation (distinguishing between individual objects, e.g., different contrails). In addition to separating contrails from clear sky, it ~~could~~ can handle complex backgrounds, such as ~~low-altitude~~ ~~low-altitude~~ cloud layers that partially or fully obscure contrails, by assigning appropriate ~~“cloud”~~ “cloud” labels while still maintaining unique instance identities. For example, in a single image, panoptic segmentation can identify all visible contrail pixels, correctly label intervening clouds, and assign consistent instance masks to each contrail, even when they overlap, intersect, appear fragmented, or are seen through thin cloud cover. In fact, contrails often break into multiple disconnected components due to atmospheric conditions and natural dissipation processes. A robust monitoring system must not only identify these fragments, but also associate them with the correct contrail instance.

It is worth noting that, fragmentation poses a significant challenge for contrail analysis based solely on images or videos: visually disjointed segments from the same flight must be grouped without external data. Moreover, ~~low-altitude~~ ~~low-altitude~~ cloud obscuration and sun glare can further interrupt or mask contrail continuity, producing ~~multi-polygon~~ ~~multi-polygon~~ annotations even for a single physical contrail. In operational settings, however, it is possible to first perform single-polygon instance segmentation and then associate multiple instances with the same flight using auxiliary data such as aircraft trajectories and wind fields. This post-processing step enables grouping across time and space based on flight identity rather than visual continuity. In this work, we restrict ourselves to purely image-based analysis and defer the integration of external data sources to future work.

Mask2Former, originally designed for individual images, can be easily extended to video data to improve the consistency of panoptic segmentation across frames (Cheng et al., 2021a). By leveraging temporal information, Mask2Former for videos performs semantic segmentation, instance segmentation, and tracking in an integrated manner. In this paper, we study both the frame-based and video-based versions of Mask2Former, comparing their performance on our dataset.

The remainder of this paper is structured as follows. Section 2 provides the necessary background on contrail formation and computer vision techniques, establishing the foundation for the challenges addressed in this work.

149 Section 3 reviews related work on contrail datasets and segmentation mod-
 150 els, highlighting current limitations and motivating our approach. Section 4
 151 introduces our newly developed video-based dataset, detailing its annotation
 152 methodology and unique instance-level structure. Section 5 describes our
 153 panoptic segmentation framework based on the Mask2Former architecture.
 154 Section 6 presents and analyses the experimental results. Finally, Section 7
 155 summarises our main contributions and outlines future research directions.

156 2. Background

157 This section introduces the key concepts necessary to understand the chal-
 158 lenges addressed in this work. We begin by outlining the physical processes
 159 behind contrail formation and ~~the~~their implications for climate, focusing on
 160 why contrails are particularly difficult to detect and track. We then review
 161 relevant computer vision techniques, specifically object detection and image
 162 segmentation, and assess their suitability for analysing contrails.

163 2.1. The Science of Contrails

164 Contrails are artificial clouds that form behind aircraft when hot, humid
 165 engine exhaust mixes with the cold, low-pressure air at cruising altitudes, typ-
 166 ically between 8 and 12 km. If ~~the~~their atmospheric conditions are ~~right,~~suitable
 167 — specifically, if the temperature falls below -40°C and a critical threshold
 168 (typically around -40°C , depending on pressure and humidity) and the air is
 169 sufficiently humid, ~~the~~the water vapour in the exhaust condenses and freezes
 170 into ice crystals. ~~This process, modelled and quantified by~~ The physical
 171 mechanism underlying this process was first explained by Schmidt (1941),
 172 who recognized that contrails form when ambient temperature is low enough
 173 to cause the humidity inside the aircraft plume to reach saturation with
 174 respect to liquid water, triggering condensation. Appleman (1953) provided
 175 further quantitative analysis, though without fully accounting for engine
 176 characteristics. Schumann (1996) later developed a comprehensive treatment
 177 incorporating engine efficiency and practical application methods, formalizing
 178 what is now known as the Schmidt–Appleman criterion ~~(Appleman, 1953),~~
 179 . This process produces the familiar thin, white trails visible in the sky.
 180 ~~Some contrails dissipate rapidly, while others persist and spread, eventually~~
 181 ~~forming larger ice cloud structures known as contrail cirrus.~~

182 Like natural clouds, contrails influence ~~the~~ Earth’s radiation budget: they
 183 ~~trap~~reduce outgoing long-wave radiation, leading to warming, while also

reflecting incoming solar radiation, which has a cooling effect. The net result depends on the contrail's altitude, optical properties, lifespan, and the time of day. The ~~precise relative impact~~ magnitude of contrail climate forcing relative to aviation's CO₂ emissions depends on the climate metric chosen (Borella et al., 2024); however, contrails are thought to warm the climate at a level of the same order of magnitude as aviation's CO₂ emissions (Lee et al., 2021; Teoh et al., 2023). This makes the monitoring and characterization of contrails an essential part of understanding aviation's full environmental footprint (Teoh et al., 2023) and developing mitigation strategies (Teoh et al., 2020).

Quantifying this radiative forcing requires understanding both contrail optical properties and their spatial and temporal distribution. Early satellite-based studies provided first estimates of regional contrail radiative effects (Meyer et al., 2002) and developed parametric models linking contrail properties to radiative forcing (Schumann et al., 2009). Climatological analyses of persistent contrails revealed dependencies on atmospheric conditions and aircraft traffic patterns (Iwabuchi et al., 2012; , while ground-based observations offered validation of satellite-derived contrail properties (Mannstein et al., 2010).

As mentioned above, the observational viewpoint offers an alternative perspective that focuses on detecting and analysing contrails directly ~~in~~ atmospheric imagery using satellite and ground-based remote sensing instruments. However, detecting and tracking contrails presents several technical challenges, which helps explain the growing research interest in the topic. Satellite imagery often lacks the spatial and temporal resolution needed to detect contrails in their early stages (~~Ng et al., 2023~~) (Ng et al., 2023; Mannstein et al., 2010) . Geostationary satellites have a nominal spatial resolution of about 0.5 to 2 km and a temporal resolution of 5 to 15 min, which is often insufficient to capture the narrow, faint, and short-lived nature of freshly formed contrails unless they persist and grow. Even when contrails ~~do~~ spread into detectable cloud structures, they are difficult to distinguish from natural cirrus, particularly in scenes with complex cloud layers. Moreover, by the time a contrail is visible in satellite images, it has often drifted and deformed, complicating ~~the~~ attribution to the flight that produced it (Chevallier et al., 2023; Sarna et al., 2025). This linkage is crucial, as identifying the originating flight enables researchers to retrieve essential details such as aircraft type and engine model, key inputs for assessing contrails' environmental impact and improving physical models through comparison with empirical observations.

Ground-based cameras (Schumann et al., 2013; Low et al., 2025) offer

222 a complementary perspective with critical advantages. Positioned beneath
223 flight paths, these systems can capture high-resolution images and video with
224 far greater spatial and temporal fidelity than satellites. Crucially, they can
225 detect contrails immediately after formation, while they are still thin, linear,
226 and visually distinct. This early visibility simplifies the task of associating
227 observed contrails with the specific flight responsible, especially when com-
228 bined with precise trajectory data. The main drawback is, naturally, their
229 restricted spatial coverage, which hinders the ability to monitor contrails
230 from ~~their~~ formation to dissipation.

231 This attribution advantage is particularly significant compared to satellite-based
232 approaches. Geostationary satellites face several challenges: their coarse
233 spatial resolution ($\sim 0.5\text{--}2$ km/pixel) means contrails must persist and spread
234 before becoming detectable, by which time wind advection has displaced
235 them substantially from their formation location; their temporal resolution
236 (5–15 minutes) means the originating aircraft may be far away when the
237 contrail first appears; and multiple aircraft may have traversed similar airspace
238 during this window, creating ambiguity. Attribution from satellite data
239 therefore requires sophisticated algorithms accounting for wind fields, parallax,
240 and probabilistic matching (Chevallier et al., 2023; Riggi-Carolo et al., 2023; Geraedts et al., 2024
241 . In contrast, ground-based cameras observe contrails at formation with
242 high spatial resolution (~ 73 m/pixel at 10 km altitude in our system) and
243 30-second sampling, enabling straightforward contrail-to-flight attribution
244 without the ambiguities inherent in satellite-based approaches.

245 While not the focus of this paper, one promising direction involves com-
246 bining ground-based and satellite observations into a unified monitoring
247 framework. In such a system, contrails would first be detected in high-
248 resolution ground-based imagery and attributed to specific flights using tra-
249 jectory and weather data, providing access to key aircraft and engine param-
250 eters. Crucially, to enable continuous tracking beyond the limited field of
251 view of the ground-based camera, these contrails would then need to be re-
252 liably linked to their evolving counterparts in satellite imagery as they drift,
253 expand, and age. Successfully associating contrails across these two modali-
254 ties, ~~ground~~ and ~~satellite~~ would allow monitoring of their full lifecycle
255 from formation to dissipation while preserving information about the specific
256 aircraft and flight responsible for creating them.

2.2. Computer Vision Techniques for Contrail Monitoring

Contrails are visually challenging targets for computer vision due to their thin, elongated shapes, variable curvature, and tendency to fragment or fade over time. These characteristics make them fundamentally different from the objects typically addressed in standard object detection benchmarks, such as vehicles and animals in datasets like the Common Objects in Context (COCO) dataset (Lin et al., 2014), which features well-defined, discrete objects.

Traditionally, object detection methods localise targets using bounding boxes, usually axis-aligned rectangles. [Standard approaches such as Faster R-CNN \(Ren et al., 2017\) and YOLO \(Redmon et al., 2016\) exemplify this paradigm.](#) This approach works well for objects like cars or animals, which are compact and roughly rectangular, but performs poorly for contrails. A single axis-aligned bounding box may inadvertently include multiple contrail segments or large amounts of background sky, while missing parts of curved or fragmented trails. Oriented bounding boxes offer some improvement by allowing rotation, which better fits the geometry of elongated contrails. However, they still fall short in capturing fine-grained shapes, gaps, or fading segments. Figure 1 shows the limitations of axis-aligned and oriented bounding boxes for object detection on contrails.

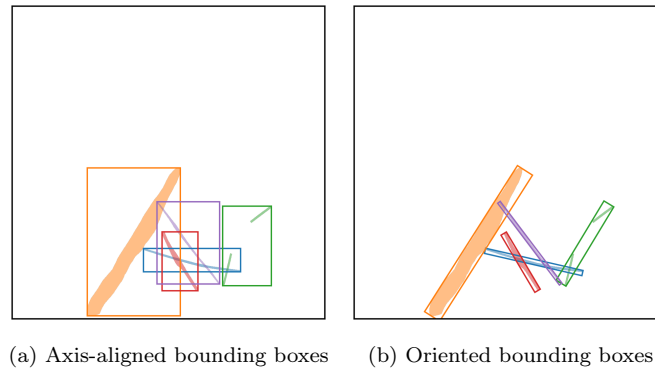


Figure 1: Illustration of bounding box detection on contrails. Each detected contrail is highlighted with a distinct color. Note how elongated or fragmented trails challenge bounding box alignment and separation.

Instance segmentation provides a more precise solution by predicting pixel-level masks for each individual object. This approach is particularly beneficial for contrails, as it can delineate each trail accurately even when

they intersect, overlap, or dissipate unevenly. For instance, two overlapping contrails that fade at different rates can still be assigned to distinct instances.

It is important to note that instance segmentation has been addressed in atmospheric science for decades using classical computer vision techniques. Early work by Mannstein et al. (1999) detected contrail pixels and grouped spatially connected regions into distinct objects. Similarly, Schumann et al. (2013) used ground-based cameras with automated algorithms to identify, track, and characterize individual contrails. These methods achieved instance-level contrail separation through feature-based detection, connectivity analysis, and trajectory matching. Our work builds on this foundation by applying modern deep learning architectures that perform instance segmentation through learned feature representations rather than hand-crafted rules.

Semantic segmentation, in contrast, labels each pixel by class (e.g., "contrail" or "sky") but does not distinguish between individual contrails. This is insufficient when studying temporal evolution or interactions between specific contrails, since it treats all contrails as a single undifferentiated class.

Panoptic segmentation combines the strengths of both approaches: it assigns a class label to every pixel (semantic segmentation) and an instance identifier where appropriate (instance segmentation). In this framework, "things" such as individual contrails are assigned unique instance labels, while "stuff" like the background sky or natural clouds is labelled only by class. This unified view is well-suited to contrail monitoring, enabling fine-grained analysis of individual contrails within the broader atmospheric context. Moreover, the framework can be readily extended to additional classes (e.g., cirrus, cumulus) for more comprehensive scene understanding, provided ~~of course, that~~ these classes have been effectively and consistently labelled during dataset creation, which introduces an additional layer of complexity to the annotation campaign. Figure 2 illustrates the instance, semantic, and panoptic segmentation methods.

An important but often overlooked issue in the literature is how contrails are geometrically represented. In reality, a single contrail may consist of several disconnected segments, for example, due to fading or occlusion making it a multi-polygon shape. See, for instance, the green contrail in Figs. 2a and 2c. However, the most natural approach is to simplify this by treating each segment as a separate, independent polygon, effectively assuming that each fragment belongs to a different contrail.

