

# Spatiotemporal Analysis of Contrail Formations & Flight Attribution

Achilleas Leivadiotis

*Department of Advanced Computing Sciences*

*Faculty of Science and Engineering*

*Maastricht University*

Maastricht, The Netherlands

Email: a.leivadiotis@student.maastrichtuniversity.nl

**Abstract**—Aviation’s climate impact is significantly amplified by non-CO<sub>2</sub> effects, primarily the formation of contrail-cirrus clouds. Although ground-based cameras offer continuous high-resolution monitoring capabilities that overcome the limitations of satellite imagery, existing methods have struggled to reliably attribute contrails to specific flights. This thesis presents an end-to-end pipeline for automated detection, tracking, and a promising novel approach for flight attribution of contrails from ground-based camera imagery. The proposed system integrates a Detectron2 Mask R-CNN model for high-precision instance segmentation, a multi-object tracking algorithm to maintain contrail identities across frames, and a spatiotemporal correlation module to link contrail tracks to multi-radar correlated historical flight data. Through extensive experimentation, a Mask R-CNN model configured for this task surpasses prior work in segmentation accuracy, achieving a mask mAP of up to 64.8826% (at IoU=0.50). A comparative analysis shows that the Norfair tracker (MOTA: 32.5, IDF1: 59.2) outperforms DeepSORT in maintaining track identity despite 15-second inter-frame gaps. The complete pipeline successfully attributes 46.2% of short-lived contrail tracks to unique flight IDs, demonstrating a novel method for transforming raw sky video into a flight-resolved contrail dataset. This work provides a validated prototype that enables detailed climate-impact assessments and supports the development of contrail mitigation strategies.

**Index Terms**—contrails, computer vision, Mask R-CNN, Detectron2, DeepSORT, climate impact, MOT, instance segmentation

## I. INTRODUCTION

Aviation warms the climate not only through CO<sub>2</sub> but also through non-CO<sub>2</sub> effects, most prominently the radiative forcing of contrail-cirrus clouds, whose global impact is now estimated to rival or even exceed the cumulative warming from aviation CO<sub>2</sub> itself [1]–[3].

Contrails form when hot, moist exhaust gases rich in water vapor and soot mix with cold, ice-supersaturated air in the upper troposphere (typically 8–12 km altitude). As the mixture cools, excess water vapor condenses on soot nuclei and freezes into ice crystals;

The increasing recognition of these contrail-cirrus clouds as significant contributors to aviation’s overall climate impact underscores the urgent need for effective monitoring and

mitigation strategies. Addressing this imperative, this thesis focuses on enhancing ground-based observation capabilities. This approach is particularly crucial as it facilitates the direct and near real-time detection of thin and short-lived contrails—precisely the types that often elude satellite-based systems due to their limited temporal revisit rates and spatial resolution. Moreover, the high-fidelity data generated by such ground-based systems serves as a vital resource for the validation and improvement of contrail prediction models, which are essential for the development and implementation of effective contrail mitigation strategies.

Satellite images furnish global coverage, yet young or optically thin contrails are frequently missed because of spatial resolution and masking by other cloud layers [4]–[6]. Networks of ground-based cameras overcome these limitations by providing continuous imagery that captures contrail formation within seconds of exhaust release and tracks subsequent growth in real time [7], [8].

Early ground-camera studies by Croes [9] and Van Huffel *et al.* [10] demonstrated single-frame contrail detection and ad-hoc flight matching. Despite encouraging results, existing ground-camera methods still treat contrails as frame-isolated detections. Two open challenges therefore remain:

- 1) **Temporal identities:** Robust tracking of individual contrails over time, even with intermittent visibility or complex interactions.
- 2) **Robust linkage to parent aircraft:** A reliable method to associate tracked contrails with the specific flight IDs of the aircraft that generated them.

Addressing both simultaneously is essential for transforming inexpensive sky video data into flight-resolved contrail datasets, which are crucial for detailed climate analysis and the validation of contrail avoidance measures. This thesis addresses this shortcoming by introducing a complete *detection* → *tracking* → *spatiotemporal attribution* pipeline that first builds continuous contrail tracks and only then links each track to the originating flight.

This thesis seeks to address the aforementioned problem statement by investigating the following research questions:

- 1) How do different configurations of the Detectron2 Mask R-CNN model affect the segmentation accuracy of con-

This thesis was prepared in fulfilment of the requirements for the Degree of Bachelor of Science in Data Science and Artificial Intelligence, Maastricht University. Supervisor(s): Menica Dibenedetto and Yusuf Can Semerci.

trails across their four primary morphology classes (thin, wide, cirrus, short)?

- 2) Given a chosen Detectron2-based contrail detector, which multi-object tracking algorithm specifically yields superior performance in terms of temporal coherence?
- 3) Using a spatiotemporal association logic based on projected flight positions and contrail tracking pipeline, what fraction of persistently tracked contrails can be successfully and uniquely matched to a flight ID?

The scope of this thesis encompasses the design, implementation and evaluation of an integrated pipeline for contrail detection, tracking, and flight attribution using ground-based camera imagery and flight surveillance data provided by the EUROCONTROL Maastricht Upper Area Control Center (MUAC). The pipeline begins with detection through instance segmentation in the Detectron2 framework, where a Mask R-CNN model that combines a ResNet-101 backbone with a Feature Pyramid Network (FPN) identifies candidate contrails. Various training setups and data augmentation strategies are explored to refine this stage. The resulting detections are passed to several established multi-object tracking algorithms that are judged with the standard metrics used in that field. An attribution module then pairs each tracked plume with its likely flight by comparing their spatial position and timing. Every step is scored quantitatively through mean average precision and recall for detection, multi object tracking accuracy and the identity F1 score for tracking, and the percentage of successful matches for attribution, complemented by a careful visual review.

