Satellite-based detection of contrails using deep learning

by

Luke Kulik

B.Sc., Aerospace Engineering, Delft University of Technology (2016)

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of

Master of Science in Aeronautics and Astronautics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2019

Signature redacted

Author .....
Department of Aeronautics and Astronautics
Aug 20, 2019

Signature redacted

Certified by .....
Steven R.H. Barrett
Associate Professor of Aeronautics and Astronautics
Thesis Supervisor

Signature redacted

Accepted by .....
Sertac Karaman
Associate Professor of Aeronautics and Astronautics
Chair, Graduate Program Committee

Satellite-based detection of contrails using deep learning

by

Luke Kulik

Submitted to the Department of Aeronautics and Astronautics
on Aug 20, 2019, in partial fulfillment of the
requirements for the degree of
Master of Science in Aeronautics and Astronautics

Abstract

Aircraft condensation trails, known as contrails, are estimated to account for approximately half of all climate warming resulting from aviation emissions, but uncertainty in coverage is high [12, 2]. These artificial clouds form behind aircraft and trap outgoing infrared radiation, leading to a warming effect [12, 3, 28, 40, 3]. With aviation forecast to contribute up to 15% of total anthropogenic warming by 2050 [18], an accurate estimate of total contrail-attributable warming is required. Automated detection of contrails using satellite images has not been feasible to date due to the complex evolution of contrail shapes and their similarity to cirrus clouds and land features [21, 12, 39]. This has been a limiting factor in quantifying contrail climate impacts [39, 3]. We use a deep learning approach to overcome these difficulties. Instead of relying on fragile, hand-engineered features, we train a deep convolutional neural network directly on a dataset of annotated infrared images from a geostationary satellite. We estimate year-round contrail coverage over the continental United States, based on automated classification of imagery extracted from the Geostationary Operational Environmental Satellite system. We find that contrails cover an average of 0.55% of the contiguous United States, an area equivalent to a quarter of the size of Florida. We show that the relationship between total contrail coverage and air traffic varies as a function of time and location, with greater contrail area produced in early morning than in the afternoon even if traffic volume is the same. We also find a seasonal dependence, with contrail coverage reaching a minimum in late summer at the same time that air traffic peaks. We create contrail maps showing maxima over major air traffic hubs like San Francisco, Chicago, Dallas and highlighting major contrail-producing airways. We release the first publicly available benchmark dataset of contrail annotations to serve as a platform for future research of contrail formation and impacts.

Thesis Supervisor: Steven R.H. Barrett
Title: Associate Professor of Aeronautics and Astronautics

Acknowledgments

I would like to thank Professor Steven Barret and Doctor Sebastian Eastham for their guidance through this thesis. Thank you to my parents for their unwavering support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Contrails are thin ice clouds produced by aircraft engines during cruise, at altitudes of 8 to 13 km [12]. Contrails disturb the balance between outgoing thermal radiation from the Earth's surface and downwelling solar radiation, leading to a net warming effect [14]. Radiative imbalances such as these can be quantified as a positive radiative forcing (RF) [29], from which climate damages are calculated. Contrails have been estimated to be the largest single RF component associated with aviation [2, 12], potentially exceeding impacts due to cumulative aviation CO_2 emissions [12]. With aviation traffic forecast to double in the next 18 years, these impacts are expected to grow with total aircraft traffic, while changes in aircraft efficiency and background climate are thought to result in further increases in contrail formation likelihood [10, 18]. However, significant uncertainties still exist regarding the magnitude of the RF contribution, properties, and lifecycle of contrails [12, 40, 2, 18, 8, 14]. These have been attributed to scarce and low-quality observational data [12, 2, 3]. In particular, no observational data yet provides continuous contrail coverage information. The current state-of-the-science for contrail detection is performed using hand-calibrated legacy algorithms, which rely on post-processing using external flight routes databases [39]. Research is also hindered by a lack of data sharing, as existing datasets have not been made publicly available. These issues need to be addressed if effective policy decisions for reducing the impact of aviation on the climate are to be made.

Recently-launched, high-resolution geostationary satellites (GOES-East, GOES-

West) offer an opportunity to study contrail evolution over the Western Hemisphere in near real-time. However, with the expected increases in the daily volume of satellite imagery (50 PB of data per year by 2023 [27]), automated algorithms are needed to quickly and accurately extract scientific variables of interest. Deep learning is a possible solution to this problem, given the progress that has been made in recent years in developing robust vision systems based on deep convolutional neural networks (CNNs), which are able to learn the features in images without direct human supervision [13]. Machine learning using satellite imagery is a growing area of research [4, 11, 16, 17, 15]. However, neural networks for contrail detection were previously deemed infeasible due to classification difficulty [12] and excessive data and training time requirements [21, 22]. In this work, we overcome these limitations by revisiting key assumptions made in previous analyses - we collect a new, high resolution contrail dataset; use modern convolutional neural network architectures, and train our models on Graphical Processing Units (GPUs), optimized for this task. We aim to quantify diurnal variations in contrail coverage by applying deep learning methods to high temporal-resolution satellite observations. Comparing these observations to schedule data, will enable us to determine the empirical relationship between contrail coverage and flight volume by time of day, season, and region.