While this simplification avoids the complexity of handling multi-polygons directly, it introduces a significant challenge: to reconstruct the full contrail,

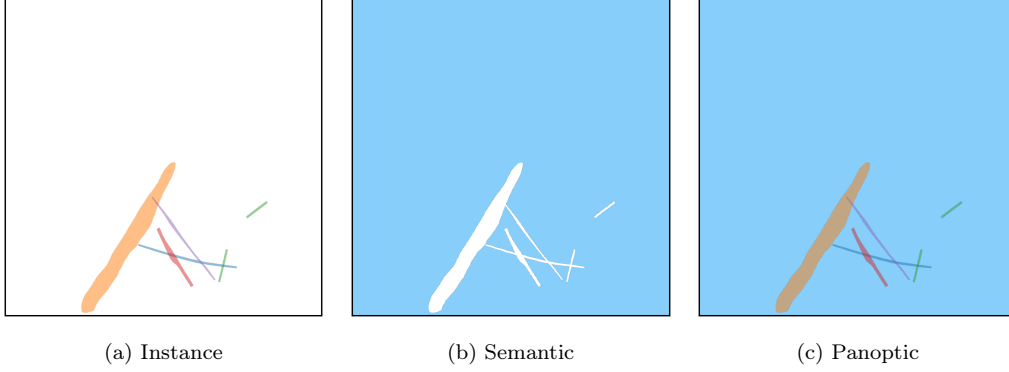


Figure 2: Comparison of segmentation methods applied to illustrative contrails. Distinct colours indicate different (a) Instance segmentation assigns unique identifiers (colors) to each contrail instances or classes, depending on the method enabling individual tracking but without classifying non-contrail regions. (b) Semantic segmentation identifies all contrail pixels as a single class (white) versus background (blue), without distinguishing between individual contrails. (c) Panoptic segmentation combines both approaches: each contrail receives a unique identifier while all pixels are classified (contrails in color, background in blue). This unified representation enables both instance-level tracking and scene-level understanding.

one must find a way to link fragmented pieces together. This requires ad-hoc linking strategies, which vary in complexity and accuracy. Some methods rely purely on the geometric properties of the fragments, such as their proximity or alignment, while others incorporate external data, such as aircraft flight paths or meteorological information, to make more informed associations.

In this work, we adopt panoptic segmentation as the foundation for segmenting and tracking contrails. This choice is motivated by its ability to simultaneously achieve instance-level precision and maintain contextual awareness of the surrounding scene. Moreover, by explicitly addressing the issue of fragmented contrails, our method enables instance-level identification of contrails without requiring external sources of information, such as flight or weather data. This is particularly valuable in scenarios where such data may be unavailable or incomplete. However, we also explore an alternative version of the model that treats each contrail fragment as an independent instance, under the assumption that a downstream algorithm, leveraging external traffic and meteorological data, will later associate these segments with their corresponding flights. The comparative evaluation of these two strategies ; self-contained instance identification versus externally supported

336 post-association ~~;~~  will be presented in future publications. In this paper,
337 we focus solely on presenting the contrail segmentation models themselves.

338 3. State of the Art

339 This section presents an overview of prior work in contrail segmenta-
340 tion and analysis, focusing first on the datasets that have been developed
341 to support this research, and then on the computational models used for
342 contrail segmentation and flight attribution. The scope and key features of
343 existing datasets are outlined, with particular attention given to the lim-
344 ited availability of temporal annotations and flight attribution ground truth.
345 Subsequently, we examine state-of-the-art segmentation and tracking meth-
346 ods, particularly deep learning-based approaches, assessing their applicability
347 and performance in contrail analysis. This review highlights gaps in current
348 research and motivates the contributions presented in this paper.

349 3.1. Datasets

350 Recent advances in contrail detection have been supported by the devel-
351 opment of annotated datasets, primarily based on satellite imagery. These
352 datasets have facilitated the application of computer vision techniques for
353 contrail identification, although aspects such as temporal continuity and in-
354 tegration with flight metadata remain limited in most cases. In this section,
355 we review the most relevant publicly available datasets and place our contri-
356 butions within this context.

357 ~~(Kulik, 2019) and (Meijer et al., 2022) are~~ Kulik (2019) and Meijer et al. (2022)
358 are, to our knowledge, the first studies to leverage a modern, data-driven,
359 deep learning framework for large-scale contrail segmentation. The authors
360 developed and applied convolutional neural networks, which were trained
361 using a manually curated dataset comprising over 100 ~~manually~~-annotated
362 geostationary GOES satellite images with instance segmentation.

363 One of the first large-scale labelling efforts in contrail detection was led by
364 Google Research, beginning with the development of a contrail dataset based
365 on high-resolution Sentinel satellite imagery (McCloskey et al., 2021). Hu-
366 man experts manually annotated the images using structured guidelines, pro-
367 ducing ~~polygonal masks for each visible contrail segment~~ masks that identify
368 contrail pixels at the semantic segmentation level, distinguishing contrail
369 from non-contrail regions without tracking individual contrail instances. Mul-
370 tiple annotators independently labelled each image, and the dataset includes

all individual annotations, with the option to filter results by majority consensus. This methodology improved both the spatial precision and overall quality of the labels.

Building on this work, Google released the OpenContrails dataset (Ng et al., 2023), which is based on images from the GOES-16 Advanced Baseline Imager (ABI). ~~Thanks to the 10-minute temporal resolution provided by the geostationary orbit of GOES-16, the dataset is well suited to study contrails at large scales.~~ OpenContrails offers temporal context by including short sequences of unlabelled images surrounding each annotated frame, providing valuable information to annotators for more accurate labelling. Only the central frame in each sequence is annotated, therefore not allowing a direct comparison of contrail dynamics with physical models. ~~Notably, a 2025 update introduced instance-level labels, enabling the use of the dataset for instance-based models and expanding its potential for more advanced contrail analysis.~~

In the domain of ground-based data for contrail research, significant resources have been developed to support computer vision tasks. Gourgue et al. (2025) introduce an ~~open-aceess~~ open-access corpus of around 1,600 ~~polygon-annotated~~ polygon-annotated hemispheric sky images acquired at the SIRTa atmospheric laboratory, near Paris, offering class labels that distinguish ~~“young,” “old,” and “very old”~~ “young,” “old,” and “very old” contrails as well as several confounding artefacts. By capturing ~~high-resolution~~ high-resolution ground views minutes after formation, ~~the dataset fills the~~ this dataset fills a temporal-spatial gap left by satellite benchmarks. Complementary to this data provision, Pertino et al. (2024) focus on the development of detection methodology, providing a comprehensive comparison of computer vision models applied to both visible and infrared images.

Rather than creating a dataset for training modern convolutional networks on segmentation tasks, ~~(Low et al., 2025)~~ Low et al. (2025) manually annotated the correspondence between contrail waypoints ~~;~~ derived from the application of the CoCiP model and observations from their wide-angle ground camera system. This approach is particularly well-suited for directly assessing and parametrizing physical models.

~~Meijer et al. (2024) is~~ Earlier studies have successfully collocated contrails using various combinations of sensors, including ground-based observations, satellite imagery, and lidar data (Iwabuchi et al., 2012; Mannstein et al., 2010). For example, Vázquez-Navarro et al. (2010) demonstrated tracking contrails first identified in high-resolution MODIS imagery through time sequences of

409 [Meteosat data, leveraging complementary spatial and temporal resolution.](#)
410 [Building on this foundation, Meijer et al. \(2024\)](#) is, to our knowledge, the
411 first example of ~~dataset collocating images on two different~~ [a dataset specifically](#)
412 [designed for contrail altitude estimation by collocating images from two](#)
413 [distinct](#) remote sensors: they assembled a dataset ~~specifically for contrail~~
414 ~~altitude-altitude estimation, comprising over 3000~~ [comprising over 3,000](#) cases
415 over the contiguous United States (2018–2022). Contrails were first lo-
416 cated via automated detection in ~~GOES-16~~ [GOES-16](#) ABI infrared imagery,
417 then precisely collocated, correcting for parallax and wind advection, with
418 CALIOP lidar ~~cross-sections~~ [cross-sections](#). The team then conducted man-
419 ual inspections of the matched imagery to verify and validate alignment. This
420 benchmark dataset linking geostationary contrail signatures to ~~high-resolution~~
421 [high-resolution](#) vertical profiles enables supervised ~~deep-learning~~ [deep-learning](#)
422 approaches to predict contrail top heights from ABI data.

423 A significant advance in contrail detection has been the development of
424 synthetically ~~labeled datasets~~ [\(Chevallier et al., 2023\)](#) ~~labelled datasets~~ [Chevallier et al. \(2023\)](#)
425 generated a synthetic dataset using CoCiP (Schumann, 2012) to overlay con-
426 trail polygons onto GOES-16 imagery, enabling the first instance segmen-
427 tation pipeline for contrail detection. The performance of flight assignment
428 algorithms was validated using actual GOES data, through manual inspec-
429 tion rather than synthetic reference ground truth. Building on this synthetic
430 foundation, ~~(Sarna et al., 2025)~~ [Sarna et al. \(2025\)](#) introduced a benchmark
431 dataset, SynthOpenContrails, with sequences of synthetic contrail detections
432 tied to known flight metadata, providing the first opportunity to quantita-
433 tively evaluate and improve contrail-flight attribution algorithms. To our
434 knowledge, this is the only dataset providing localized and tracked contrails
435 with attributable ground truth, albeit synthetic. While the use of synthetic
436 datasets represents a modern and cutting-edge technique for training algo-
437 rithms, the use of manually labelled data as test sets is ~~still~~ theoretically
438 preferable to objectively assess algorithmic performance. However, obtaining
439 such datasets on geostationary satellite images, with their coarse resolution,
440 remains very difficult at this stage, which motivates the approach adopted by
441 the authors. As mentioned in ~~(Sarna et al., 2025)~~ [Sarna et al. \(2025\)](#), obtain-
442 ing such a reference dataset with ground truth for flight attribution based on
443 human annotations is ~~definitely~~ feasible in principle with higher resolution
444 ~~low-orbit~~ [low-orbit](#) satellites or ground-based cameras, which is the focus of
445 the present work.

446 Overall, while existing datasets have contributed valuable resources, there

447 is a lack of comprehensive, human-labelled data containing temporally re-
 448 solved, instance-level, and flight-attributed annotations. Our work addresses
 449 this issue by introducing a dataset designed to provide these annotations,
 450 collected using our ground camera system.

451 3.2. Models

452 Contrail monitoring with computer vision was first pioneered in the early
 453 nineties (Forkert et al., 1993; Mannstein et al., 1999), using ~~non-data-driven~~
 454 ~~traditional~~ image-analysis techniques. Their work applied ~~linear-kernel~~ ~~linear-kernel~~
 455 methods, direct thresholding of brightness temperature difference channels,
 456 and early Hough-transform operators (Pratt, 2007) optimized for linear shape
 457 detection, to identify contrails in AVHRR satellite imagery. This ~~approach~~
 458 ~~was further improved by (Vázquez-Navarro et al., 2010) and (Duda et al., 2013)~~
 459 ~~foundational work was extended through improved detection algorithms (Meyer et al., 2002, 2007)~~
 460 ~~, automated tracking methods (Vázquez-Navarro et al., 2010), and enhanced~~
 461 ~~cirrus detection capabilities (Ewald et al., 2013; Mannstein et al., 2012). Parallel~~
 462 ~~advances in cloud property retrieval from geostationary satellites (Bugliaro et al., 2012; Hamann et~~
 463 ~~al., 2017b,a) further~~
 464 ~~refined contrail and cirrus characterization. Ground-based validation studies (Mannstein et al., 2010)~~
 465 ~~provided essential verification of these satellite-based methods. These classical~~
 466 ~~computer vision approaches were later complemented by improvements from Duda et al. (2013)~~
 467 ~~and eventually by modern deep learning techniques.~~

468 To the best of our knowledge, ~~Kulik (2019); Meijer et al. (2022)~~ ~~Kulik (2019)~~
 469 ~~and Meijer et al. (2022)~~ represent the earliest applications of modern con-
 470 volutional networks to pixel-level classification and semantic segmentation.
 471 Building on the OpenContrails dataset, Ng et al. (2023) employed semantic
 472 segmentation algorithms, specifically DeepLabV3 (Chen et al., 2017, 2018),
 473 to identify contrails in ~~ash-rgb~~ ~~ash-RGB~~ composites using brightness tem-
 474 perature differences. Their work demonstrated that adding temporal context
 475 via a 3D encoder, incorporating the time dimension, led to improved perfor-
 476 mance. Moreover, results from the subsequent Kaggle competition showed
 477 that ~~UNet~~ ~~U-Net~~ models (Ronneberger et al., 2015) equipped with modern
 478 transformer backbones, such as MaxViT (Tu et al., 2022) and CoatNet (Dai
 479 et al., 2021), achieved even stronger results (Jarry et al., 2024).

480 Using an ensemble approach, Ortiz et al. (2025) combined six neural
 481 networks, including U-Net, DeepLab, and transformer architectures, and ap-
 482 plied optical-flow-based corrections to maintain temporal consistency across

consecutive satellite frames. Meanwhile, Sun and Roosenbrand (2025) introduced a Hough-space line-aware loss for few-shot scenarios, supplementing Dice loss with a global alignment term to encourage predictions to align with linear structures.

Shifting from pixel-level masks to instance-level contrail segmentation and making use of synthetic data (Chevallier et al., 2023), Chevallier et al. (2023) introduced the first algorithmic pipeline focused on instance segmentation for contrail detection, utilizing the Mask R-CNN algorithm (He et al., 2017). Similarly, Van Huffel et al. (2025) adopted Mask R-CNN to process images captured by their wide-angle ground camera system.

The difficult-challenging task of attributing detected contrails to individual flights (typically using ADS-B information) in geostationary satellite imagery, typically using automatic dependent surveillance-broadcast (ADS-B) data, has been the focus of several recent studies. (Chevallier et al., 2023) Chevallier et al. (2023) introduced a pipeline that combines contrail detection, tracking, and matching with aircraft using geometric criteria and wind-corrected trajectories. Riggi-Carolo et al. (2023) proposed an probabilistic matching methods that account Riggi-Carolo et al. (2023) proposed a probabilistic matching method that accounts for uncertainties in flight data and atmospheric conditions leveraging as well on Hough-based line detection. (Geraedts et al., 2024) , incorporating features derived from Hough-based line detection to improve alignment. Geraedts et al. (2024) presented a scalable system designed to assign contrails to flights on a large scale, enabling routine monitoring of contrail formation and supporting climate assessments. (Sarna et al., 2025) Sarna et al. (2025) systematically benchmarked and refined these attribution algorithms, highlighting common challenges and proposing improved association metrics, building on the release of the synthetically generated SynthOpenContrails dataset.