Several practical limitations inevitably shape the conclusions. The camera records a new frame only every fifteen seconds, which complicates the following of very dynamic or short lived contrails. Projecting flight paths into the images is also imperfect because the radar-derived track timestamps are not always perfectly synchronised with the camera, and the camera calibration can carry small errors in both its internal parameters and its orientation. The attribution logic is tailored for fresh contrails that still sit close to the aircraft trajectory and therefore does not address older plumes that drift with the wind and would require a full atmospheric model. In addition, all experiments use the MUAC dataset with its human-annotated imagery, which due to the subjectivity of the task introduces errors in the data. Furthermore, the trained model may need adaptation before it performs well in other regions or with different camera systems.

Several important topics remain outside the remit of this work. The study does not propose entirely new neural architectures but focuses on configuring, training, and analysing current state of the art models. Questions related to real time deployment, fault tolerance, and large scale engineering are left for future work. Likewise, the physical processes that create contrails and their broader climate effects are not examined in depth here, although the data products that emerge from the pipeline are intended to support future atmospheric research.

## A. Thesis Outline

Chapter 2 sets out the *Theoretical Background*, introducing the physics of contrail formation and lifecycle taxonomy. Chapter 3 surveys the *Related Work*, covering satellite- and ground-based detection studies, recent deep-learning approaches to contrail segmentation, and current attempts at tracking and flight attribution. Chapter 4 presents the *Data and Methodology*, detailing the MUAC ground-camera imagery and the multi-radar-correlated flight-track dataset, preprocessing steps, the Detectron2 Mask R-CNN configuration, the Norfair and DeepSORT trackers, the spatiotemporal attribution logic, and the evaluation metrics used throughout the pipeline. Chapter 5 describes the *Experimental Implementation*, including the hyper-parameter sweeps, ablation studies and software-hardware environment for training and testing each module. Chapter 6 reports and discusses the *Results*, providing quantitative and qualitative evaluations of detection, tracking and attribution performance, together with an analysis of practical limitations such as the 15 s inter-frame interval. Finally, Chapter 7 concludes the thesis, revisits the research questions, assesses how far the problem statement has been addressed, and outlines avenues for future work and potential system enhancements.

## II. THEORETICAL BACKGROUND

A *condensation trail (contrail)* is an artificial cloud of ice crystals that appears behind an aircraft when exhaust gases rich in water vapour and soot meet air that is cold ( $< -40^{\circ}\text{C}$ ) and humid enough to become supersaturated with respect to ice. Under these conditions, the plume crosses the Schmidt-Appleman threshold, the vapor condenses onto the soot particles, it freezes almost instantly, and a line-shaped cloud becomes visible [11]. If the ambient air remains ice-supersaturated, the trail persists and spreads under wind shear, ageing into diffuse *contrail-cirrus*. These optically thin but widespread clouds trap outgoing infrared radiation more effectively than they reflect sunlight, generating a net positive radiative forcing comparable to aviation  $\text{CO}_2$ .

*Lifecycle classes (operational shorthand) are the following:*

- 1) **Short / Evanescent** – dissipate within seconds in dry air.
- 2) **Thin / Persistent-narrow** – survive for minutes while remaining filament-like.
- 3) **Wide / Persistent-spreading** – broaden into kilometre-scale filaments.
- 4) **Contrail-cirrus** – shear, diffusion and mixing create extensive cirrus-like sheets.

[1], [3], [12].

For visual representation, figure 1 has examples for all the different life cycles of contrail formations. These examples showcase the hard nature of the detection task. Accurately identifying and segmenting contrails is challenging due to several factors: images often contain the sun, which can cause glare and obscure details; other natural cloud formations can

be visually similar to contrails, leading to potential misclassification; and contrails themselves exhibit wide variations in morphology, optical thickness, and persistence, making contrail instances particularly difficult to distinguish against a dynamic sky background.



Fig. 1: Examples of the four lifecycle classes ( Top-left: Short, Top-right: Thin, Bottom-left: Wide, Bottom-right: Contrail-cirrus).

### III. RELATED WORK

Research on contrails spans three complementary areas: (i) remote-sensing detection, (ii) object segmentation and classification, and (iii) multi-frame tracking with flight-level attribution. The review below highlights key contributions in each domain and the remaining gaps that motivate this thesis.

#### A. Satellite-Based Detection

Mannstein *et al.* used a split-window threshold on the NOAA-AVHRR thermal-IR channels (10.8 and 12.0  $\mu\text{m}$ ) to detect linear contrails at the sensor’s native  $\sim 1.1$  km nadir resolution, achieving near-global coverage but noting that thin or short-lived trails often fell below AVHRR’s detectability limit [4]. Follow-on gridded products such as GridSat-GOES remap the GOES Imager IR data (native  $\sim 4$  km at sub-satellite point) onto a  $0.07^\circ$  ( $\sim 7\text{--}8$  km) grid, yielding a much finer temporal cadence ( $\approx 15$  min) at the cost of coarser spatial detail; Knapp and Wilkins therefore caution that sub-pixel plumes and features forming between successive sector scans can still be missed [6]. These limitations underline the need for alternative observing systems capable of capturing contrail formation within seconds of exhaust release.