1.1 Related Work

In order to understand the approach being employed, we first outline the approach used for a similar problem - the detection of roads from satellite imagery. We then discuss the current state-of-the-science for contrail detection, and highlight three key areas in which these approaches are failing.

1.1.1 Road detection from satellite images

A task similar to linear contrail detection that has garnered more interest from the computer vision community is road detection from satellite images. Roads on satellite images are, similarly to contrails, slender and sparse features. Detecting roads is

a relatively simpler task to learn because of their predictable, uniform color and sharp edges. However, it still took 30 years of work on the problem to solve it, with Mnih et al. neural-network based solution reaching feasibility first in 2010 [26]. Figure 1-1 shows road segmentation maps closely resembling the contrail segmentation task. Since then, a number of more modern approaches using fully convolutional encoder-decoder neural networks have emerged [1, 45]. This work builds on the recent generative adversarial network (GAN) based approach that requires minimum post-processing [6]. The modern, GAN approach promises an opportunity to exploit the semantic structure of images, intuitively understood by humans, directly during training.

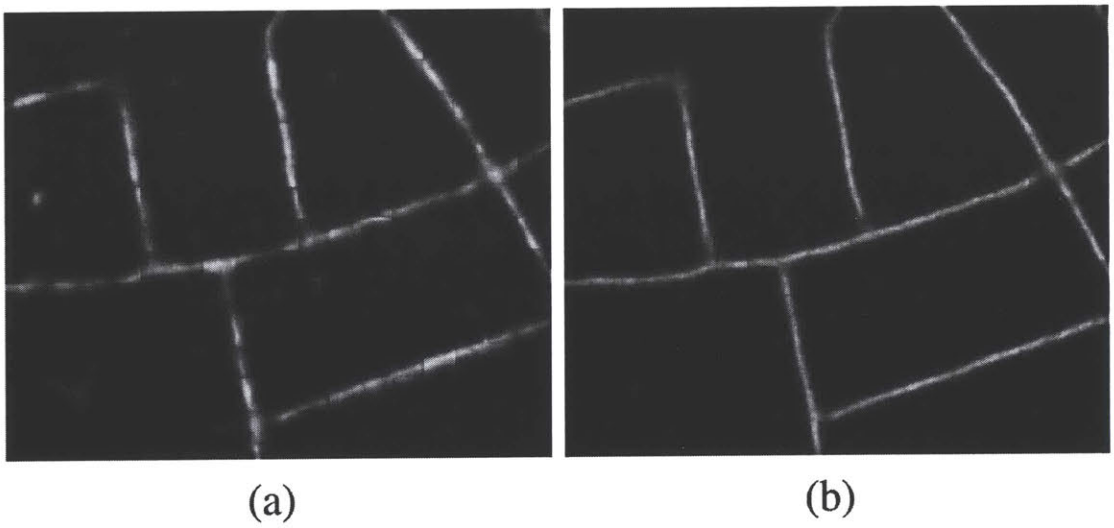

Figure 1-1: Road segmentation maps (a) before and (b) after post-processing by Mnih et al [26], closely resembling contrail segmentation task

1.1.2 Contrail detection from observational data

The previous observational studies of contrail coverage suffered from at least one of the three fundamental drawbacks. These are:

- Relying on low earth orbit satellites, which provide at most two overpasses a day, making it impossible to capture the diurnal trends in contrail coverage

- Using algorithms with noise characteristics that drown out contrail patterns and necessitated filtering with flight datasets
- Lack of scalability due to human labor involved (manual parameter tuning or manual data collection)

All but four previous studies were based on a single traditional computer vision algorithm [21, 39, 31, 41, 43, 30, 42]. Two other works attempted neural network based methods, but were unable to produce a comprehensive contrail inventory [22, 44]. The remaining two used sparse ground observations [24, 20].

The traditional approach relied on hand-engineered features designed to detect contrails of a set width on a particular type of satellite imagery and required extensive pre-processing [21]. Due to the low signal-to-noise ratio in the traditional contrail detection algorithm, contrail patterns reveal some broad trends, but do not appear accurate enough to resolve any flight corridors, as shown in Figure 1-2. The algorithm also constrained the viewing zenith angles to less than 50° for its optimal performance [31] and had a tendency to pick up linear land features and elongated cirrus as false positives [25]. This necessitated filtering with external flight path datasets [39], which makes analysis of how flight traffic affects contrails very difficult, because of the introduced coupling.

Neural networks are another potential solution, but were previously deemed infeasible due to classification difficulty [12] and excessive data and training time requirements [21, 22]. Meinert et. al failed to produce any results [21], while Zhang et al. (2018) claimed to have successfully trained a shallow neural network to recognize contrails over Japan, but no coverage estimates, trends, or datasets were provided [44], making this finding of limited usefulness.