By contrast, our work targets ground-based imagery, capturing contrails immediately after formation and enabling near-instantaneous flight attribution via ADS-B ADS-B data. We harness panoptic segmentation using Mask2Former, trained on high-resolution video, to extract pixel-accurate masks of individual contrails and track them over time. This fills the gap in early-stage contrail detection and provides richer spatial and temporal detail than existing satellite-based models.

518 4. Dataset

519 The primary contribution of this paper is the introduction of a new
520 dataset designed to support contrail detection, tracking, and attribution.
521 This section provides a detailed overview of the dataset. Section 4.1 de-
522 scribes the data collection and labelling campaign. Section 4.2 summarizes
523 the structure and content of the dataset.

524 4.1. Data collection and labelling campaign

525 To support the development of machine learning models for contrail detec-
526 tion, we conducted an extensive labelling campaign as part of the *ContrailNet*
527 project. Visible-spectrum image sequences were acquired using ~~a~~an all-sky
528 ground-based camera installed on the roof of the EUROCONTROL Innova-
529 tion Hub ~~, capturing (Location: 48°36'1.87" N, 2°20'48.46" E). The camera~~
530 captured the sky every 30 seconds at a resolution of 1976×2032 pixels.

531 Our camera provider, Reuniwatt, ~~has~~ delivered a dual ~~all-sky~~all-sky cam-
532 era system: the first unit, CamVision, operates in the visible spectrum, cap-
533 turing ~~high-resolution~~high-resolution fisheye images every 30 seconds with
534 on-board processing and self-calibration, ensuring reliable daytime operation
535 even in dusty or wet conditions. The second unit, SkyInsight, uses long-
536 wave infrared (~~8–13 μ m~~8–13 μ m) imaging via a ~~chrome-coated~~chrome-coated
537 hemispherical mirror and will be used in future research.

538 The raw all-sky images were first geometrically projected onto a square
539 grid. This projection process ~~, uses~~ camera-specific calibration files to asso-
540 ciate each pixel with its corresponding azimuth and zenith angles, effectively
541 removing lens distortions and re-mapping the sky onto a uniform Cartesian
542 representation. A ~~75km \times 75km~~75 km \times 75 km grid of georeferenced points
543 was computed at a fixed cloud altitude (10 km), and a linear interpolation
544 scheme was used to assign raw pixel values to the projected frame. The out-
545 put is a square image of size 1024×1024 pixels that preserves the spatial
546 geometry of the sky above the camera.

547 To improve the visual clarity and consistency of the sequences, each
548 projected image ~~then~~ undergoes a three-step enhancement process. First,
549 brightness is increased using a linear scaling operation to compensate for
550 underexposure in certain atmospheric conditions. Second, local contrast is
551 enhanced via CLAHE (Contrast Limited Adaptive Histogram Equalization),
552 which boosts ~~features like~~fine features like faint or fragmented contrails with-
553 out overexposing ~~the image~~bright regions. Finally, colour warmth is reduced

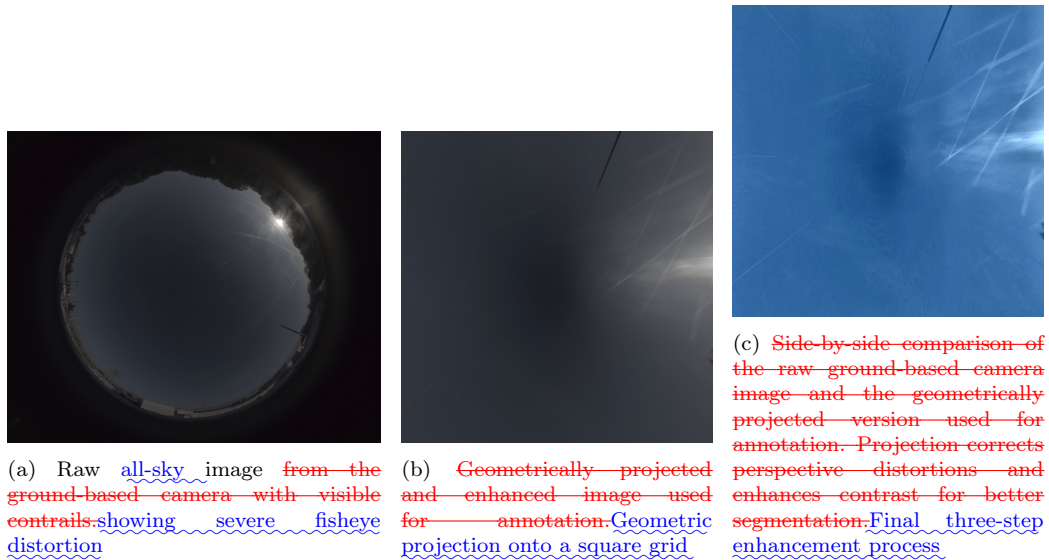


Figure 3: Impact of preprocessing pipeline on contrail visibility.

by rebalancing the blue and red channels, mitigating the effects of high solar glare and improving contrail visibility in high-sunlight challenging lighting conditions. This preprocessing pipeline proved essential in highlighting fine contrail structures, especially in complex atmospheric scenes. Both raw and enhanced projected images are displayed in Figure 3 for two reasons: it enables annotators to identify and delineate contrails consistently across diverse atmospheric scenes, and it simplifies the learning task for computer vision models by removing camera-specific distortions and enhancing the natural linear structure of contrails. Figure 3 illustrates the impact of geometric projection and enhancement, demonstrating how the pipeline reveals contrails that would otherwise be difficult or impossible to annotate reliably. All models presented in this work are trained and evaluated exclusively on preprocessed images.

The process of labelling video sequences included in the dataset were not randomly sampled from the full archive. To ensure sufficient contrail instances for effective model training while maintaining seasonal and atmospheric diversity, we applied a two-stage selection strategy. First, the complete year-long archive was processed using a lightweight binary classifier to distinguish contrail-present from contrail-absent images. This automated filtering efficiently identified candidate periods by excluding extended intervals of clear sky

or heavy low-altitude cloud cover. Second, video sequences were manually selected from these filtered periods, prioritizing scenes with visible, persistent contrails suitable for detailed temporal annotation. This approach deliberately oversamples contrail-positive cases, enhancing the dataset’s utility for segmentation and tracking but introducing a selection bias that should be considered when evaluating model performance on unfiltered operational data. The final dataset spans the full calendar year, ensuring coverage of diverse seasonal and atmospheric conditions.

The labelling process was applied to video sequences; each sequence comprised between 60 and 480 images, corresponding to durations of 30 minutes to 4 hours, enabling the temporal tracking of contrails throughout their formation and dissipation phases.

The labelling process was carried out using a dedicated annotation tool developed by Encord, who also provided a professional team of annotators. We maintained close collaboration with this team through regular coordination meetings, during which the annotation guide was developed and iteratively refined. The labelling platform was specifically configured to overlay flight trajectory data above the camera’s field of view, assisting annotators in identifying “new” contrails, “new” contrails—those forming above the camera and visibly associated with a known aircraft trajectory. In contrast, “old” “old” contrails were defined as those already present at the start of a sequence or likely formed outside the camera’s field of view, making flight association impossible.

Each contrail was annotated using high-precision polygons that tracked its spatial extent throughout its visible evolution, from early linear stages to advanced spreading phases. When contrails became fragmented or partially obstructed by clouds, multiple polygons were used and linked using relational attributes (~~Fragmented contrail and Cloud obstruction~~ “fragmented contrail” and “cloud obstruction”) to preserve temporal continuity.

To ensure the highest annotation quality, the campaign incorporated a multi-stage review protocol. An initial calibration phase was conducted using a sample dataset to harmonise interpretation and identify edge cases. Each labelled sequence then underwent a two-step quality control process: a ~~technical review~~ made ~~technical review~~ by the labelling team, followed by an ~~expert review~~ made ~~expert review~~ by EUROCONTROL to ensure final quality. In total, 4,536 hours of labelling and 431 hours ~~or~~ ~~of~~ reviewing were performed.

611 4.2. Dataset Description

612 The GVCCS dataset ~~Jarry et al. (2025)~~ (Jarry et al., 2025) is the first
613 open-access, instance-level annotated video dataset designed for contrail de-
614 tection, segmentation, and tracking from visible ground-based sky camera
615 imagery. It consists of 122 high-resolution video sequences (totaling 24,228
616 images) captured at the EUROCONTROL Innovation Hub in Brétigny-sur-
617 Orge, France, using Réuniwatt’s CamVision sensor. Each sequence has been
618 carefully annotated with temporally consistent polygon masks for visible con-
619 trails, including multi-instance tracking and, where possible, attribution to
620 specific flights using aircraft trajectory data.

621 In total, the annotation team labelled ~~4651~~ 4,651 individual contrails
622 with a total of ~~176,194~~ 176,194 polygons. The sequences cover a wide range
623 of durations (from 0.5 to 142.5 minutes per contrail), with each contrail com-
624 prising between 1 and 589 polygons (mean: 37.8). On average, each video
625 sequence spans 96.6 minutes and contains approximately 193 annotated im-
626 ages. About ~~3346~~ 3,346 contrails are associated with unique flight identifiers
627 derived from synchronized flight trajectory data filtered above 15,000 ft.

628 The GVCCS dataset is structured into `train/` and `test/` folders, each
629 containing `images`, `annotations.json` (COCO format), and associated flight
630 data in `parquet` format. The dataset supports a range of research tasks
631 including semantic and panoptic segmentation, temporal tracking, lifecycle
632 analysis, and contrail-flight attribution, and is released under the CC BY
633 4.0 license.

634 5. Segmentation Models

635 This section reviews the segmentation models evaluated for identifying,
636 and for some also tracking, contrails. As established in Section 2.2, our
637 primary objective is instance segmentation (detecting individual contrails
638 and assigning them unique identifiers) which is essential for temporal tracking
639 and flight attribution. The models presented here are capable of panoptic
640 segmentation (jointly handling instance identification and scene classification),
641 though our evaluation focuses primarily on contrail instance quality rather
642 than exhaustive scene parsing.

643 We focus on two model families: Mask2Former, a state-of-the-art transformer-
644 based segmentation model, and a U-Net using a discriminative embedding
645 loss. Both are evaluated on individual images, while only Mask2Former is
646 additionally evaluated on videos.

Table 1: Descriptive statistics of the annotated contrail dataset

Metric	Value
Total sequences (labelled)	122
Total images	24,228
Average sequence duration in minutes (<u>minutes</u>)	96.6
Images per sequence (min / max / mean)	41 / 600 / 198.6
Total annotated contrail instances	4651 <u>4,651</u>
Total unique flight IDs assigned	3354 <u>3,346</u>
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

647 We also explore two problem formulations: in the single-polygon case,
648 each visible contrail fragment is treated as an independent instance; in the
649 multi-polygon case, all fragments of a given contrail are labelled as a single
650 instance, even if they are spatially disconnected. The single-polygon setting
651 assumes that a subsequent linking algorithm, not implemented in this work,
652 could later group fragments into full contrails. The multi-polygon formula-
653 tion, in contrast, expects the model to infer such groupings implicitly.

654 5.1. Mask2Former

655 Mask2Former is a universal segmentation architecture that unifies seman-
656 tic, instance, and panoptic segmentation within a single model. It is built
657 around a hierarchical encoder-decoder structure comprising three main com-
658 ponents: a convolutional backbone for multi-scale feature extraction, a pixel
659 decoder that generates dense spatial embeddings, and a transformer decoder
660 with learnable mask queries that iteratively refines segmentation predictions.

661 A central innovation in Mask2Former is its use of ~~the so-called~~-masked
662 attention in the transformer decoder. Unlike standard cross-attention, which
663 considers the entire image, masked attention limits attention to regions sur-
664 rounding the current predicted masks. This localized focus enables more pre-
665 cise refinement of object boundaries, which is particularly beneficial for thin,
666 high-aspect-ratio structures like contrails. The model’s learnable queries act
667 as object proposals and are refined through multiple decoding layers to gen-

668 erate final instance masks and class labels in an end-to-end manner.

669 An important aspect of Mask2Former’s effectiveness lies in its loss func-
670 tion (~~i.e., the training objective~~), which guides the model to learn accurate
671 ~~segmentation masks and their corresponding classes~~ — the mathematical objective
672 that the model seeks to minimize during training. A loss function quantifies
673 the difference between predicted outputs (e.g., segmentation masks) and
674 ground truth annotations, providing the learning signal that guides iterative
675 parameter updates. The loss function used by Mask2Former combines sev-
676 eral components. First, it uses a classification loss that helps the model
677 assign the correct class to each predicted mask (e.g., contrail vs. sky). Sec-
678 ond, it includes a mask loss, which measures how closely the predicted mask
679 matches the ground-truth mask for that object, commonly using a pixel-
680 wise binary cross-entropy or Dice loss. Finally, Mask2Former incorporates
681 a matching step based on the Hungarian algorithm ~~to align predictions with~~
682 ~~ground truth in an optimal~~, (Kuhn, 1955) — a combinatorial optimization
683 method that solves the assignment problem by finding the optimal one-
684 to-one way. ~~This correspondence between two sets given a cost matrix.~~
685 In this context, the algorithm matches each predicted mask with its most
686 appropriate ground-truth object by minimizing a combined cost based on
687 classification and mask similarity. This optimal matching ensures that each
688 ~~predicted mask prediction~~ is evaluated against the ~~most appropriate reference object~~ correct
689 reference, avoiding duplicate ~~assignments or ambiguous assignments, which is~~
690 particularly important when multiple contrails with similar appearance are
691 present in the same image.