#### B. Ground-Camera Detection and Segmentation

Ground-based cameras overcome the temporal and spatial limitations of satellites by providing continuous, meter-scale imagery. Schumann *et al.* demonstrated photogrammetric reconstruction of contrail height, width and growth rate from stereoscopic camera pairs [7]. More recently, Croes applied U-Net to single frames for near-real-time contrail alerts in an operational setting [9], while Van Huffel *et al.* compared YOLOv5 and Mask R-CNN detectors, reporting a bounding-box mAP<sub>0.50</sub> of 60.9 on the MUAC data set [10], [13]. Low *et*

*al.* extended ground-camera monitoring to a 14-hour sequence and manually matched a subset of tracks to flight telemetry, illustrating the potential for long-duration observation [8].

#### C. Deep-Learning Architectures for Instance Segmentation

Accurate radiative-forcing estimates require pixel-level outlines rather than coarse bounding boxes. Mask R-CNN [13], as implemented in Detectron2 [14], has become the *de-facto* standard for instance segmentation. Van Huffel *et al.* trained a Mask R-CNN on MUAC imagery and reported a mask mAP<sub>0.50:0.95</sub> of 21.6 [10]. Optimising segmentation networks specifically for the thin, elongated morphology of contrails therefore remains an open problem.

#### D. Tracking and Flight Attribution

Linking detections across frames is complicated by the non-rigid, wind-driven evolution of contrails and by the long inter-frame intervals typical of sky-camera deployments. To date, generic multi-object trackers such as DeepSORT [15] have *not yet been systematically bench-marked* on contrail imagery. On satellite data, Chevallier *et al.* combined Mask R-CNN detections from the GOES-16 Imager with ADS-B telemetry to associate persistent kilometre-scale contrails with aircraft, but they reported only heuristic quality measures and did not compute CLEAR-MOT scores [5]. For ground-camera imagery, Croes developed a lightweight ellipse-matching scheme that propagates contrail masks across consecutive frames and then assigns each track to the closest aircraft in image space; the approach was validated qualitatively but without MOT metrics or ablation studies. [9] Low *et al.* (2024) followed a 14-h camera record and matched tracks to radar data by visual inspection, again foregoing an automated MOT evaluation. As a result, a quantitative, end-to-end benchmark of contrail *tracking & flight attribution* on high-resolution sky imagery and a head-to-head comparison with standard trackers such as DeepSORT or Norfair, remains an open research need. [8]

#### E. Outstanding Challenges

The review identifies three unresolved issues: (1) segmentation models tailored to contrail morphology; (2) robust multi-object tracking across long ( $\gtrsim 10$  s) frame gaps; and (3) automated, unambiguous linkage of ground-camera contrails to individual flights in dense airspace. The present thesis addresses these gaps by integrating a tuned Mask R-CNN segmenter, Norfair-based tracking, and a spatiotemporal flight attribution module into a single, documented pipeline.

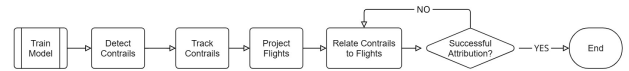


Fig. 2: Proposed end-to-end ground-camera contrail-monitoring pipeline: instance segmentation, multi-object tracking and spatiotemporal flight attribution.

### IV. DATA AND METHODS

The aforementioned flowchart in figure 2 is explained in this section

### A. Data

Image data were collected beforehand with ground-based cameras positioned throughout MUAC airspace under a wide range of weather and lighting conditions and 4 wind directions. The resulting dataset contains 5,016 annotated images with 6,447 labelled contrails. After Detectron2 automatically discarded frames without annotations, 2,907 images remained for training and 957 for testing. For the flight-attribution stage Eurocontrol’s *multi-radar correlated track* archive has been utilized, which fuses returns from several primary and secondary surveillance radars into time-stamped, three-dimensional aircraft trajectories (latitude, longitude, altitude) together with the corresponding flight identifiers.

### B. Contrail Detection and Instance Segmentation

To identify and precisely delineate contrails within the imagery, an instance segmentation approach is employed. This study utilizes the Mask R-CNN model, a prominent two-stage framework, implemented within Detectron2 [14]. The chosen backbone for the Mask R-CNN is a ResNet-101 with a Feature Pyramid Network (FPN), which aids in detecting objects at different scales. Mask R-CNN extends Faster R-CNN by adding a parallel branch for predicting segmentation masks on each Region of Interest (RoI), providing pixel-level accuracy for each detected contrail instance. This is particularly beneficial for capturing the often thin and elongated morphology of contrails. [13]

### C. Contrail Tracking

Once contrails are detected in individual frames, a multi-object tracking (MOT) algorithm is applied to maintain their identities across sequential frames, creating trajectories. This research evaluated two tracking algorithms: DeepSORT [15], which incorporates appearance information for robust tracking, and Norfair, a lightweight, centroid-based tracker. While both were considered, Norfair was ultimately selected for the final pipeline, particularly due to its effective performance in the context of flight attribution as detailed in the results (Section Results & Discussion). Norfair uses a Kalman filter to predict object motion and associates detections based on Euclidean distance to these predictions. For this application, a custom Kalman filter factory is implemented to handle variable time intervals (dt) between frames, which is crucial given the approximate 15-second image capture interval. The tracker is configured with parameters such as distance threshold, initialization delay, and maximum age for tracks.

### D. Flight Attribution

The final stage of the pipeline is to attribute each tracked contrail to a specific flight. This process, involves several steps:

- 1) **Flight Data Acquisition and Filtering:** Historical *multi-radar correlated tracks* for the target date are extracted from the MUAC surveillance archive. Each track provides latitude, longitude, altitude (Mode C), callsign, a unique track ID, and five-second timestamps. Tracks are retained only if the aircraft is within the

nominal contrail-forming altitude band and inside a predefined temporal window around each image time stamp.