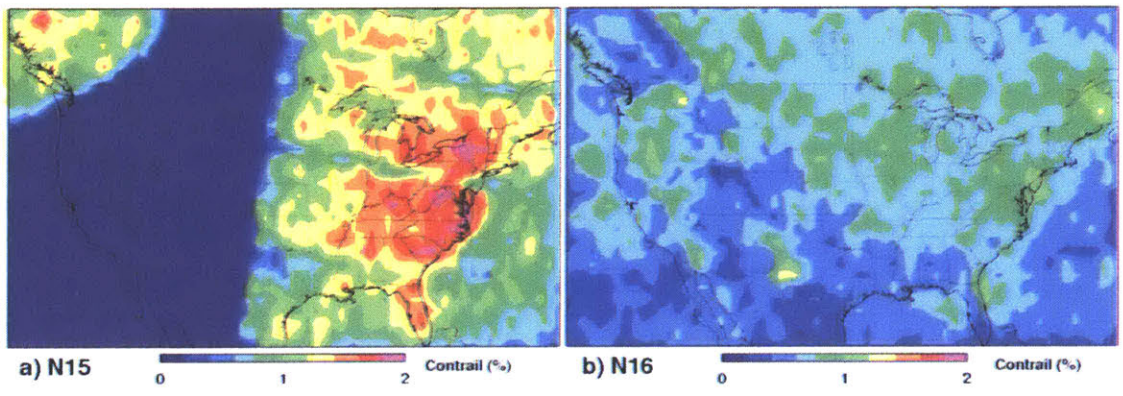

Figure 1-2: Mean 2001 daytime contrail coverage over USA domain estimated using Mannstein's contrail detection algorithm on NOAA-15 and NOAA-16 satellite images [31]

1.2 Method outline and key concepts

This section introduces the methods used in this thesis and provides an overview of the following chapters. The key idea of this work is to connect the developments in remote sensing and computer vision in light of the pressing concerns of aviation's climate impact. To date, no public datasets of contrail annotations are available, impeding progress and creating a barrier to entry for improving the automated contrail detectors (referred to here as "classifiers" or "detectors"). Existing work on observational estimates of contrails is not sufficiently temporally resolved to capture high frequency trends, is too noisy to accurately resolve contrail maxima and not scalable to new instruments without manual algorithm tuning or heavy post-processing.

In this work, we train a deep, convolutional neural network to identify contrails using a dataset of hand-annotated satellite images. We use high-frequency satellite images from easily-accessible cloud-based sources for easy replicability and extensibility of results. We leverage established computer vision methods from similar tasks e.g., road detection from satellite images with a focus on end-to-end learning (no post-processing). As an additional product we create a public dataset of contrail annotations to serve as an algorithm's benchmark.

Chapter 2 describes how we acquire and prepare the data for training and inference. We provide a rationale for why GOES-East satellite is selected as a source of

imagery and explain the details behind dataset acquisition and pre-processing. We also describe the labeling process for a large dataset of binary ("yes/no") classifications, in addition to a smaller set of pixel-level classifications, and discuss dataset split and verification. In Chapter, 3 we show how we selected and trained architectures for a simple binary classifier and pixel-level segmentation network. We provide quantitative proof that the networks are able to capture the complex structure of contrails using infrared satellite images. Finally, Chapter 4 outlines the scientific results achieved with the help of the developed contrail classifiers. We show daily and seasonal trends in contrail cover and compare it to air traffic. We end with a discussion of limitations and future work in Chapter 5.

Chapter 2

Data acquisition and preparation

To generate high-quality contrail coverage maps, two components are required: a high-resolution satellite dataset; and a high-precision, generalizable classification algorithm. This section describes the steps taken to select and prepare the most suitable satellite dataset for both classifier training and generating contrail maps for scientific use.

2.1 Satellite selection

Satellite imagery selection is driven by a trade-off between spatial and temporal resolution. For a fixed optical instrument quality, an imager in a lower orbit will create a higher resolution image at the cost of decreased swath width (and thus decreased temporal resolution). In the extreme case, geostationary (GEO) orbits provide a near-real time view of a large fraction of the Earth's surface, but at the cost of decreased spatial resolution. The satellites considered at the time of writing of this thesis are shown in Table 2.1.