692 A detailed technical description of the model is beyond the scope of this
693 paper, as our focus is on applying Mask2Former to contrail segmentation; we
694 refer the reader to the original work by ~~Cheng et al. (2022)~~ Cheng et al. (2022)
695 for a comprehensive overview of the architecture and performance on popular
696 datasets.

697 To capture temporal dynamics inherent in contrail evolution, we ex-
698 tend Mask2Former to process short video sequences. Although designed
699 for single images, the model can handle multiple consecutive frames as a 3D
700 spatio-temporal volume by treating time as an additional axis alongside spa-
701 tial dimensions, following the extension introduced by ~~Cheng et al. (2021a)~~
702 Cheng et al. (2021a).

703 Compared to traditional segmentation models, Mask2Former offers sub-
704 stantial architectural advantages. Mask R-CNN (He et al., 2017), while ef-
705 fective, performs detection and segmentation as separate stages, which can

introduce spatial misalignment and inefficiencies, especially when segmenting long, disconnected objects. DETR (DEtection TRansformer) (Carion et al., 2020), though end-to-end and transformer-based, primarily focuses on object detection and lacks the fine-grained spatial modelling needed for precise mask prediction. MaskFormer (Cheng et al., 2021b) introduces transformer-based decoding for segmentation, but relies on global attention, which can dilute spatial precision. Mask2Former refines this approach with masked attention and iterative refinement, leading to improved accuracy, especially in challenging tasks where objects are often thin, faint, and visually ambiguous.

5.2. U-Net with Discriminative Loss

As a baseline, we implement a ~~two steps~~ two-step instance segmentation model. First, we use a ~~classical U-net architecture~~ Jarry et al. (2024) U-Net architecture (Ronneberger et al., 2015) for segmentation. U-Net is ~~designed specifically for image segmentation tasks and a convolutional neural network originally developed for biomedical image segmentation, characterized by its distinctive U-shaped architecture.~~ The network features a symmetrical encoder-decoder structure. ~~The encoder part of the network gradually reduces the spatial size of the input image, extracting:~~ the encoder progressively downsamples the input to capture high-level ~~features that capture the overall context.~~ features that capture the overall context. ~~The decoder then progressively restores the spatial resolution by upsampling these features to produce a segmentation map that matches the original image size. Importantly~~ semantic features, while the decoder upsamples to recover spatial resolution. Crucially, U-Net ~~uses skip connections that directly link corresponding layers in the~~ employs skip connections—direct pathways that link corresponding encoder and decoder layers, bypassing intermediate processing. These connections allow fine-grained spatial details ~~(such as exact contour boundaries) that are~~ lost during downsampling to be ~~recovered~~ directly recovered in the decoder, improving the quality and precision of segmentation outputs.

Second, we use a similar architecture that learns a unique feature representation, or embedding, for each pixel in an image by using a discriminative loss function ~~—~~ a training objective specifically designed to encourage pixels from the same instance to have similar embeddings while pushing apart embeddings from different instances. In this model, the final head of the U-Net does not produce a typical segmentation map with class labels. Instead, it produces an embedding for each pixel ~~;~~ (a vector in a high-dimensional feature space). The goal is for pixels that belong to the same object instance

to have similar embeddings (meaning they are close together in this feature space), while pixels belonging to different instances have embeddings that are far apart. This way, the model effectively learns to group pixels based on their learned features.

The process of identifying individual instances is performed in two separate steps. The first step is to generate these pixel embeddings with the U-Net, and the second step is to group or cluster these embeddings into individual instances. For clustering, we use ~~the HDSCAN algorithm, to find the clusters and a final k-means to associate outliers with closest cluster.~~ HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) (Campello et al., 2013)—a density-based clustering algorithm that automatically identifies clusters of arbitrary shape without requiring a predetermined number of clusters. HDBSCAN groups pixels with similar embeddings (high local density in the embedding space) into the same instance while identifying outliers that do not belong to any clear cluster. These outliers are subsequently assigned to the nearest cluster using k -means, ensuring complete instance coverage. This approach is particularly suitable for contrails, which often exhibit irregular, fragmented, or elongated shapes that are difficult to cluster using traditional methods like k -means alone.

The discriminative loss function used to train the model is composed of three parts. The first part, known as the pull term, encourages embeddings of pixels that belong to the same instance to be close together, making the cluster compact. The second part, called the push term, forces embeddings of different instances to be sufficiently separated from each other, preventing clusters from overlapping. The third part is a regularization term that prevents the embeddings from growing too large in magnitude, which stabilizes the training process and embedding space. This combination allows the model to learn meaningful and well-separated pixel embeddings without relying on explicit object bounding boxes or pre-defined region proposals. For readers interested in the mathematical formulation and detailed rationale behind the discriminative loss, we refer to the original paper by ~~De Brabandere et al. (2017)~~ De Brabandere et al. (2017).

It is important to note that this model operates only on single images. Unlike models such as Mask2Former for videos mentioned in the previous section, it does not incorporate any temporal or sequential information, nor does it include recurrent layers or mechanisms to handle videos. Extending this approach to process video sequences and incorporate temporal consistency would require significant changes to both the architecture and the algorithms

781 used, which is outside the scope of this work.

782 The embedding-based approach is well suited to segmenting objects that
783 may not be spatially continuous, such as contrails with fragmented shapes.
784 Since the model does not require spatial continuity, it can learn to embed
785 separate, disconnected parts of the same contrail into a similar region of
786 the feature space if they share common visual characteristics and belong to
787 the same label. However, this approach has its challenges. If parts of the
788 same contrail differ significantly in appearance ~~;~~ due to factors like changes
789 in lighting, atmospheric conditions, or variations in the background texture:
790 ~~They~~, they may be embedded differently and incorrectly assigned to separate
791 clusters. Conversely, visually similar but unrelated contrail fragments could
792 be mistakenly grouped together, as the model relies solely on the learned
793 embeddings for clustering.

794 Figure 4 illustrates ~~a qualitative result of the instance discriminative~~
795 ~~segmentation model~~ how the discriminative embedding approach learns to
796 separate contrail instances. On the left, the ground truth labels are dis-
797 played, highlighting the pixel-wise assignment to contrail instances. On the
798 right, we show the corresponding discriminative embedding space, ~~reduced to~~
799 ~~two dimensions using~~. Since each pixel is represented by a high-dimensional
800 embedding vector (typically 32 dimensions), we apply Principal Component
801 Analysis (PCA) ~~for visualization purposes to reduce this to two dimensions for~~
802 visualization: PCA identifies the two orthogonal directions that capture the
803 most variance in the embedding space, effectively projecting the high-dimensional
804 clusters onto a 2D plane. Each point ~~represents a pixel embedding, and~~
805 ~~colors indicate the instance it belongs to in this plot~~ represents a single pixel,
806 colored according to its ground-truth instance label. This visualization pro-
807 vides insight into how the model, trained with a discriminative loss, learns
808 to embed pixels from the same instance close together in the feature space,
809 while separating those from different instances. The separation observed in
810 the embedding space confirms the model’s ability to cluster fragmented con-
811 trail structures, although visually similar but unrelated segments may still
812 partially overlap in the embedding due to shared appearance features.

813 6. Results

814 This section presents the performance of the models introduced in Sec-
815 tion 5 on contrail segmentation tasks. Our primary goal is not to achieve
816 state-of-the-art results but to establish clear ~~examples of application~~ application

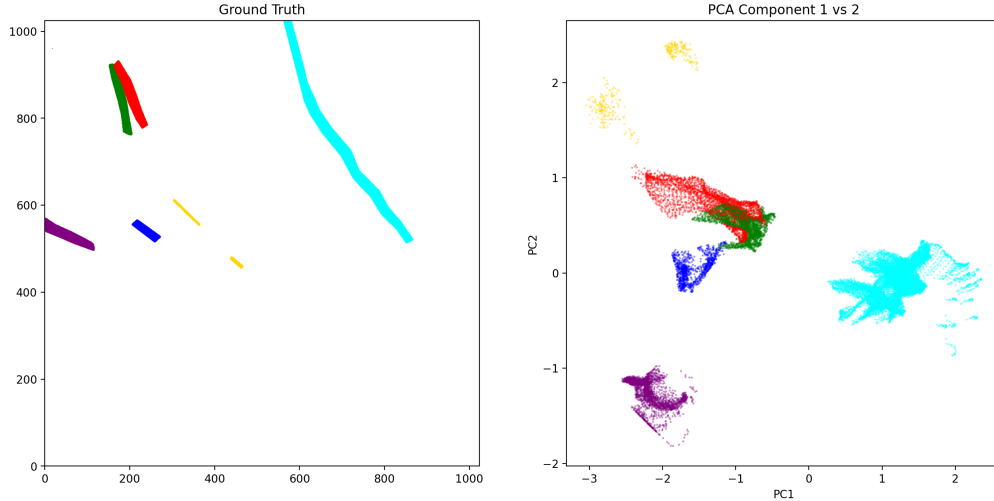


Figure 4: ~~The true label is displayed on the left and the~~ Illustration of the ~~left and the~~ discriminative embedding ~~on the right~~ method for instance segmentation. ~~The latter was created using Principal Component Analysis~~ Left panel: Ground-truth (PCA human-annotated) instance labels, where each color represents a distinct contrail. Right panel: Visualization of the learned pixel embeddings. The ~~colours reflect~~ U-Net model learns to map each pixel to a point in a high-dimensional feature space such that pixels belonging to the same contrail ~~instances~~ are positioned close together, while pixels from different contrails are far apart. For visualization, PCA reduces this high-dimensional space to two dimensions by identifying the directions of maximum variance. Each point represents one pixel, positioned according to its learned embedding and colored by its ground-truth contrail instance. Well-separated, compact clusters indicate that the model has successfully learned to group pixels from the same contrail while distinguishing different contrails.

817 examples and meaningful baseline performances. By doing so, we highlight
 818 the unique opportunities offered by this dataset and provide a foundation
 819 for the research community to build upon, encouraging rapid progress in the
 820 critical field of aviation’s climate impact.

821 6.1. Training

822 All models were ~~initialised~~ initialized from existing pretrained check-
 823 points. We trained two versions of the Mask2Former architecture for the
 824 ~~single image~~ single-image segmentation task. Both models share the same
 825 core architecture but differ in the size of their transformer backbone: one
 826 uses the Swin-Base (Swin-B) configuration and the other uses the larger
 827 Swin-Large (Swin-L). The main difference between these two lies in model

828 capacity. Swin-L has significantly more parameters, ~~which enables enabling~~
829 it to learn richer representations at the cost of higher computational require-
830 ments.

831 Both image models were initialized from publicly available pretrained
832 checkpoints in the Mask2Former Model Zoo¹. Each model was first pre-
833 trained on the ImageNet-21k (IN21k) (Ridnik et al., 2021) classification
834 dataset and then fine-tuned on the COCO panoptic segmentation dataset.
835 While COCO (Lin et al., 2014) does not include contrails, it spans a wide
836 range of natural (including clouds and sky) and man-made objects, offering
837 useful general-purpose segmentation features. This two-stage pretraining
838 (IN21k followed by COCO) has been widely validated in the literature and
839 provides a strong initialization for fine-tuning on contrail imagery.

840 Both the Swin-B and Swin-L variants were trained on individual image
841 frames using 200 learnable object queries. Given our hardware setup, two
842 NVIDIA RTX 6000 GPUs, each with 48 GB of memory, we were able to
843 train both variants on the image dataset without significant memory limita-
844 tions.

845 For video segmentation, we used the video-specific variant of Mask2Former,
846 which extends the original architecture to handle temporal sequences. Like
847 the image-based model, it ~~also~~ uses 200 object queries and Swin Transformer
848 backbones, and ~~it~~ is initialized from a checkpoint pretrained on the YouTube-
849 VIS 2019 dataset (Yang et al., 2019). Although YouTubeVIS does not contain
850 contrails, its emphasis on learning temporally consistent object masks across
851 frames makes it well suited to capture the dynamics of contrails in video
852 data.

853 Due to GPU memory constraints, we limited both training and inference
854 to short video clips composed of a small number of consecutive frames. While
855 this restriction was necessary to fit within available hardware resources, par-
856 ticularly for memory-intensive architectures, it also shaped our training strat-
857 egy. During training, these clips are randomly sampled from longer video se-
858 quences to introduce temporal diversity ~~into the training process~~. By varying
859 the starting points of the sampled clips, the model is exposed to contrails at
860 different stages of their lifecycle (formation, elongation, dissipation) and
861 in diverse atmospheric contexts. This stochastic sampling encourages the
862 model to learn more generalizable temporal representations.

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

863 To support this setup, we trained the video Mask2Former model using
 864 both Swin-Base (Swin-B) and Swin-Large (Swin-L) backbones. However, the
 865 number of frames per clip had to be adjusted based on model capacity and
 866 memory availability. With the more lightweight Swin-B variant, we were able
 867 to train on 5-frame clips, while the higher-capacity Swin-L model could only
 868 be trained on 3-frame clips due to its significantly larger memory footprint.
 869 This reflects a trade-off between temporal context and model expressiveness:
 870 longer clips may better capture the dynamic evolution of contrails, whereas
 871 larger models like Swin-L provide richer per-frame representations. Training
 872 both configurations allows us to explore how these two dimensions ~~;~~ (temporal
 873 depth and model capacity~~;~~) interact in the context of contrail segmentation.

874 For the U-Net model, we used a backbone based on MaxViT-B, a hybrid
 875 vision transformer architecture that combines convolutional layers with self-
 876 attention mechanisms for efficient and scalable visual representation learning.
 877 This backbone was pretrained on ImageNet-21k and subsequently fine-tuned
 878 on ImageNet-1k, providing robust feature representations to support the dis-
 879 criminative loss function employed during contrail segmentation training.