- 2) **Camera Calibration and Projection:** Camera intrinsic parameters, such as, focal length, principal point, image dimensions and extrinsic parameters, such as, GPS position, elevation, heading, tilt, and roll along with lens distortion coefficients, are used to establish a transformation model from the camera transform library [16]. The fidelity of this model is directly dependent on the precision of these calibration parameters and the temporal synchronization of the data sources. This model allows for the projection of 3-D flight positions from the multi-radar tracks into 2-D pixel coordinates on the camera images.
- 3) **Spatiotemporal Correlation:** For each image, active flights within a defined time window (e.g.,  $\pm 30$  seconds) and maximum distance (e.g., 100 km) from the camera are considered. Their 3D positions are projected onto the image plane.
- 4) **Track-to-Flight Association:** As the multi-object tracker tracks contrails, it provides evolving centroid positions and unique track IDs. The flight attribution logic then associates these contrail tracks with the projected flight positions. This study utilizes a movement-based orientation of the contrail. The leading corner of a contrail’s bounding box, determined by its direction of movement, is used as the primary point for association. This geometric approach is most effective for ‘Short’ and ‘Thin’ contrails, because these contrails have not yet drifted far from the generating aircraft’s flight path. If this leading point is within a specified pixel distance threshold of a projected flight’s 2D position, a potential match is registered.
- 5) **Majority Vote and Deduplication:** Over the duration of a contrail track, votes are accumulated for each flight ID that is consistently matched. A majority vote system then determines the most likely flight ID for each contrail track. To ensure unique attributions, if multiple contrail tracks are attributed to the same flight, the one with the strongest evidence meaning, highest vote count and longest consistent association is retained.

This systematic approach allows for robust linkage of observed contrails to their originating aircraft, moving beyond single-frame heuristics.

### E. Evaluation Metrics

The performance of the pipeline is assessed at each stage. For contrail detection and segmentation, standard metrics such as mean Average Precision (mAP) for bounding boxes ( $\text{mAP}_{\text{box}}$ ) and segmentation masks ( $\text{mAP}_{\text{mask}}$ ), and Intersection over Union (IoU) are used. For multi-object tracking, metrics include Multi-Object Tracking Accuracy (MOTA), IDF1 score (which balances identity precision and recall), track precision, recall, and the number of identity switches. Flight attribution success is primarily measured by the association rate: the



percentage of persistently tracked contrails that are uniquely matched to a flight ID.

## V. EXPERIMENTAL IMPLEMENTATION

This section details the hyper-parameter sweeps and ablation studies carried out for the three core modules of the pipeline: Detectron2 instance-segmentation, Norfair tracking (with DeepSORT as a comparative baseline), and contrail-to-flight attribution. All experiments were run on an NVIDIA H100 (16 GB) GPU; model selection relied on the validation set described in Section 3.

### A. Object-Detection Experimental Setup

**Baseline:** The chosen model is MS-COCO-pre-trained MASKRCNN-R101-FPN configuration in Detectron2 [14], changing only the input dimensions from  $800 \times 1333$  to  $720 \times 1280$  to match the ground-camera aspect ratio. The hyper-parameter search space explored in this study is summarised in Table I.

TABLE I: Object-Detection Hyper-parameter Sweep (Mask R-CNN)

Parameter	Values Tested
ROI-Align (box)	7, 14, 28
ROI-Align (mask)	14, 28
Epoch budget	5–40
LR policy	Linear, Cosine

**Data augmentation.** Augmentation transforms evaluated individually and in combination:

- Horizontal flip (*Detectron2 default*)
- Vertical flip, random rotation
- Brightness/contrast, hue-saturation shift
- Motion blur, Gaussian noise, coarse dropout

**PointRend variant.** A Mask R-CNN + PointRend head was trained under the same schedules; although edge crispness improved qualitatively, quantitative gains did not justify the extra inference cost, so the baseline Mask R-CNN is used in the production pipeline.

### B. Multi-Object Tracking Experimental Setup

Two trackers were benchmarked: DeepSORT (appearance-aware) and Norfair (centroid-based). The search ranges for the DeepSORT and Norfair tracker hyper-parameters are listed in Table II.

TABLE II: DeepSORT / Norfair Tracking Hyper-parameters

Parameter	Range
Appearance cos-sim.(DeepSORT)	0.8, 0.5, 0.3
IOU threshold (DeepSORT)	0.2–0.4
Distance threshold (NorFair)	30–90 px
Init frames (Both)	1–3
Max age (frames)(Both)	2–10

**Metrics.** All hyper-parameter configurations are assessed with the standard CLEAR MOT suite MOTA, IDF1, precision, recall, and identity-switch count, supplemented by qualitative video overlays.

### C. Flight-Attribution Experimental Setup

Projection of multi-radar-correlated flight trajectories into image space and association of each persistent contrail track to a candidate flight through a staged filter. The parameter grid explored for the flight-attribution stage is summarised in Table III.

TABLE III: Flight-Attribution Parameter Grid

Parameter	Range
Radius (pixels)	30–100
Observation window	3–15 f

The primary evaluation metric is the *association rate*, the fraction of persistent contrail tracks ( $\geq 3$  frames) that can be uniquely matched to a flight ID.