We select GOES-16 (a.k.a. GOES-East) as the source of imagery. It satisfies our requirements, because of its very high temporal resolution, potential for scalability, and good spatial resolution and coverage. For a comprehensive estimate of contrail cover over the United States all of these are required. Good temporal resolution is needed to capture high-frequency air traffic and meteorological effects on contrail

Satellite name	Coverage	Spatial res. [m]	Temporal res.	Cloud availability
GOES-16	Americas (GEO)	500 - 2000	5 - 15 minutes	Yes
Himawari-8	Australia, Asia (GEO)	500 - 1000	10 minutes	No
MODIS Aqua+Terra	Global (LEO)	250 - 1000	0.5 - 2 days	Selected products
AVHRR	Global (LEO)	1100	0.5 days	No
Landsat 8	Global (LEO)	30	16 days	Yes
Sentinel-2	Global (LEO)	10	10 days	Yes

Table 2.1: Satellite imagers considered as data inputs to the contrail detector. GEO indicates geostationary orbit, LEO indicates low Earth orbit. Cloud availability is determined through Amazon Earth.

cover. Moreover, a temporally-coherent, static view of the whole domain provided by a GEO satellite will let us see coherent patterns of contrail distribution. Algorithms developed for a portion of the GOES-16 domain will be readily extensible to other areas covered by the satellite (e.g., South America), as well as other satellites in the fourth generation of GOES family (e.g., GOES-17). The planned mission duration for GOES-16 and GOES-17 reaches until year 2031, ensuring at least 12 years of additional data. With small adjustments the developed detectors should also be applicable to similar instruments such as Himawari-8 as well as satellites that will be launched in the future.

2.2 Dataset retrieval

We use 28 terabytes (TB) of full-disk images acquired by Advanced Baseline Imager (ABI) on board of GOES-East and downloaded from NOAA's Comprehensive Large Array-data Stewardship System (March 1, 2017 - July 9, 2017) and Open Commons Consortium (OCC) Environmental Data Commons (July 10, 2017 - July 1, 2019). We download the data using Amazon Web Services Command Line interface:

```
aws s3 sync s3://noaa-goes16/ABI-L2-MCMIPF/
↪ /workspace/data/noaa-goes16/ABI-L2-MCMIPF/ --no-sign-request
↪ --endpoint-url https://osdc.rcc.uchicago.edu --no-verify-ssl
```

Based on the physical properties of contrails, we use four out of 16 infra-red frequency bands: 8.5 µm ("Cloud-Top Phase"), 10.35 µm, 11.2 µm, 12.3 µm ("IR Longwave Windows") all at a spatial resolution of 2 km. We also use three additional

visible light frequency bands for labeling: 0.64 µm ("Red"), 0.87 µm ("Veggie"), 0.47 µm ("Blue") at higher spatial resolution between 0.5-1.0 km. Full-disk images of the Earth are captured at 15 minute intervals throughout most of the dataset, with 10 minute intervals from 2 April 2019 onward. In the interval of interest 98.2% of data is available before outlier filtering, excluding the Nov 30 - Dec 14, 2017 period when the satellite was repositioned from 89.3° W to 75.2° W.

2.3 Data preparation

We select our region to include the contiguous US, capturing 60% of the air traffic (by fuel burnt in 2016) in the GOES-16 domain. We reproject, re-center, and crop the original full disk images to minimize the distortion in this area. We warp the GOES-East images to orthographic projection centered on the geographic center of the contiguous United States (39.8283° N, 98.5795° W) using nearest neighbor interpolation at single-precision floating point to avoid introducing sampling artifacts. Figure 2-1 shows a sample reprojected image, warped to minimize distortions in contrail area and shape.

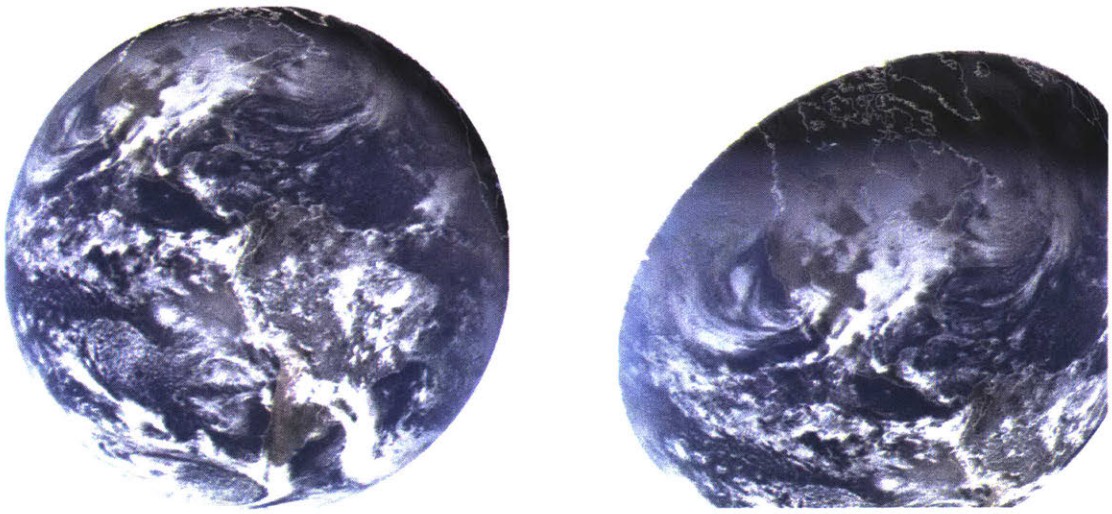

Figure 2-1: US-centered ortographic reprojection sample (right). Source full disk image (left) taken on Sunday, Dec 17, 2017, 18:00 UTC