880 The training procedure for each model involved several epochs of su-
 881 pervised learning, with early stopping applied based on performance on a
 882 validation set. The dataset was partitioned into training, validation, and
 883 test sets using a 70-10-20 random split ~~;~~ ~~done~~ at the video level. This means
 884 that all frames from a given video were assigned exclusively to one of the
 885 three sets to avoid any potential data leakage. To ensure ~~a~~ fair and unbiased
 886 evaluation, we also balanced the number of empty sequences ~~;~~ ~~videos that~~
 887 ~~contain no contrails~~ ~~;~~ videos containing no contrails across the three
 888 subsets.

889 It is important to note that the reported metrics reflect model performance
 890 on contrail-rich scenarios, as the dataset construction deliberately oversampled
 891 contrail-positive sequences to maximize training signal. While this choice
 892 enhances the dataset’s utility for contrail detection and tracking tasks, generalization
 893 to unfiltered operational data with arbitrary sky conditions may differ and
 894 warrants further investigation.

895 We did not perform exhaustive hyper-parameter tuning for any of the
 896 models. Instead, our goal with this experimental setup was to establish
 897 baseline results and to analyze model performance both qualitatively and
 898 quantitatively under realistic computational and data constraints. All mod-
 899 els were trained using the default hyper-parameters reported in their orig-
 900 inal publications. Tables 2 and 3 summarize the most important training

901 parameters for each model. Note that the models differ in the specific hyper-
 902 parameters relevant to their architecture and training setup. Future work
 903 will focus on exploring more sophisticated modeling strategies, systematic
 904 hyper-parameter optimization, and additional training refinements.

Table 2: Default hyper-parameters for Mask2Former models.

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 <u>Rotation (90°); vertical flip; horizontal flip</u>	Applied at image level (Image); applied at clip level (Video) <u>Applied at image level (Image); applied at clip level (Video)</u>

905 ~~Remember that each~~ Each model was trained and evaluated on two dis-
 906 tinct formulations of the instance segmentation task. The first formulation
 907 treats a contrail as a single object, even if it is composed of multiple dis-
 908 connected regions or fragmented segments. In this setup, the model must
 909 learn to group visually and spatially separated regions that correspond to
 910 the same physical contrail. The second ~~task~~ formulation simplifies the prob-
 911 lem by treating each visible polygon as an independent instance. In this
 912 formulation, the model is not required to group disjoint segments belonging
 913 to the same contrail; instead, it simply detects and segments each distinct
 914 region. This approach corresponds to a modular processing pipeline where
 915 instance merging and flight attribution occur at a later stage, as will be
 916 discussed in future work.

Table 3: Default hyper-parameters for U-Net model trained with discriminative loss.

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. Evaluation

We evaluate both semantic and instance-level segmentation performance using a combination of standard and task-adapted metrics.

For semantic segmentation, we report ~~pixel-wise scores such as mean intersection over union~~ the mean Intersection over Union (mIoU) and the Dice coefficient. ~~For instance segmentation, we adopt the COCO evaluation protocol with modifications to better reflect the thin, elongated structure of contrails. All metrics are computed globally over the full test set. In the sections that follow, we describe our evaluation procedure, sliding window inference strategy for video models, and the rationale behind our choice of metrics. The presentation and interpretation of~~ Both metrics quantify the overlap between predicted and ground-truth masks, with values ranging from 0 (no overlap) to 1 (perfect match).

The mIoU is calculated as:

$$\text{mIoU} = \frac{\text{Area of Intersection}}{\text{Area of Union}},$$

where the intersection is the set of pixels correctly predicted as contrail, and the union includes all pixels predicted as contrail plus all true contrail pixels. This metric equally penalizes both false positives (predicting contrail where there is none) and false negatives

935 (missing actual contrail pixels).

936 The Dice coefficient is calculated as:

$$\text{Dice} = \frac{2 \times \text{Area of Intersection}}{\text{Size of Prediction} + \text{Size of Ground Truth}}.$$

937 The factor of 2 in the numerator makes the Dice coefficient emphasize
938 correct overlap more strongly than mIoU. It is particularly sensitive to small
939 or thin structures, making it well-suited for evaluating contrails, which often
940 appear as narrow, elongated features that occupy a small fraction of the
941 image.

942 *Temporal Evaluation Strategy*

943 For video-based models, inference is performed using a sliding window
944 approach, where each video is divided into overlapping short clips of fixed
945 length, matching the clip length used during training (e.g., 3 frames for the
946 Swin-L model, 5 frames for the Swin-B model). These clips advance by one
947 frame at a time (stride one), allowing the model to leverage temporal context
948 effectively while respecting memory constraints during inference. Crucially,
949 segmentation accuracy is computed only on the central frame of each short
950 clip. This design ensures that each frame in the video contributes exactly
951 once to the evaluation metrics, only when it appears as the center frame
952 of a clip. This prevents duplicate evaluation and enables ~~a~~ fair comparison
953 with image-based models, which predict on single frames independently. For
954 example, if a 5-frame clip is used on a video with frames numbered 1 through
955 10, the first evaluation clip spans frames 1–5 with evaluation on frame 3;
956 the next clip covers frames 2–6 (evaluated on frame 4), and so on. This
957 guarantees unique evaluation for frames 3 to 8, each exactly once.

958 It should be noted that the video-based Mask2Former model maintains
959 temporally consistent instance identifiers within each clip. That is, if a con-
960 trail is labelled as instance #3 in one frame of a clip, it retains this identifier
961 across all frames in the same clip. However, since clips are processed inde-
962 pendently, these identifiers are not guaranteed to remain consistent between
963 consecutive clips. A given contrail may receive a different identifier in adja-
964 cent clips. To enable continuous tracking of contrails throughout the entire
965 video, we introduce a simple post-processing method that links and recon-
966 ciles these instance identifiers to generate coherent, continuous tracks; this
967 method is described in detail in Appendix A.

Instance Segmentation Metrics

Model performance is evaluated using both semantic and instance-level segmentation metrics. All metrics are computed globally by aggregating predictions and ground truths across the entire test set before applying the metric calculations. This global computation prevents biases that can arise from averaging metrics computed independently on each observation (i.e., frame), which is particularly important in settings with imbalanced or sparse data such as contrail segmentation.

~~For semantic segmentation, we report the mean Intersection over Union (mIoU) and the Dice coefficient. The mIoU measures the overlap between the predicted and ground truth binary masks by calculating the ratio of the intersection area to the union area of the masks, thus penalizing both false positives and false negatives. The Dice coefficient, defined as twice the area of overlap divided by the total size of the predicted and ground truth masks, emphasizes the correct overlap and is especially sensitive to thin or fragmented structures, making it a suitable metric for evaluating contrails.~~

Instance segmentation performance is assessed using COCO-style metrics (Lin et al., 2014) computed globally over the dataset. To accommodate the specific challenges posed by contrails, we adapt the IoU threshold range and denote metrics with the following notation: $\text{AP}_{\text{X@}[\text{IoU range} \mid \text{size category} \mid \text{size category} \mid \text{max detections}]}$, where IoU range specifies the range of IoU thresholds over which

- IoU range : The range of Intersection over Union thresholds used. A prediction is considered a "true positive" only if its IoU with a ground-truth object exceeds the threshold. Average Precision (AP) ~~or Average Recall (AR) is computed~~, size category indicates the object size subset considered, and max detections ~~is the~~ is computed across multiple thresholds and averaged.
- Size category : Filters objects by area — "small" ($< 32^2$ pixels), "medium" (32^2 to 96^2 pixels), "large" ($> 96^2$ pixels), or "all" (no filtering).
- Max detections : The maximum number of ~~detections per image considered~~ predicted instances considered per image (e.g., 100).

For example, $\text{AP}_{\text{X@[0.25:0.75] \mid \text{all} \mid \text{all} \mid 100]}}$ denotes ~~the mean average precision calculated~~ Average Precision computed over IoU thresholds ranging

from 0.25 to 0.75, ~~considering across~~ all object sizes ~~and up to~~, with a maximum of 100 detections evaluated per image. ~~Object size categories follow the standard definitions used in COCO-style metrics: small objects have an area less than $32^2 = 1,024$ pixels; medium objects range between 32^2 and $96^2 = 9,216$ pixels; large objects exceed 96^2 pixels. Metrics such as AP@0.25:0.75 | small | 100~~ then reflect the performance specifically on small-sized objects, under the specified IoU and detection constraints. In the results that follow, we report both Average Precision (AP) and Average Recall (AR) using this notation.

We restrict the IoU threshold range to $[0.25, 0.75]$, rather than the standard COCO range of $[0.50, 0.95]$, to better accommodate the elongated and thin geometry of contrails, where very high IoU thresholds are overly strict. Contrails are thin, irregular, and may extend across large image portions, making exact mask overlap challenging. ~~Under typical COCO metrics, a prediction with partial but semantically correct overlap might be unfairly penalized. For example, a predicted mask overlapping only~~ A prediction overlapping 30% of a contrail would be ignored under COCO’s default minimum IoU of 0.5, but counted as a true positive under our more lenient thresholds. This adaptation better reflects practical segmentation quality for contrails.

By adjusting the IoU range, the metrics better reflect practical segmentation quality for contrails, balancing sensitivity to spatial accuracy with tolerance for slight misalignments and fragmentations inherent to this domain. It is important to note that these adapted metrics are not directly comparable to standard COCO scores but are specifically tailored to provide meaningful evaluation in the context of contrail segmentation.

This evaluation framework, combining semantic and instance segmentation metrics computed globally with appropriate IoU thresholds and size categories, offers a comprehensive and interpretable means of assessing model performance. It facilitates fair comparisons across models and supports future benchmarking on our contrail dataset.

Tables 4 and 5 summarize the results for the semantic and instance segmentation tasks, respectively. All results are reported for both single-image and video-based models. Instance segmentation results are further disaggregated by annotation style: **M** refers to multi-polygon annotations, and **S** refers to single-polygon annotations. For Mask2Former models, values without parentheses correspond to the Swin-B backbone, while those in parentheses refer to Swin-L.

Table 4: Semantic segmentation metrics. For the Mask2Former variants, values without parentheses refer to Swin-B; values in parentheses refer to 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)

In the semantic segmentation task, performance remains consistent across all models and variants, with Dice and mIoU scores showing little variation. This stability is expected, as semantic segmentation only requires classifying each pixel as either contrail or sky, without distinguishing between separate contrail instances. The U-Net model achieves results on par with the more advanced Mask2Former models, indicating that per-pixel contrail detection is largely driven by local visual features, such as shape, brightness, and texture, which U-Net captures effectively.

These results also reflect the quality and consistency of our dataset: although based on ground-level imagery, the segmentation performance is in line with results reported in previous studies using satellite data (Jarry et al., 2024; Ortiz et al., 2025). Although differences in imaging modality and scene geometry preclude direct comparisons, the consistency in results suggests that semantic contrail segmentation is a well-posed task for modern architectures, with strong performance achievable across diverse data sources.

Instance segmentation results reveal clear differences between model architectures. These differences are more substantial than those observed in the semantic segmentation task, highlighting the added complexity introduced by instance-level reasoning. Mask2Former, which is designed for panoptic segmentation through object-level queries and global spatial reasoning, consistently outperforms U-Net across all instance metrics. The performance gap is particularly pronounced in the multi-polygon setting, where contrails appear fragmented and must be correctly grouped into coherent instances. These results highlight the value of architectures specifically built for instance-aware tasks: Mask2Former’s ability to reason globally and associate disjoint segments makes it better suited for detecting and tracking individual contrails.

A more nuanced comparison emerges when evaluating image-based versus video-based Mask2Former models. For the Swin-B backbone, the image-

Table 5: Instance segmentation metrics. "M" refers to multi-polygon, whereas "S" indicates single-polygon. For the Mask2Former variants, values without parentheses refer to Swin-B; values in parentheses refer to Swin-L.

Type	Metric	Single Images		Videos
		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)

1069 based model achieves higher instance segmentation performance, while the
1070 video-based model slightly outperforms it on semantic segmentation metrics.
1071 This suggests that although video models benefit from temporal consistency

1072 and motion cues, the added complexity of enforcing cross-frame coherence
1073 may introduce challenges that slightly hinder instance-level prediction accu-
1074 racy, particularly when using a lower-capacity backbone like Swin-B.

1075 In the Swin-L setting, the image-based model performs best overall. It
1076 achieves both the highest instance segmentation score and slightly superior
1077 semantic segmentation performance. These results indicate that temporal
1078 ~~modelling~~ modeling does not always yield performance improvements, espe-
1079 cially when the temporal context is limited (e.g., 3-frame clips) or when the
1080 spatial representation capacity of the model is already high. The image-
1081 based model benefits from pretraining on COCO, which may ~~favour~~ favor
1082 precise spatial delineation, while the video-based variant relies on pretrain-
1083 ing on YouTubeVIS, which is more focused on temporal coherence. However,
1084 it is important to note that the video-based model performs an additional
1085 task: tracking. By maintaining consistent instance identities across frames, it
1086 enables temporally coherent segmentation that is not achievable with image-
1087 based models. ~~All in all, the~~ The metrics reported here are computed on a
1088 per-frame basis and do not account for flickering or instance identity consis-
1089 tency over time. These temporal aspects are particularly important in video
1090 applications and are not captured by the conventional frame-level evaluation
1091 scores presented herein.