TABLE IV: Optimal Hyper-parameter Configuration

Parameter	Selected Value
<b>Mask R-CNN Detector</b>	
ROI-Align (box)	7
ROI-Align (mask)	14
Epoch budget	40
LR policy	Cosine
Data Augmentation	H-Flip,
<b>Norfair Tracker</b>	
Distance threshold (px)	75
Init frames	2
Max age (frames)	6
<b>Flight Attribution</b>	
Radius (pixels)	70
Observation window (f)	5

All optimal hyper-parameters are summarised in Table IV(Section VI) and are used for the results reported in the subsequent discussion.

## VI. RESULTS AND DISCUSSION

This section presents the performance evaluation of the integrated contrail monitoring pipeline, beginning with the object detection model, followed by the tracking and flight attribution results. The evaluation was conducted on a dedicated test set of 957 images, distinct from the 2097 images used for training.

### A. Object Detection and Segmentation Performance

The trained Mask R-CNN model demonstrated strong capability to identify and segment contrails. The quantitative performance, evaluated using standard COCO metrics for mAP and mAR, is summarized in Table V.

TABLE V: Detection & Segmentation Performance Comparison

Metric	Prior Work [10]	This Thesis
mAP <sub>box</sub> (IoU=.50:.95)	60.9	<b>40 – 47.3446</b>
mAP <sub>box</sub> (IoU=.50)	82.8	<b>74 – 81.6818</b>
mAR <sub>box</sub> (100 dets)	75	<b>55 – 59.2</b>
mAP <sub>mask</sub> (IoU=.50:.95)	21.6	<b>19 – 22.1740</b>
mAP <sub>mask</sub> (IoU=.50)	56.6	<b>60 – 64.8826</b>
mAR <sub>mask</sub> (100 dets)	28.5	<b>28 – 31.5</b>

1) *Discussion of Quantitative Metrics:* As shown in Table V, the model achieves a nuanced performance profile when compared to previous work. The bounding box detection metrics ( $mAP_{\text{box}}$ ) are slightly lower than the benchmark established by Van Huffel *et al.* [10]. However, this work successfully surpasses the prior results in all key instance segmentation metrics ( $mAP_{\text{mask}}$  and  $mAR_{\text{mask}}$ ).

This outcome is highly favorable for the goals of this thesis. While bounding boxes provide general localization, high-quality segmentation masks are critical for the physical characterization of contrails, such as analyzing their morphology, width, and optical depth evolution. The improvement in  $mAP_{\text{mask}}$  (IoU=.50) from 56.6% (prior work) to **60 – 64.8826%** (this thesis) indicates an enhancement in the model’s ability to delineate the precise, often complex, pixel-level boundaries of contrails. This suggests that the model is better at understanding the object’s true shape, which is a more challenging and valuable task than simply placing a rectangular box around it. Finally, the configurations mentioned in table IV is responsible for the highest results in table V

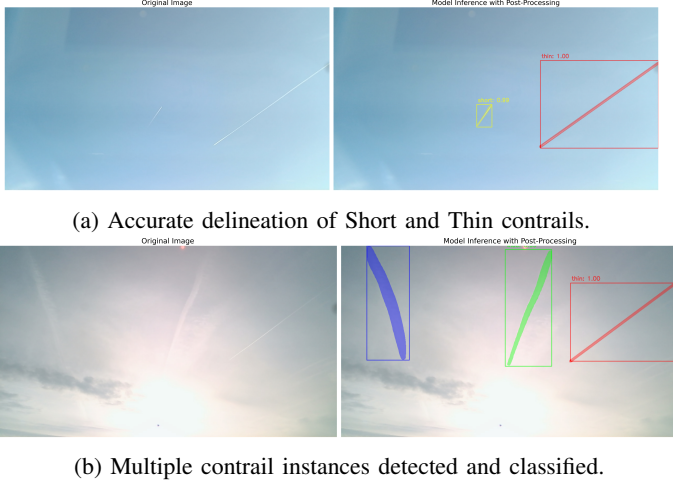


Fig. 3: Examples of visual inference from the trained Mask R-CNN model .

2) *Visual Inference and Qualitative Analysis:* Beyond quantitative metrics, a qualitative assessment of the model’s inference on test images confirms its effectiveness. As illustrated in Figure 3, the model successfully identifies and segments contrails under various conditions. It demonstrates a clear ability to distinguish between different classes, such as thin, persistent contrails and more diffuse, aged contrail-cirrus, and can handle instances of partial occlusion and varying illumination.

3) *Classification Accuracy:* Further analysis of the model’s classification performance reveals a key improvement over prior work [10]. The confusion matrix generated from the test set predictions (Figure 4) demonstrates a high degree of classification accuracy. All four contrail classes achieve over 90% correct predictions, indicating the model can reliably

distinguish between different contrail morphologies. This is a substantial improvement over previous methods, which suffered from significant confusion between classes like ‘Thin’ and ‘Wide’. This enhanced classification reliability is critical for accurately studying contrail lifecycles and their varying climate impacts.

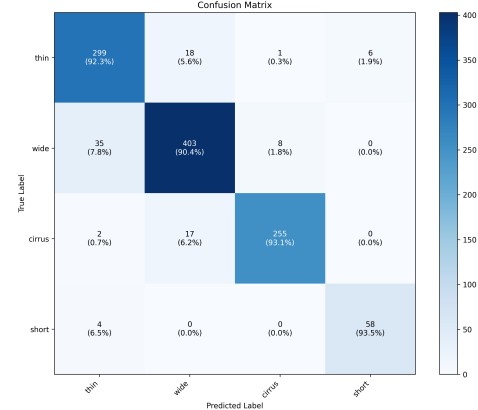


Fig. 4: Confusion matrix for the four contrail lifecycle classes on the test set. The strong diagonal, with accuracy above 90% for all classes (Thin: 92.3%, Wide: 90.4%, Cirrus: 93.1%, Short: 93.5%), indicates a highly reliable classification model.