The spatial extent of the resulting image is 63.13° W - 134.03° W and 10.41° N - 50.08° N, projected onto a six megapixel bitmap. The total area covered by our domain is 23.73 million km^2, covering 26% of Canada, 87% of Mexico and all of the contiguous United States (i.e., excluding Hawaii and Alaska). The maximum viewing angle in the total domain is 72.4° (61.8° for the United States). 99% of the contiguous United States area is distorted $\left(\left|\frac{A_{\text{projected}}}{A_{\text{actual}}}\right|\right)$ by less than 15.5%. Re-centering and cropping allow us to decrease the dataset size by 80%, from 28 TB to 6TB. High resolution maps of viewing angle and area distortion are available in Appendix A.

The target projection is defined in Geospatial Data Abstraction Library as:

```
t_p = '+proj=ortho +ellps=GRS80 +lat_0=39.8283 +lon_0=-98.5795
↪ +sweep=x +no_defs'
```

Whereas the boundaries of the domain are calculated as:

```
1 tempPrj.ImportFromProj4('+proj=longlat +ellps=GRS80 +no_defs
  ↪ +sweep=x')
2 tx_r = osr.CoordinateTransformation(tempPrj, t_p)
3
4 (urx, ury, ulz) = tx_r.TransformPoint(max_lon, max_lat)
5 (llx, lly, lrz) = tx_r.TransformPoint(min_lon, min_lat)
6
7 dest = gdal.GetDriverByName('MEM').Create('', (int((urx - llx) /
  ↪ pixel_spacing)),int((ury - lly) / pixel_spacing), 1,
  ↪ gdal.GDT_Float32)
```

For `[min_lon,min_lat,max_lon,max_lat,pixel_spacing]=[-134.03,10.41,-63.13,50.08,2004]` this results in a 2000 by 3000 pixel image.

2.4 Contrail labeling

Deep learning classification algorithms aim to robustly separate the features describing each class ("contrail"/"no contrail") . Precision and generalization of this separation depend on the model architecture and the training dataset. This section explains how we create the latter.

In order to train a neural network we start with a model architecture, consisting of filters arranged in consecutive layers. Filters consist of weights that are randomly initialized. As we pass the training images through all of the filters we get an initial (bad) prediction. We compare this prediction against a training set label using a predefined loss function (e.g., mean squared error) and "backpropagate it" i.e., take a derivative with respect to all of the weights. Each of the weights is then appropriately updated to minimize the prediction error. The process is repeated until the error on the validation set (separate from the one we train on) is minimized.

When showing data to labelers for the purposes of creating an annotated training set, we use seven image channels, displayed to labelers as three images: "True Color" (wavelengths: 0.64 µm, 0.87 µm, 0.47 µm) , "Infra-red" (difference between 12.3 µm and 10.35 µm) and "Infra-red False Color" (8.5 µm, 11.2 µm, 12.3 µm). For ease of interpretation we augment "True Color" images with infra-red (IR) 10.35 µm channel for pixels where IR intensity exceeds color intensity (at night). Contrails appear as white streaks in "True Color" and "Infra-red" imagery and dark blue lines in "IR False Color", where they can be also easily distinguished from non-ice clouds - see Figure 2-2. The process of generating "IR False Color" images is described in more detail at the end of this chapter.

We create two datasets, each containing a comparable number of pixels. The first is a "coarse" corpus, consisting of 24,321 250ÃŮ250 pixel tiles, each of which has a single binary label indicating whether or not any contrails are present in the tile. This corpus was created by a group of 16 atmospheric science and engineering experts, with self-consistency verification performed through labeling 10% of the images twice. The second set is of 110 2,000×3,000 pixel-level annotations. In this set,

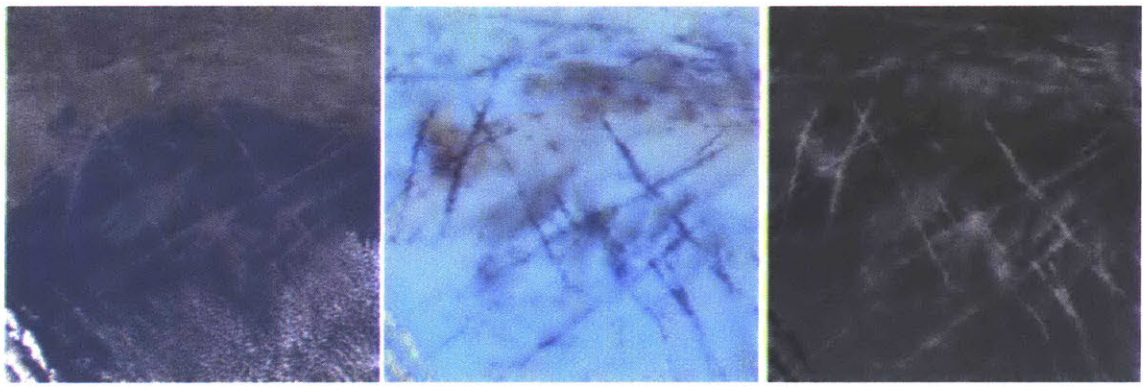

Figure 2-2: Cross-hatching contrail breakout south of Houston and Louisiana in RGB, IR False Color and Infra-red. Tile dimension is approximately 350km. Image taken on Monday, Jan 15, 2018, 17:15 UTC

individual contrails are outlined, providing pixel-level contrail annotations. This set was produced by two labelers, each provided with dedicated training to recognize and annotate contrails.