1092 An important caveat is that all reported metrics are computed independently
1093 for each frame and do not account for temporal consistency of instance
1094 identities over time. Video-based models are explicitly trained to maintain
1095 coherent instance tracks across frames through end-to-end temporal modeling,
1096 jointly optimizing segmentation and tracking within a unified objective. In
1097 contrast, image-based models require post-hoc association algorithms (such
1098 as the Hungarian matching method described in Appendix A) to link instances
1099 temporally based on spatial overlap alone. While both approaches can achieve
1100 tracking, video models learn temporal correspondences from motion cues
1101 and appearance features during training, potentially offering more robust
1102 handling of occlusions, fragmentations, and brief disappearances. However,
1103 the per-frame metrics reported here (AP, AR, Dice, mIoU) primarily assess
1104 spatial segmentation quality and do not reward temporal consistency. As
1105 a result, while video models do not uniformly outperform image models
1106 in per-frame scores, they provide qualitative benefits in terms of reduced
1107 instance ID flickering and smoother temporal transitions that are not captured
1108 by these metrics. Future work should incorporate video-specific evaluation
1109 metrics (e.g., tracking accuracy, ID switches, fragmentation) to fully characterize

the advantages of temporal modeling. Additionally, the short clip lengths used in this study (3–5 frames) were dictated by hardware constraints; longer temporal contexts may yield further improvements and warrant investigation with more capable architectures.

Overall, Swin-L outperforms Swin-B across all setups, reinforcing the benefit of increased model capacity for fine-grained spatial understanding and instance-level reasoning. Nonetheless, this comes at the cost of higher computational requirements, particularly in the video setting, underscoring a trade-off between performance and scalability.

Another important trend observed in the evaluation is that model performance is strongly influenced by contrail size and detection caps. Generally speaking, larger contrails are segmented more accurately due to their higher pixel counts and lower ambiguity, while allowing more predicted instances (e.g., increasing the detection limit) improves recall by removing constraints on how many objects can be reported. These trends are consistent with general findings in object detection and reinforce the shared challenges between contrail segmentation and broader instance segmentation tasks.

Comparing the multi-polygon and single-polygon formulations reveals a difference in task difficulty: the single-polygon setting is inherently easier. Across all models and data modalities, instance segmentation metrics are consistently higher when using the single-polygon formulation. This is because the task removes the need to group fragmented or spatially disjoint contrail segments into separate instances. Instead, all parts of a contrail, regardless of their separation, are treated as a single mask, greatly simplifying the model’s objective. The model is no longer required to learn complex grouping strategies or reason over spatial and temporal discontinuities. Note that semantic segmentation metrics remain virtually unchanged between the two formulations, indicating that identifying contrail pixels is equally feasible in both cases. The difference lies solely in how those pixels are grouped into instances. This distinction confirms that the main challenge in the multi-polygon task is not pixel classification but instance association.

These results have important practical implications for different contrail detection scenarios. For older contrails, such as those typically observed in satellite imagery or in ground-based images when the contrail formed outside the camera’s field of view, it is extremely difficult to associate the contrail with its source flight. In these cases, the only viable option is to group visible fragments into instances based solely on visual information. This makes multi-polygon instance segmentation essential, as it allows models to

detect and associate disjoint contrail segments without relying on external data. Our dataset and Mask2Former-based models are specifically designed for this setting, enabling effective instance-level detection even when contrails are fragmented, occluded, or spatially disconnected.

In contrast, when a contrail forms directly above the camera and additional data such as aircraft trajectories and wind fields are available, a different approach becomes feasible. In these situations, one can perform single-polygon instance segmentation, where contrail fragments are grouped into a single instance using post-hoc association based on flight paths and advection. This formulation is simpler from a computer vision perspective and is commonly used in the literature (Ortiz et al., 2025; Chevallier et al., 2023; Van Huffel et al., 2025), mainly because multi-polygon annotated datasets have not been available until now. However, this method depends on access to external data and is only applicable to contrails formed during the observation window, after the aircraft has entered the scene.

By supporting both the multi- and single-polygon formulations, our dataset enables training and evaluation across a broader set of operational use cases. The multi-polygon task is essential for vision-only detection of older contrails or those in satellite imagery, while the single-polygon formulation may be more suitable when additional metadata enables contrail-to-flight attribution. This distinction will be further explored in future work focused on linking contrails to their source aircraft.

6.3. Illustrative examples

We present two test-set examples to illustrate the challenges of the multi-polygon contrail segmentation task. In both cases, we compare predictions from image-based and video-based versions of the Mask2Former model, trained from pretrained Swin-L backbones. These examples highlight how temporal context affects instance predictions and expose typical failure modes, including contrail fragmentation, occlusion by clouds, and confusion between contrails and visually similar cloud structures.

Figure 5 shows a frame from April 25th, 2024 at 05:51:00 (UTC), under clear-sky conditions. The background is uniformly blue, providing ~~favourable~~ favorable conditions for both human and machine segmentation. The corresponding ground-truth annotations include several contrails labelled as fragmented (e.g., identifiers 0, 1, and 5), based on known flight trajectories available to annotators during the labelling process. This makes the example suitable for evaluating instance-level understanding in the multi-polygon setting.

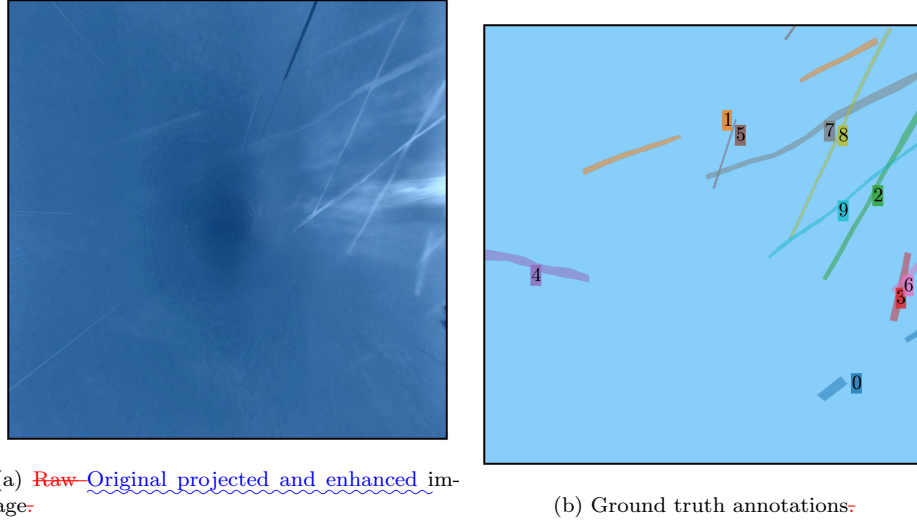


Figure 5: Raw image Original projected and enhanced as well as ground truth annotations for April 25th, 2024 at 05:51:00:00 (UTC).

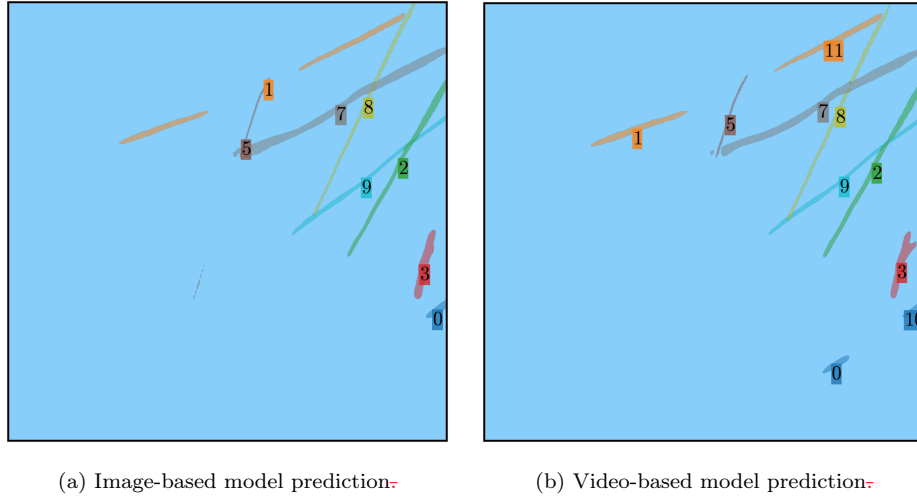


Figure 6: Predicted instances for the frame shown in Fig. 5, using Swin-L models with image and video inputs.

Despite the favourable Figure 6 shows predictions from both models for this scene. Despite the favorable background, both models exhibit instance-level errors. The image-based model correctly infers that contrail 1 is fragmented but detects just one segment of contrail 0, missing the other entirely.

1189 It completely misses contrail 4 and erroneously merges contrails 5-3 and 6
 1190 into a single prediction. The video-based model makes similar mistakes: it
 1191 also merges contrails 5-3 and 6, and fails to detect contrail 4. Additionally,
 1192 it predicts the second fragment of contrail 0 but assigns it to a different
 1193 instance, and it incorrectly splits contrail 1 into two separate instances.

1194 From a semantic segmentation perspective, both models perform rela-
 1195 tively well, as expected in a high-contrast scene. The image-based model
 1196 achieves a Dice score of 0.76 and a mean IoU of 0.64, while the video-based
 1197 model slightly outperforms it with a Dice of 0.79 and mean IoU of 0.67. How-
 1198 ever, due to the instance grouping errors, the image model achieves a slightly
 1199 higher AP@[0.25:0.75 all-all] 100] (0.62) than the video model (0.55).

1200 Figure 7 shows a more challenging frame captured on November 19th,
 1201 2023 at 08:49:30-30 (UTC). Here, several cirrus clouds are present in the
 1202 background, which introduces ambiguity, as some of these cloud structures re-
 1203 semble contrails. This scene also includes multiple contrails that are spatially
 1204 aligned and fragmented, increasing the complexity of the instance segmenta-
 1205 tion task.

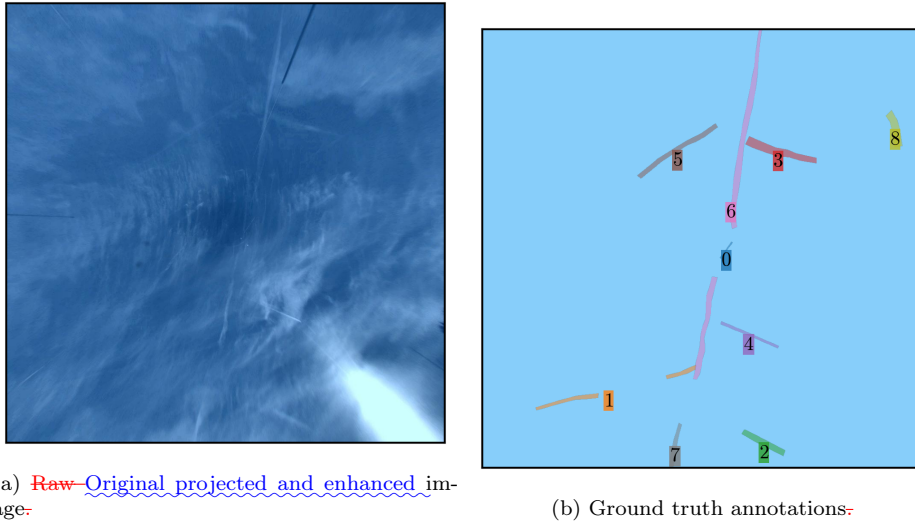
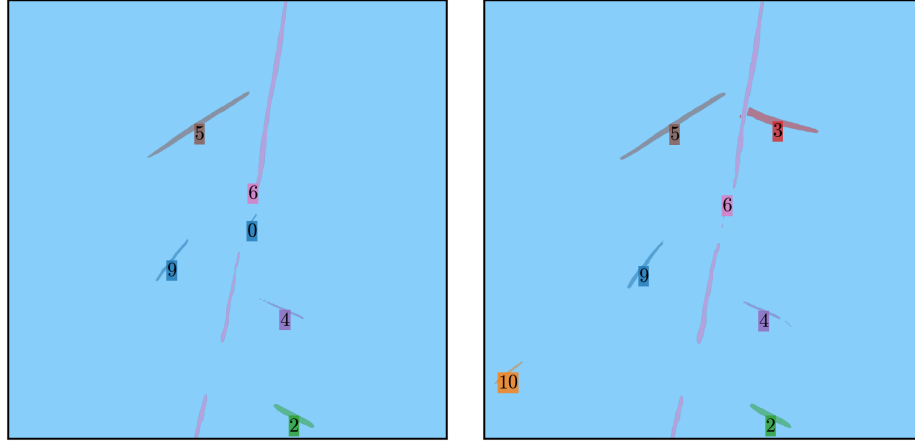


Figure 7: Raw image-Original projected and enhanced image as well as ground truth annotations for November 19th, 2023 at 08:49:30-30 (UTC).

1206 This scene illustrates a common failure mode: fragmentation and mis-
 1207 grouping of visually aligned but semantically distinct contrails. Contrail 6
 1208 is split into two segments with contrail 0 lying in between; although they ap-



(a) Image-based model prediction

(b) Video-based model prediction

Figure 8: Predicted instances for the frame shown in Fig. 7, using Swin-L models with image and video inputs.

pear collinear, contrail 0 is a distinct instance generated by a separate flight. Contrail 7 appears shortly after and may be misassociated with contrails 6 and 0 in the absence of flight metadata. The image-based model correctly separates contrail 0 from 6, but incorrectly merges contrails 6 and 7. The video model groups all three (6, 0, and 7) into a single prediction. Interestingly, this error reflects a plausible human interpretation without flight context, highlighting the challenge of the task.

Both models fail to detect contrails 1 and 8, which are partially occluded by clouds. They also produce a false positive (labelled as contrail 9), segmenting a cirrus structure that resembles a contrail. While the dataset is of high quality and was carefully annotated with access to flight information, some visually ambiguous cases, such as the one discussed, remain inherently difficult to label with certainty. In this example, the predicted region resembles a contrail in both structure and intensity, making it unclear whether the false positive stems from a model error or an understandable omission in the ground truth. These rare edge cases highlight the potential influence of mild label noise in visually complex scenes. Future work could benefit from complementary strategies such as confident learning (Northcutt et al., 2021) to further refine annotations and improve robustness in borderline cases.