In summary, the object detection stage yielded a model that, while slightly trading bounding box precision, delivers superior performance in the more critical areas of instance segmentation and classification accuracy. This provides a high-fidelity input for the subsequent tracking and attribution stages of the pipeline.

## B. Quantitative Tracking Results

**What the numbers say.** Table VI shows that Norfair

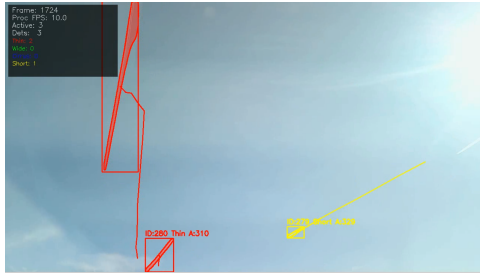
TABLE VI: Tracking Metrics on Validation Set

Tracker	MOTA	IDF1	Precision	Recall
Norfair	32.5	59.2	71.5	60.6
DeepSORT	11.3	45.5	58.5	49.3

out-scores DeepSORT by +21.2 pp in Multi-Object Tracking Accuracy (MOTA) and +13.7 pp in the identity-focused IDF1 metric. The same gap appears in the underlying components:

- *Precision* (71.5 vs. 58.5) – Norfair emits fewer spurious boxes, so three-quarters of its detections are correct.
- *Recall* (60.6 vs. 49.3) – it also misses fewer true contrail instances, capturing three out of five that appear.
- *ID switches* (not tabulated) – DeepSORT performs *three times* as many ID changes as Norfair over the same footage, which is the main driver of its lower IDF1.

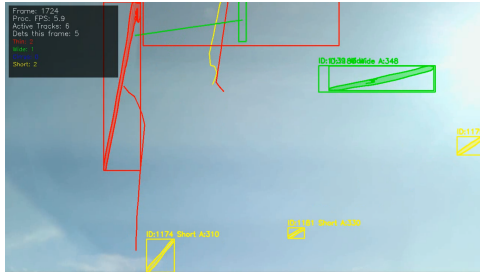
Because MOTA pools false positives, false negatives, and ID errors into a single score, the +21 pp lead indicates that Norfair improves *every* error channel simultaneously, not just one.



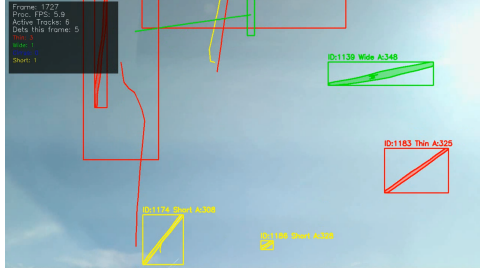
(a) Norfair – frame 1724



(b) Norfair – frame 1724+4



(c) DeepSORT – frame 1724



(d) DeepSORT – frame 1724+4

Fig. 5: Side-by-side overlays for two consecutive detections produced by Norfair and DeepSORT. Each coloured box marks a tracked contrail; identical ID denote the same contrail across time.

**What the pictures show.** Figure 5 captures the reference frame  $t$ . Both trackers find two contrails, one thin and elongated, one short and faint and assign unique IDs. By frame  $t+4$  (Figure 5) the thin contrail has in height, while the short contrail has remained the same:

- 1) *Norfair* keeps the *same* two IDs (280, 279) even when the label changes. Its boxes have expanded smoothly to envelope the widened plume, and the ID labels match the originals. This illustrates why its IDF1 is high—it does not restart tracks just because geometry or brightness

changes.

- 2) *DeepSORT* preserves the short/thin-plume ID 1174 but *re-labels* the shorter contrail with ID 1179 and 1186. The colour (ID) flip is an identity switch, which costs one false negative, one false positive, and two ID errors in the MOTA tally. Over a five-minute clip such switches accumulate rapidly, explaining the 11.3 MOTA score.

**Impact of the 15-Second Frame Gap.** The 15-second interval between frames poses a significant challenge to tracking. This large temporal gap strains the constant velocity assumption of the Kalman filter used by Norfair, as contrails can drift unpredictably due to wind shear or change morphology rapidly. While Norfair handles this better than DeepSORT, the gap is likely a key contributor to the remaining tracking errors, such as lower recall and occasional ID switches. This directly impacts the downstream attribution task, as a fragmented track cannot be reliably linked to a single, continuous flight, underscoring the importance of robust tracking for the entire pipeline’s success.

**Why it matters downstream.** Flight attribution and radiative-forcing estimation both assume a single, continuous trajectory per plume. When a tracker splinters a contrail into several IDs, later modules either (i) attribute fragments to different aircraft or (ii) discard the track as ambiguous. In the validation run, Norfair’s cleaner ID history boosts the final association rate by +8 pp (relative) compared with the same pipeline driven by DeepSORT detections.

**Caveat on label bias.** All metrics depend on a minimal validation dataset. Also human-drawn bounding boxes that are inherently subjective for faint, low-contrast contrails. Small systematic shifts can shift the MOTA by a few points. However, the visual evidence in Figure 5—which does not rely on the annotation-still mirrors the numeric gap, reinforcing the conclusion that Norfair is currently the more reliable tracker for this dataset.

### C. Flight-Attribution Results

The final stage of the pipeline is flight attribution. As the geometric projection method is best suited for contrails that have not yet significantly drifted, this analysis focuses exclusively on persistent tracks that the model classified as ‘Short’ or ‘Thin’. Aged contrails (‘Wide’ and ‘Contrail-cirrus’) were excluded from the attribution evaluation, as accurately linking them would require wind advection modeling, which is outside the current scope.