Labelers of both datasets were trained to recognize and focus on persistent linear contrails and contrails in the initial diffusion phase, which maintain linear shape - see Appendix B for details. They were also trained on how to recognize and avoid labeling cloud edges, static features, stratified clouds, and high aspect ratio cirrus.

We verify the coarse dataset through self-consistency and adversarial checks. Each image is labeled at least twice and 10% of the dataset is replicated. This allows us to assess self-consistency (how likely is it that the same image will get labeled the same way twice) and adversarial consistency (how likely is it that labelers will agree) - the results of this study are summarized in Table 2.2.

	#1	**#2**	**#3**	**#4**	**#5**	**#6**	**Mean**
Self-consistency #1	89	90	81	89	85	87	88
Self-consistency #2	88	93	88	92	91	88	
Adversarial consistency	81	85	79	87	77	81	82

Table 2.2: Percentage agreement between tile labels for six coarse data packages

We remove all tiles that are not adversarially matching from the dataset. The final dataset consists of 21,888 training labels, 1,946 validation labels, and 487 test labels. Of this set of tiles, 42% include contrails. For the pixel-level annotations 0.65% of pixels are labeled as contrails. A random selection of labeled tiles is shown in Figure 2-3.

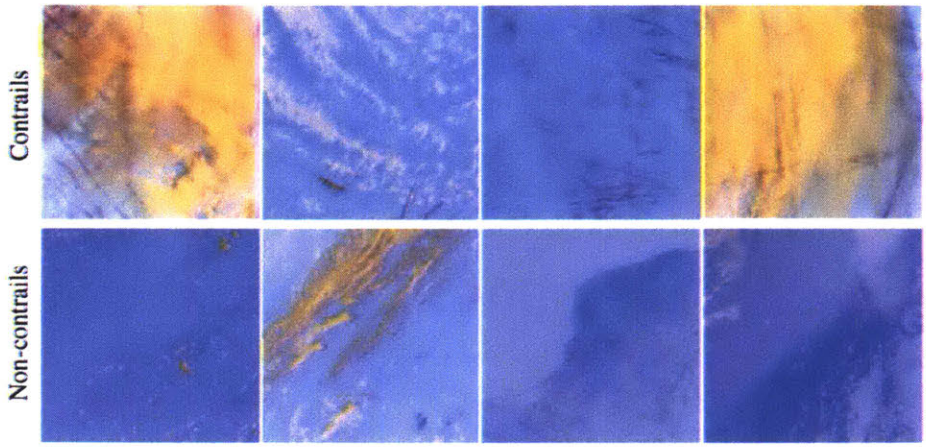

Figure 2-3: Examples of contrail and non-contrail labeled tiles

Similarly, each pixel-level annotation is verified by two labelers independently. The result, plotted by aggregating across all observations, shows expected spatial patterns (see Figure 2-4). For example, we see more contrails on the coasts and around major airports, consistent with both our expectations based on aircraft traffic and with prior studies of contrail behavior.

IR False Color

Ash false-color (referred to in this work as "IR False Color") RGB product was originally designed to identify volcanic ash and sulphur dioxide clouds [32], but can also make contrail identification easier. To enhance ash features the 12.3 µm - 11.2 µm difference, the 11.2 µm - 8.5 µm difference, and the 11.2 µm channels are constrained to a predefined range when the color shading is assigned. The color intensity is given by Equation 2.1 [5].