Semantic segmentation performance in this scene is lower than in the previous one, reflecting increased difficulty. The image model achieves a

Dice score of 0.61 and mIoU of 0.43, while the video model scores 0.70 and 0.54, respectively. Instance-level AP@[0.25:0.75 ~~all~~ all | 100] ~~s~~ scores are 0.35 and 0.37, respectively, similar to the average metrics, making this a representative case.

These examples illustrate several key challenges in multi-polygon contrail segmentation: (1) correct grouping of fragmented contrail segments from the same flight; (2) visual ambiguity due to clouds that resemble contrails; (3) occlusion; and (4) spatial overlap of contrails from different flights. While video-based models benefit from temporal information, they may over-group distinct instances. Image-based models avoid this but often fail to connect fragmented segments. Overall, these examples demonstrate the inherent difficulty of the task and the limitations of current models.

7. Conclusions

This work introduces a new dataset (Jarry et al., 2025) and baseline models for contrail segmentation from ground-based camera imagery. Our experiments show that modern computer vision methods, particularly panoptic segmentation models like Mask2Former, can be effectively applied to this task, especially when using large pretrained models and temporal information. However, performance gains often come at the cost of increased computational and memory demands, highlighting a trade-off between accuracy and practicality.

The main contribution of this study is the release of the first ~~video annotated~~ video-annotated dataset specifically designed for instance-level contrail segmentation, tracking, and flight attribution in the visual spectrum. Along with detailed evaluation metrics, including average precision and recall across multiple intersection-over-union thresholds and object size bins, this benchmark provides a reproducible baseline for further research in this emerging field.

A key limitation of our current setup is that the visible-light camera restricts observations to daytime conditions. Yet contrails often have their greatest radiative impact at night, when they ~~trap~~ reduce outgoing long-wave radiation and contribute to atmospheric warming. To address this, we are ~~deployed~~ deploying a co-located infrared imaging system that enables continuous, day-and-night monitoring. This may also allow us to begin estimating the radiative forcing of individual contrails under real atmospheric conditions.

1266 In parallel, we are working on a contrail-to-flight attribution algorithm
1267 that links observed contrails to specific aircraft using ~~automatic dependent~~
1268 ~~surveillance broadcast~~ (ADS-B) trajectory data. This tool, and the associ-
1269 ated data and code, will be openly released in a future publication. Attribu-
1270 tion is of utmost importance because it allows each contrail to be linked to
1271 detailed aircraft and engine parameters, such as aircraft type, engine model,
1272 fuel burn rate, flight altitude, and ambient conditions. These inputs are nec-
1273 essary to reproduce the contrail using physical models like CoCiP, assess its
1274 expected properties (e.g., ice crystal number, optical depth, lifetime), and
1275 ultimately validate or refine these models using real-world observations.

1276 We are also extending this work by annotating a new dataset of con-
1277 trails in satellite imagery, with instance-level and sequence-based labels.
1278 This dataset will allow us to test and evaluate the full multi-scale track-
1279 ing pipeline proposed in this paper: starting from high-resolution, ground-
1280 based detection, followed by attribution to flights, and finally linking to the
1281 same contrails as they evolve in satellite imagery. This approach offers a
1282 unique opportunity to study contrail formation, spreading, and dissipation
1283 over time and at scale. We also plan to use our ground-based dataset to eval-
1284 uate the predictions of physical models such as CoCiP. Direct comparisons
1285 between observed and simulated contrail evolution will help assess model
1286 accuracy and potentially inform improvements in contrail forecasting and
1287 climate ~~modelling~~modeling.

1288 Ideally, contrail detection, tracking, and attribution should be addressed
1289 by a single deep learning architecture capable of jointly processing video,
1290 flight trajectory data, and meteorological fields. ~~A model such as~~ For instance,
1291 ~~a variant of Mask2Former could be adapted for this purpose.~~ Integrating
1292 ~~these tasks into one architecture~~ Such an integrated approach would en-
1293 able end-to-end learning and exploit the complementary nature of the in-
1294 puts, as weather conditions and aircraft traffic data are highly informative
1295 for both detecting and tracking contrails. However, this integration is not
1296 straightforward. It requires careful design of input data representations to
1297 handle ~~spatiotemporal~~ spatio-temporal and multi-modal inputs, the creation
1298 of aligned and consistent annotations for all tasks, and the development of
1299 loss functions that balance competing objectives across detection, segmenta-
1300 tion, tracking, and attribution. Despite these challenges, we encourage the
1301 research community to explore this unified approach.

1302 Additionally, deploying multiple cameras in a spatially distributed network
1303 would enable stereographic height analysis: contrails observed simultaneously

from different viewing angles could be triangulated to determine altitude directly, rather than assuming a fixed height. This would provide crucial validation data for contrail formation models and improve flight attribution accuracy by eliminating altitude uncertainty.

More broadly, we hope this work encourages the development of similar ground-based contrail monitoring systems in other regions. A collaborative, open-science approach, sharing datasets, models, and observational infrastructure, will be essential to building a geographically diverse and temporally continuous picture of contrail ~~behaviour~~ behavior. We view this paper as a first step toward a ~~data-driven ecosystem collaborative, open-science framework~~ for contrail research: one that integrates physical ~~modelling~~ modeling with observational data through openly shared datasets and tools, spans spatial and temporal scales through multi-platform monitoring, and supports long-term efforts to better understand and reduce aviation's impact on ~~the climate~~ climate. By providing high-quality ground-based data alongside baseline computer vision models, we aim to facilitate model-data comparison, enable validation of physical models, and encourage the development of complementary monitoring systems worldwide.

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., Belouin, N., 2024. The importance of an informed choice of CO₂-equivalence metrics for contrail avoidance. Atmospheric Chemistry and Physics 24, 9401–9417.
- Bugliaro, L., Mannstein, H., Kox, S., 2012. Ice clouds properties from space, in: Schumann, U. (Ed.), Atmospheric Physics – Background, Methods, Trends. Springer, Heidelberg. doi:10.1007/978-3-642-30183-4_25.
- Campello, R.J.G.B., Moulavi, D., Sander, J., 2013. Density-based clustering based on hierarchical density estimates, in: Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), Springer. pp. 160–172. doi:10.1007/978-3-642-37456-2_14.

- 1337 Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko,
1338 S., 2020. End-to-end object detection with transformers, in: European
1339 Conference on Computer Vision (ECCV), pp. 213–229.
- 1340 Chen, L.C., Papandreou, G., Schroff, F., Adam, H., 2017. Rethinking
1341 atrous convolution for semantic image segmentation. arXiv preprint
1342 arXiv:1706.05587 .
- 1343 Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-
1344 decoder with atrous separable convolution for semantic image segmenta-
1345 tion, in: Proceedings of the European Conference on Computer Vision
1346 (ECCV), pp. 801–818.
- 1347 Cheng, B., Choudhuri, A., Misra, I., Kirillov, A., Girdhar, R., Schwing,
1348 A.G., 2021a. Mask2former for video instance segmentation. arXiv preprint
1349 arXiv:2112.10764 Extends Mask2Former to video by processing 3D seg-
1350 mentation volumes.
- 1351 Cheng, B., Misra, I., Schwing, A.G., Kirillov, A., Girdhar, R., 2022. Masked-
1352 attention mask transformer for universal image segmentation, in: Pro-
1353 ceedings of the IEEE/CVF Conference on Computer Vision and Pattern
1354 Recognition (CVPR).
- 1355 Cheng, B., Schwing, A.G., Kirillov, A., 2021b. Per-pixel classification is not
1356 all you need for semantic segmentation, in: Advances in Neural Information
1357 Processing Systems (NeurIPS).
- 1358 Chevallier, R., Shapiro, M., Engberg, Z., Soler, M., Delahaye, D., 2023.
1359 Linear contrails detection, tracking and matching with aircraft using geo-
1360 stationary satellite and air traffic data. Aerospace 10, 578.
- 1361 Dai, Z., Liu, H., Le, Q.V., Tan, M., 2021. CoAtNet: Marrying convolution
1362 and attention for all data sizes. Advances in Neural Information Processing
1363 Systems 34, 3965–3977.
- 1364 De Brabandere, B., Neven, D., Van Gool, L., 2017. Semantic instance seg-
1365 mentation with a discriminative loss function. CoRR abs/1708.02551.
1366 URL: <http://arxiv.org/abs/1708.02551>, arXiv:1708.02551.

1367 Duda, D.P., Minnis, P., Khlopenkov, K., Chee, T.L., Boeke, R., 2013. Esti-
1368 mation of 2006 northern hemisphere contrail coverage using MODIS data.
1369 Geophysical Research Letters 40, 612–617.

1370 Ewald, F., Bugliaro, L., Mannstein, H., Mayer, B., 2013. An improved
1371 cirrus detection algorithm MeCiDA2 for SEVIRI and its evaluation with
1372 MODIS. Atmospheric Measurement Techniques 6, 309–322. doi:10.5194/
1373 amt-6-309-2013.

1374 Forkert, T., Strauss, B., Wendling, P., 1993. A new algorithm for the au-
1375 tomated detection of jet contrails from NOAA-AVHRR satellite images,
1376 in: Proc. of the 6th AVHRR Data Users’ Meeting, EUMETSAT-Joint
1377 Research Centre of the Commission of the EC. pp. 513–519.

1378 Fritz, T.M., Eastham, S.D., Speth, R.L., Barrett, S.R., 2020. The role of
1379 plume-scale processes in long-term impacts of aircraft emissions. Atmo-
1380 spheric Chemistry and Physics 20, 5697–5727.

1381 Geraedts, S., Brand, E., Dean, T.R., Eastham, S., Elkin, C., Engberg, Z.,
1382 Hager, U., Langmore, I., McCloskey, K., Ng, J.Y.H., et al., 2024. A scalable
1383 system to measure contrail formation on a per-flight basis. Environmental
1384 Research Communications 6, 015008.

1385 Gierens, K., Matthes, S., Rohs, S., 2020. How well can persistent contrails
1386 be predicted? Aerospace 7, 169. doi:10.3390/aerospace7120169.

1387 Gourgue, N., Boucher, O., Barthès, L., 2025. A dataset of annotated ground-
1388 based images for the development of contrail detection algorithms. Data
1389 in Brief 59, 111364.

1390 Hamann, U., Walther, A., Baum, B., Bennartz, R., Bugliaro, L., Derrien, M.,
1391 Francis, P., Heidinger, A., Joro, S., Kniffka, A., Le Gléau, H., Lockhoff,
1392 M., Lutz, H.J., Meirink, J.F., Minnis, P., Palikonda, R., Roebeling, R.,
1393 Thoss, A., Platnick, S., Watts, P., Wind, G., 2014. Remote sensing of
1394 cloud top pressure/height from SEVIRI: analysis of ten current retrieval
1395 algorithms. Atmospheric Measurement Techniques 7, 2839–2867. doi:10.
1396 5194/amt-7-2839-2014.

1397 He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN, in:
1398 Proceedings of the IEEE International Conference on Computer Vision
1399 (ICCV), pp. 2961–2969.

1400 Iwabuchi, H., Yang, P., Liou, K.N., Minnis, P., 2012. Physical and optical
1401 properties of persistent contrails: Climatology and interpretation. *Journal of Geophysical Research: Atmospheres* 117, D06215. doi:10.1029/
1402 2011JD017020.
1403

1404 Jarry, G., Very, P., Ballerini, F., Dalmau, R., 2025. GVCCS: Ground visible
1405 camera contrail sequences. URL: [https://doi.org/10.5281/zenodo.](https://doi.org/10.5281/zenodo.16419651)
1406 16419651, doi:10.5281/zenodo.16419651.

1407 Jarry, G., Very, P., Heffar, A., Torjman-Levavasseur, V., 2024. Deep semantic
1408 contrails segmentation of GOES-16 satellite images: A hyperparameter
1409 exploration .

1410 Kox, S., Bugliaro, L., Ostler, A., 2014. Retrieval of cirrus cloud optical
1411 thickness and top altitude from geostationary remote sensing. *Atmospheric
1412 Measurement Techniques* 7, 3233–3246. doi:10.5194/amt-7-3233-2014.

1413 Kuhn, H.W., 1955. The Hungarian method for the assignment problem.
1414 *Naval Research Logistics Quarterly* 2, 83–97.

1415 Kulik, L., 2019. Satellite-based detection of contrails using deep learning.
1416 Ph.D. thesis. Massachusetts Institute of Technology.

1417 Lee, D.S., Fahey, D.W., Skowron, A., Allen, M.R., Burkhardt, U., Chen, Q.,
1418 Doherty, S.J., Freeman, S., Forster, P.M., Fuglestad, J., et al., 2021. The
1419 contribution of global aviation to anthropogenic climate forcing for 2000
1420 to 2018. *Atmospheric Environment* 244, 117834.

1421 Lin, T.Y., Maire, M., Belongie, S.J., Bourdev, L.D., Girshick, R.B., Hays, J.,
1422 Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft COCO:
1423 Common objects in context. *CoRR abs/1405.0312*. arXiv:1405.0312.

1424 Low, J., Teoh, R., Ponsonby, J., Gryspeerdt, E., Shapiro, M., Stettler, M.E.,
1425 2025. Ground-based contrail observations: comparisons with reanalysis
1426 weather data and contrail model simulations. *Atmospheric Measurement
1427 Techniques* 18, 37–56.

1428 Mannstein, H., Brömser, A., Bugliaro, L., 2010. Ground-based observa-
1429 tions for the validation of contrails and cirrus detection in satellite im-
1430 agery. *Atmospheric Measurement Techniques* 3, 655–669. doi:10.5194/
1431 amt-3-655-2010.