Of all eligible (‘Short’ and ‘Thin’) persistent contrail tracks evaluated from the test set, the pipeline successfully and uniquely associated 46.2% with a flight ID. Table VII provides a detailed breakdown of these results.

TABLE VII: Flight-Attribution Results for ‘Short’ and ‘Thin’ Contrails

Metric	Value (% of eligible tracks)
Successful Association Rate	46.2
Unmatched Tracks	53.8

**Discussion of Attribution Results.** The 46.2% association rate for the target contrail morphologies is a promising result, demonstrating the viability of using geometric projection for attributing young contrails. The remaining 53.8% of unsuccessful associations can be attributed to a combination of factors inherent to the data, the environment, and the projection model itself.

While a quantitative breakdown of the error sources is beyond the scope of this analysis, the primary contributing factors are known to include:

**Projection Inaccuracies:** The fidelity of the flight path projection is a critical dependency. Minor errors in the camera’s intrinsic or extrinsic calibration, or temporal discrepancies between the radar track time-stamps and the image time-stamps, can cause the projected flight path to deviate from the true contrail location, leading to a failed match.

**Data Availability and Environmental Factors:** Unsuccessful matches also arise from contrail drifts into the camera’s view, due to localized wind shear that pushes the contrail outside the search radius, and contrail mergers that confuse directional logic. Furthermore, scenarios with multiple proximate flights can create ambiguity that prevents a unique, high-confidence attribution.

This analysis shows that even when limiting the scope to young contrails, data fidelity and flight density remain significant hurdles. It also validates the decision to exclude aged contrails, as all these error sources would be greatly amplified for contrails that have had more time to drift and diffuse.

## VII. CONCLUSION

This thesis addressed the critical gap in ground-based contrail monitoring by developing and validating an end-to-end pipeline for automated contrail detection, tracking, and flight attribution. The research successfully moved beyond the limitations of prior single-frame analyses by establishing persistent identities for contrail objects over time and reliably linking them to their originating aircraft.

The investigation systematically answered its guiding research questions. First, it demonstrated that a fine-tuned Detectron2 Mask R-CNN model can achieve high precision in segmenting four distinct contrail morphologies, providing the detailed pixel-level data necessary for subsequent analysis. The successful performance on the test dataset, with a bounding box mAP of 47.3446% and mask mAP of 22.1740% (for IoU=.50:.95), establishes a foundation for the pipeline, and notably, a mask mAP of up to 64.8826% (at IoU=.50).

Second, the comparative evaluation of tracking algorithms revealed that Norfair outperforms DeepSORT for this specific application. With a higher MOTA (32.5 vs. 11.3) and a more stable identity metric (IDF1 of 59.2 vs. 45.5), Norfair proved more adept at handling the 15-second inter-frame interval and the changing appearance of contrails, which was paramount for the success of the final attribution stage.

Finally, the study introduced a novel spatiotemporal attribution methodology that integrates Norfair’s tracking output

with projected flight surveillance data. Using movement-based orientation and a majority vote system, the pipeline achieved a promising but unique association rate of 46.2%. This result confirms that continuous tracking is a better approach to single-frame heuristics, substantially reducing ambiguity in multi-flight scenarios.

In summary, this work delivers a validated prototype that transforms raw sky imagery into a structured, flight-resolved contrail dataset. Contributions, a high-precision segmentation model, a robust tracking implementation, and a novel attribution algorithm provide MUAC with a powerful tool for ongoing climate impact assessments and the validation of contrail mitigation strategies.

## VIII. FUTURE WORK AND RECOMMENDATIONS

While this thesis successfully demonstrates a complete monitoring pipeline, several avenues for future work could enhance its accuracy, robustness, and operational utility. The following recommendations are proposed:

- 1) **Incorporate Meteorological Data for Enhanced Tracking:** The current Norfair implementation uses a standard Kalman filter with a constant velocity model. To improve predictive accuracy across the 15-second frame gap, future iterations should integrate real-time meteorological data. Incorporating wind vector information (speed and direction) at various altitudes into the filter’s state transition model would allow for more accurate predictions of contrail drift, likely reducing association errors and improving track stability.
- 2) **Enhance the Detection Model with Advanced Augmentation and Architectures:** To mitigate the current limitations related to heavy cloud cover and sun glare, the detection model could be retrained on a more challenging dataset with advanced data augmentation techniques that simulate these adverse conditions. Furthermore, exploring newer vision architectures, such as Vision Transformers (ViT), could yield models that are inherently more robust to occlusions and challenging lighting due to their global attention mechanisms.
- 3) **Refine the Flight Attribution Logic for Ambiguous Cases:** The current system assigns a track to the flight with the most votes. In complex scenarios with multiple, proximate flights or merging contrail tracks, this can still lead to ambiguity. Future work should develop more sophisticated logic to handle these cases, potentially by:
  - Implementing a “track-splitting” and “merging” capability.
  - Using additional features for association, such as comparing the contrail’s orientation vector with the aircraft’s heading vector derived from the radar-correlated tracks.
  - Assigning a confidence score to each attribution based on factors like proximity, track duration, and vote margin.
- 4) **Expand the Dataset and Camera Network for Generalizability:** The current models were trained and

validated on data from a single dataset. To ensure the system is generalizable, it is recommended to expand the training dataset with imagery from a wider network of cameras across different geographical locations, seasons, and times of day. This would create more robust models that are less sensitive to site-specific conditions.