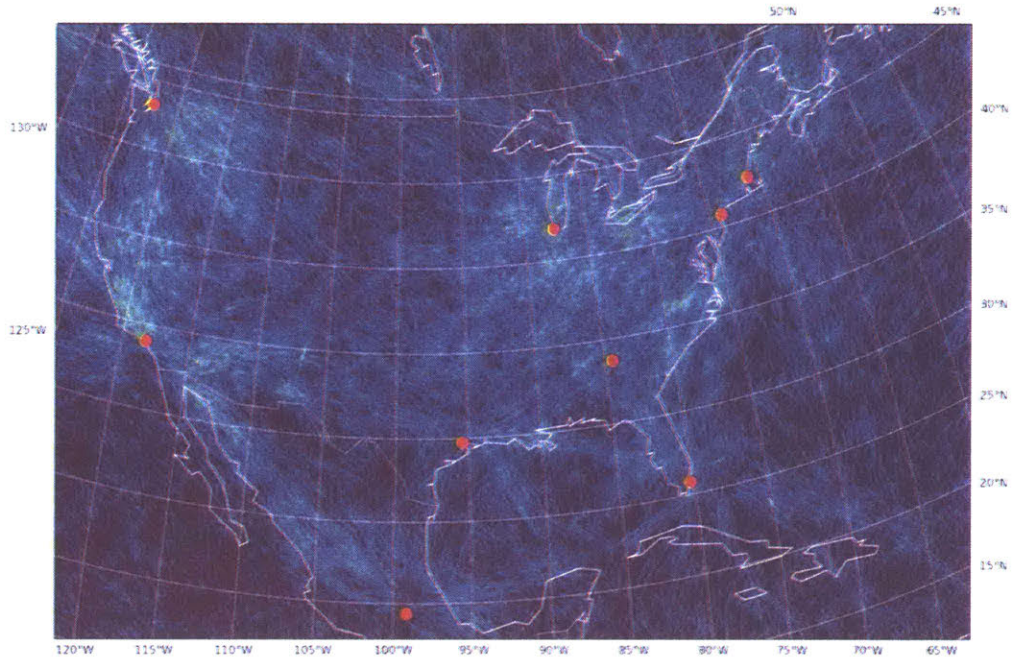

Figure 2-4: Aggregate of 110 pixel-level human annotations of contrails. Red dots represent major airports: Seattle, NYC JFK, Boston, Miami, LA, Houston, Atlanta, Chicago, Mexico City

$$(R, G, B) = 255 \left(\frac{(T_{R,G,B} \text{ or } \Delta T_{R,G,B}) - T_{\min}}{T_{\max} - T_{\min}} \right)^{\frac{1}{\gamma_{R,G,B}}} \tag{2.1}$$

Ash product is IR-based and therefore available 24 hours a day. $T_{\max}$ and $T_{\min}$ represent the range of values used to constrain each of the color components, while γ is the enhancement factor [5]. The temperature ranges for each of the channels are summarized in Table 2.3. Ash product's utility for contrail detection can be attributed to the blue-channel temperature range designed for distinguishing between surface and cloud temperatures and tuned for highlighting ice clouds [32].

Color	**Wavelength**	**Interpretation**	**Min [°K]**	**Max [°K]**	**Gamma**
Red	12.3 µm - 11.2 µm	Optical thickness	-4	2	1
Green	11.2 µm - 8.5 µm	Particle phase	-4	5	1
Blue	11.2 µm	Temperature	243	303	1

Table 2.3: Bands used to create a composite false-color IR image from [34] including physical interpretation and temperature ranges

Chapter 3

Network development

In order to generate contrail maps, we need a high precision, generalizable classifier. This chapter explains the details behind our two-stage approach to training and evaluating deep neural network architectures on the two separate datasets described in Section 2.4. We first train a binary classifier, which determines if a 250-500 km tile contains contrails or not. We then train a semantic classifier which refines its predictions on a pixel-level (approximately 2 km resolution) contrail map.

3.1 Binary classification

To train a binary classifier we implement a widely used residual neural network architecture "ResNet50". The network is first pretrained to 92% accuracy (counting top five classes) on the 1,000 object classes (1.2 million images) of the 2012 ImageNet Challenge. We then modify it for the purpose of this project by replacing ResNet's average pooling layer with a PyTorch's adaptive average pooling and substituting the last fully connected layer with a new one with 2-class output. The overview of the network architecture is shown in Figure 3-1. We finally train it on the coarse tile-level dataset using three IR channels as input: 8.5 µm, 11.2 µm and the difference between 12.3 µm and 10.35 µm.

We use Facebook's PyTorch deep learning framework to train, validate, and test our network. To match the range of ImageNet pretraining inputs we normalize (center

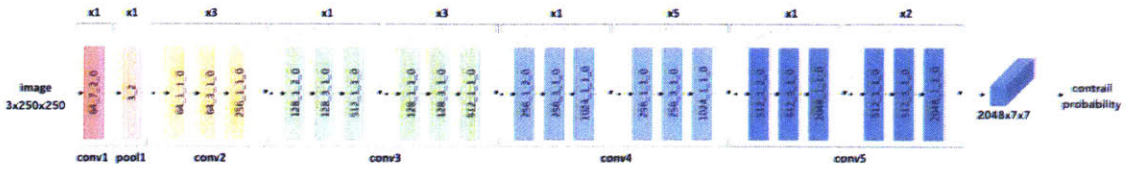

Figure 3-1: Modified ResNet50 architecture used for coarse dataset classification [7]

at zero and scale by standard deviation) the tile images with `data_mean=[274.16, 275.74,3.05]` and `data_std=[18.37,19.50,1.85]` for respective IR channels. We perform network weights update based on a gradient from 50 randomly selected training images (training batch size). We augment the dataset with random horizontal and vertical flipping. For smaller datasets, we also find that color jittering (slightly changing values of hue, contrast, saturation, brightness) is effective at preventing overfitting. We train our CNN using backpropagation with stochastic gradient descent - the optimizer parameters are `[learning_rate,momentum,weight_decay]=[1e-3, 0.9,1e-5]`. We use a learning rate schedule which decreases the learning rate by $10\times$ after the third, sixth, and 20th iteration, found empirically. We train until validation loss stops decreasing.