- 1432 Mannstein, H., Meyer, R., Wendling, P., 1999. Operational detection of con-
1433 trails from NOAA-AVHRR data. *International Journal of Remote Sensing*
1434 20, 1641–1660.
- 1435 Mannstein, H., Schumann, U., 2005. Aircraft induced contrail cirrus over Eu-
1436 rope. *Meteorologische Zeitschrift* 14, 549–554. doi:10.1127/0941-2948/
1437 2005/0058.
- 1438 Mannstein, H., Vázquez-Navarro, M., Graf, K., Duda, D.P., Schumann, U.,
1439 2012. Contrail detection in satellite images, in: Schumann, U. (Ed.), *At-
1440 mospheric Physics – Background, Methods, Trends*. Springer, Berlin, Hei-
1441 delberg. doi:10.1007/978-3-642-30183-4_26.
- 1442 McCloskey, K.J.F., Geraedts, S.D., Jackman, B.H., Meijer, V.R., Brand,
1443 E.W., Fork, D.K., Platt, J.C., Elkin, C., Van Arsdale, C.H., 2021. A
1444 human-labeled Landsat-8 contrails dataset, in: *Proceedings of the ICML
1445 Workshop on Tackling Climate Change with Machine Learning*.
- 1446 Meijer, V.R., Eastham, S.D., Waitz, I.A., Barrett, S.R., 2024. Contrail alti-
1447 tude estimation using GOES-16 ABI data and deep learning. *Atmospheric
1448 Measurement Techniques* 17, 6145–6162.
- 1449 Meijer, V.R., Kulik, L., Eastham, S.D., Allroggen, F., Speth, R.L., Karaman,
1450 S., Barrett, S.R., 2022. Contrail coverage over the united states before
1451 and during the COVID-19 pandemic. *Environmental Research Letters* 17,
1452 034039.
- 1453 Meyer, R., Buell, R., Leiter, C., Mannstein, H., Pechtl, S., Oki, T.,
1454 Wendling, P., 2007. Contrail observations over Southern and Eastern
1455 Asia in NOAA/AVHRR data and comparisons to contrail simulations
1456 in a GCM. *International Journal of Remote Sensing* 28, 2049–2069.
1457 doi:10.1080/01431160600641707.
- 1458 Meyer, R., Mannstein, H., Meerkötter, R., Schumann, U., Wendling, P., 2002.
1459 Regional radiative forcing by line-shaped contrails derived from satellite
1460 data. *Journal of Geophysical Research: Atmospheres* 107, ACL 17–1–ACL
1461 17–15. doi:10.1029/2001JD000426.
- 1462 Minnis, P., Palikonda, R., Walter, B.J., Ayers, J.K., Mannstein, H., 2005.
1463 Contrail properties over the eastern North Pacific from AVHRR data. *Me-
1464 teorologische Zeitschrift* 14, 515–523. doi:10.1127/0941-2948/2005/0056.

- 1465 Ng, J.Y.H., McCloskey, K., Cui, J., Meijer, V.R., Brand, E., Sarna,
1466 A., Goyal, N., Van Arsdale, C., Geraedts, S., 2023. OpenContrails:
1467 Benchmarking contrail detection on GOES-16 ABI. arXiv preprint
1468 arXiv:2304.02122 .
- 1469 Northcutt, C., Jiang, L., Chuang, I., 2021. Confident learning: Estimating
1470 uncertainty in dataset labels. *Journal of Artificial Intelligence Research*
1471 70, 1373–1411. doi:10.1613/jair.1.12125.
- 1472 Ortiz, I., García-Heras, J., Jafarimoghaddam, A., Soler, M., 2025. Enhancing
1473 GOES-16 contrail segmentation through ensemble neural network model-
1474 ing and optical flow corrections. *Authorea Preprints* .
- 1475 Palikonda, R., Minnis, P., Duda, D.P., Mannstein, H., 2005. Contrail
1476 coverage derived from 2001 AVHRR data over the continental United
1477 States and surrounding areas. *Meteorologische Zeitschrift* 14, 525–536.
1478 doi:10.1127/0941-2948/2005/0051.
- 1479 Pertino, P., Pavarino, L., Lomolino, S., Miotto, E., Cambrin, D.R., Garza,
1480 P., Ogliari, E.G.C., 2024. Ground-based contrail detection by means of
1481 computer vision models: A comparison between visible and infrared im-
1482 ages. 2024 IEEE 8th Forum on Research and Technologies for Society and
1483 Industry Innovation (RTSI) , 254–259.
- 1484 Pratt, W.K., 2007. Digital image processing: PIKS Scientific inside. vol-
1485 ume 4. Wiley Online Library.
- 1486 Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look
1487 once: Unified, real-time object detection, in: *Proceedings of the IEEE*
1488 *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.
1489 779–788. doi:10.1109/CVPR.2016.91.
- 1490 Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster R-CNN: Towards real-
1491 time object detection with region proposal networks. *IEEE Transactions*
1492 *on Pattern Analysis and Machine Intelligence* 39, 1137–1149. doi:10.1109/
1493 TPAMI.2016.2577031.
- 1494 Ridnik, T., Ben Baruch, E., Noy, A., Zelnik-Manor, L., 2021. ImageNet-21K
1495 pretraining for the masses. *CoRR* abs/2104.10972. arXiv:2104.10972.

- 1496 Riggi-Carolo, E., Dubot, T., Sarrat, C., Bedouet, J., 2023. AI-driven identi-
1497 fication of contrail sources: Integrating satellite observation and air traffic
1498 data. *Journal of Open Aviation Science* 1.
- 1499 Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional networks
1500 for biomedical image segmentation, in: *International Conference on Med-
1501 ical Image Computing and Computer-Assisted Intervention*, Springer. pp.
1502 234–241.
- 1503 Sarna, A., Meijer, V., Chevallier, R., Duncan, A., McConnaughay, K., Ger-
1504 aedts, S., McCloskey, K., 2025. Benchmarking and improving algorithms
1505 for attributing satellite-observed contrails to flights. *EGUsphere* 2025, 1–
1506 58.
- 1507 Schmidt, E., 1941. Die Entstehung von Eisnebel aus den Auspuffgasen
1508 von Flugmotoren, in: *Schriften der Deutschen Akademie der Luftfahrt-
1509 forschung*. Verlag R. Oldenbourg, München. volume 44, pp. 1–15. URL:
1510 <http://elib.dlr.de/107948/>.
- 1511 Schumann, U., 1996. On conditions for contrail formation from aircraft ex-
1512 hausts. *Meteorologische Zeitschrift* 5, 4–23. doi:10.1127/metz/5/1996/4.
- 1513 Schumann, U., 2012. A contrail cirrus prediction model. *Geoscientific
1514 Model Development* 5, 543–580. URL: [https://gmd.copernicus.org/
1515 articles/5/543/2012/](https://gmd.copernicus.org/articles/5/543/2012/), doi:10.5194/gmd-5-543-2012.
- 1516 Schumann, U., Hempel, R., Flentje, H., Garhammer, M., Graf, K., Kox, S.,
1517 Lösslein, H., Mayer, B., 2013. Contrail study with ground-based cameras.
1518 *Atmospheric Measurement Techniques* 6, 3597–3612.
- 1519 Schumann, U., Mayer, B., Graf, K., Mannstein, H., Meerkötter, R., 2009. A
1520 parametric radiative forcing model for cirrus and contrail cirrus, in: *ESA
1521 Atmospheric Science Conference*, ESA SP-676, Barcelona, Spain. pp. 1–6.
- 1522 Strandgren, J., Bugliaro, L., Sehnke, F., Schröder, L., 2017a. Cir-
1523 rus cloud retrieval with MSG/SEVIRI using artificial neural networks.
1524 *Atmospheric Measurement Techniques* 10, 3547–3573. doi:10.5194/
1525 amt-10-3547-2017.

- 1526 Strandgren, J., Fricker, J., Bugliaro, L., 2017b. Characterisation of the ar-
 1527 tificial neural network CiPS for cirrus cloud remote sensing with MSG/-
 1528 SEVIRI. *Atmospheric Measurement Techniques* 10, 4317–4339. doi:10.
 1529 5194/amt-10-4317-2017.
- 1530 Sun, J., Roosenbrand, E., 2025. Few-shot contrail segmentation in remote
 1531 sensing imagery with loss function in hough space. *IEEE Journal of Se-
 1532 lected Topics in Applied Earth Observations and Remote Sensing* .
- 1533 Teoh, R., Engberg, Z., Schumann, U., Voigt, C., Shapiro, M., Rohs, S.,
 1534 Stettler, M., 2023. Global aviation contrail climate effects from 2019 to
 1535 2021. *EGUsphere* 2023, 1–32.
- 1536 Teoh, R., Schumann, U., Stettler, M.E., 2020. Beyond contrail avoidance:
 1537 Efficacy of flight altitude changes to minimise contrail climate forcing.
 1538 *Aerospace* 7, 121.
- 1539 Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A., Li, Y.,
 1540 2022. MaxViT: Multi-axis vision transformer, in: *European Conference
 1541 on Computer Vision*, Springer. pp. 459–479.
- 1542 Van Huffel, J., Ehrmanntraut, R., Croes, D., 2025. Contrail detection
 1543 and classification using computer vision with ground-based cameras, in:
 1544 2025 Integrated Communications, Navigation and Surveillance Conference
 1545 (ICNS), IEEE. pp. 1–6.
- 1546 Vázquez-Navarro, M., Mannstein, H., Kox, S., 2015. Contrail life cycle and
 1547 properties from 1 year of MSG/SEVIRI rapid-scan images. *Atmospheric
 1548 Chemistry and Physics* 15, 8739–8749. doi:10.5194/acp-15-8739-2015.
- 1549 Vázquez-Navarro, M., Mannstein, H., Mayer, B., 2010. An automatic contrail
 1550 tracking algorithm. *Atmospheric Measurement Techniques* 3, 1089–1101.
 1551 doi:10.5194/amt-3-1089-2010.
- 1552 Yang, L., Fan, Y., Xu, N., 2019. Video instance segmentation, in: *ICCV*.

1553 **Appendix A. Consistent Instance Tracking Algorithm**

1554 Due to memory limitations, the video segmentation model operates on
 1555 short temporal clips of fixed length N frames, using a sliding window of

stride 1. While instance segmentation within each clip is temporally consistent (i.e., instance identifiers are maintained across frames within the clip), the model processes each clip independently. As a result, instance identifiers are not necessarily consistent across clips.

To enforce globally consistent instance identifiers across the full video sequence, we implement a deterministic post-processing method that aligns instance predictions across overlapping clips. The method uses mask overlap similarity, specifically, IoU, across shared frames and performs optimal bipartite matching using the Hungarian algorithm. Below, we provide a rigorous description of the method.

For a given frame index $t \in \{N, N+1, \dots, T\}$, we define:

- The **current clip** as the sequence $F_{t-N+1}, F_{t-N+2}, \dots, F_t, F_{t-N+1}, F_{t-N+2}, \dots, F_t$.
- The **previous clip** as the sequence $F_{t-N}, F_{t-N+1}, \dots, F_{t-1}, F_{t-N}, F_{t-N+1}, \dots, F_{t-1}$.

The two clips overlap in $N-1$ frames: $F_{t-N+1}, \dots, F_{t-1}$ frames: $F_{t-N+1}, \dots, F_{t-1}$. Only frame F_t is newly introduced in the current clip. At each step, we seek to propagate consistent instance identifiers by matching instances across the overlapping frames. Let:

- $\mathcal{I}_{\text{prev}} = \{1, \dots, K\}$, $\mathcal{I}_{\text{prev}} = \{1, \dots, K\}$: instance identifiers in the previous clip.
- $\mathcal{I}_{\text{curr}} = \{1, \dots, M\}$, $\mathcal{I}_{\text{curr}} = \{1, \dots, M\}$: instance identifiers in the current clip.

We define a cost matrix $C \in \mathbb{R}^{M \times K}$, $C \in \mathbb{R}^{M \times K}$, where each element C_{ij} encodes the negative temporal IoU between instance $i \in \mathcal{I}_{\text{curr}}$ and instance $j \in \mathcal{I}_{\text{prev}}$ over the overlapping frames:

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

where $\mathcal{M}_{i,f}^{\text{curr}}$ and $\mathcal{M}_{j,f}^{\text{prev}}$ denote the binary masks of instances i and j at frame f and j at frame f , respectively. If an instance does not appear in a given frame (e.g., missing mask), its contribution is treated as zero overlap.

1585 To eliminate unlikely or noisy matches, we apply a threshold $\tau \in [0, 1]$
 1586 $\tau \in [0, 1]$ on the mean IoU:

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

1587 where the threshold τ is selected empirically to balance precision and ro-
 1588 bustness; we recommend $\tau = 0.1$.

1589 We remove rows and columns of the cost matrix that contain only $+\infty$
 1590 $+\infty$ entries. Using the modified cost matrix, we solve the bipartite assign-
 1591 ment problem via the Hungarian algorithm (Kuhn, 1955), an optimization
 1592 method that finds the optimal one-to-one matching minimizing total cost—obtaining
 1593 a one-to-one (or partial) mapping between current and previous instances.
 1594 Let $\sigma : \mathcal{I}_{\text{curr}} \rightarrow \mathcal{I}_{\text{prev}} \cup \{\emptyset\}$ $\sigma : \mathcal{I}_{\text{curr}} \rightarrow \mathcal{I}_{\text{prev}} \cup \{\emptyset\}$ denote the resulting as-
 1595 signment. We then update the instance identifiers in the current clip to
 1596 match those of the assigned instances in the previous clip. Unmatched in-
 1597 stances are assigned new unique identifiers. The pseudo-code of the algorithm
 1598 is presented in Algorithm 1.

Algorithm 1 Post-processing for Consistent Instance Tracking

Require: Predicted instance masks for video frames F_1, \dots, F_T , ~~threshold~~
 ~~τ~~ F_1, \dots, F_T , threshold τ

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

1599 This process is applied sequentially from frame ~~$t = N$ to T~~ $t = N$ to T ,
1600 ensuring that instance identifiers are globally consistent across the video.