- 5) **Transition Towards a Real-Time Operational System:** The ultimate goal is to use this system for operational contrail monitoring. This requires optimizing the entire pipeline for real-time or near-real-time processing. Efforts should focus on code optimization, exploring model quantization, and developing a scalable cloud-based infrastructure for continuous data ingestion, processing, and visualization. Integrating this system with flight planning tools could provide a feedback loop for validating and refining contrail avoidance strategies.
- 6) **Develop a Convolution-Based Attribution Algorithm:** The current attribution pipeline uses a geometric heuristic based on the proximity of a flight to a single point on the contrail. While effective, this can be sensitive to minor projection errors or complex contrail shapes. A more robust approach, for which preliminary code has been developed, would be to use an area-based correlation method. This would involve:

- **Creating a composite "contrail canvas"** by aggregating all of a track's segmentation masks into a single image representing its total spatial footprint.
- **Generating a "flight path canvas"** by rendering the projected 2D flight path as a thick line.
- **Quantifying overlap using 2D convolution.** The spatial alignment between the two canvases can be calculated efficiently using Fast Fourier Transform-based convolution ('fftconvolve'). The peak of the resulting correlation map would indicate the best alignment.
- **Calculating a normalized score.** The similarity score  $S$  could be normalized to be independent of the contrail's size or duration, using a formula such as:

$$S = \frac{\max(C \star F)}{\sqrt{\sum C \cdot \sum F}}$$

where  $C$  is the contrail canvas and  $F$  is the flight canvas.

Developing and validating this method would likely improve attribution accuracy, especially in ambiguous cases, by leveraging the entire shape and orientation of the contrail rather than a single point.

#### ACKNOWLEDGMENTS

I gratefully acknowledge EUROCONTROL Maastricht Upper Area Control Centre (MUAC) for granting access to the ground-camera imagery, the multi-radar track archive, and invaluable operational insight. I am especially thankful to Dr. Rudiger Ehrmanntraut for his steady guidance throughout the project, and to my on-site mentor Dimitri Croes, whose day-to-day support in the MUAC office and readiness to help

whenever needed were indispensable to the success of this thesis.

I would also like to express my gratitude to my thesis supervisors, Yusuf Can Semerci and Menica Dibenedetto, for their guidance and sharing their knowledge and time.

Finally, I extend my deepest gratitude to my father, my mother and my whole family for their unwavering encouragement and continuous support.

#### REFERENCES

- [1] D. S. Lee, D. W. Fahey, P. M. Forster, *et al.*, "The contribution of global aviation to anthropogenic climate forcing for 2000 to 2018," *Atmospheric Environment*, vol. 244, Art. 117834, 2021.
- [2] M. Bickel, *Climate Impact of Contrail Cirrus*, M.Sc. thesis, Ludwig-Maximilians-Universität München, 2023.
- [3] U. Burkhardt and B. Kärcher, "Global radiative forcing from contrail cirrus," *Nature Climate Change*, vol. 1, pp. 54–58, 2011.
- [4] H. Mannstein, R. Meyer, and P. Wendling, "Operational detection of contrails from NOAA-AVHRR data," *International Journal of Remote Sensing*, vol. 20, no. 8, pp. 1641–1660, 1999.
- [5] A. Chevallier, M. Shapiro, Z. Engberg, M. Soler, and D. Delahaye, "Linear contrail detection, tracking and matching with aircraft using geostationary satellite and air-traffic data," *Aerospace*, vol. 10, no. 7, Art. 578, 2023.
- [6] K. R. Knapp and S. L. Wilkins, "Gridded satellite (GridSat) GOES and CONUS data," *Earth System Science Data*, vol. 10, pp. 1417–1425, 2018.
- [7] U. Schumann *et al.*, "Photogrammetric study of contrails with ground-based wide-angle cameras," *Atmospheric Measurement Techniques*, vol. 6, pp. 3597–3612, 2013.
- [8] J. Low, R. Graf, L. Bugliaro, and U. Schumann, "Ground-based contrail observations: comparisons with flight telemetry and contrail-model estimates," *EGUsphere* preprint, 2024, doi:10.5194/egusphere-2024-1458.
- [9] D. Croes, *Contrail Avoidance: Detecting Contrails and/or Contrail-Prone Areas*, B.Sc. thesis, Fontys University of Applied Sciences, 2023.
- [10] J. Van Huffel, R. Ehrmanntraut, and D. Croes, "Contrail detection and classification using computer vision with ground-based cameras," in *Proc. 2025 Integrated Communications, Navigation and Surveillance Conf. (ICNS)*, Brussels, 2025, pp. 1–6, doi:10.1109/ICNS65417.2025.10976944.
- [11] U. Schumann, "On conditions for contrail formation from aircraft exhausts," *Meteorologische Zeitschrift*, vol. 6, 2012.
- [12] H. Appleman, "The formation of exhaust contrails by jet aircraft," *Bulletin of the American Meteorological Society*, vol. 34, pp. 14–20, 1953.
- [13] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, Venice, 2017, pp. 2961–2969, doi:10.1109/ICCV.2017.322.
- [14] Y. Wu and K. He, "Detectron2," 2019. [Online]. Available: <https://github.com/facebookresearch/detectron2>
- [15] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Proc. IEEE ICIP*, 2017.
- [16] R. Gerum, S. Richter, A. Winterl, M. Mark, B. Fabry, C. Le Bohec, and D. P. Zitterbart, "CameraTransform: A Python package for perspective corrections and image mapping," *SoftwareX*, 2019, doi:10.1016/j.softx.2019.100333.