The resulting classifier is 90% accurate (fraction of images that are correctly classified) at a 97% precision (fraction of positives that are true). We correctly classify 99% of clear-sky images and 79% of contrail-containing images. This is shown in the confusion matrix in Figure 3-2 (left panel).

We explore the network predictions further by inspecting a sample of its outputs. Figure 3-3 shows correct predictions with the highest and lowest contrail probabilities and their corresponding images in IR False Color. Tiles with a high proportion of contrail pixels are detected with high probability, whereas tiles that lack contrails and are comprised mostly of non-ice clouds, land, and ocean are assigned 0.0 probabilities.

Figure 3-4 shows the classifications that are "the most incorrect" i.e., their estimated probability is the farthest from a human-assigned label. These false positives and false negatives visibly contain very little information to make a tile-level distinction between contrail and non-contrail images for both the classifier and human

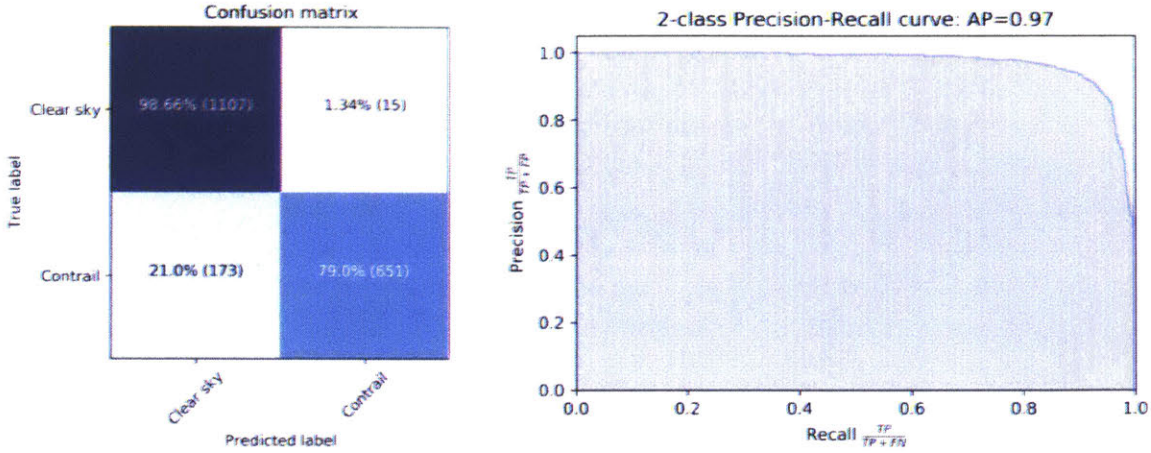

Figure 3-2: Confusion matrix and precision-recall curve for the trained binary contrail classifier

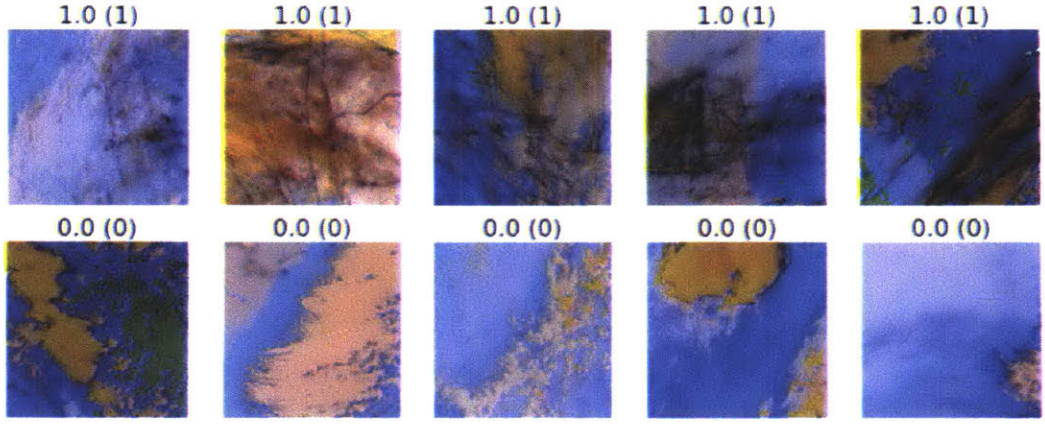

Figure 3-3: Correct classifications: first row shows true positives, second row true negatives. Number above the tile represents the estimated probability of a contrail detection, while the number in a bracket is the human-assigned label

labelers. We speculate that the classifier learns to detect difficult examples better than individual humans, as it attempts to match the common set of contrail features implicitly used by multiple human labelers.

3.1.1 Gradient visualization

For a trained neural network, the output probability depends on how particular input pixels or pixel patches "activate" the learned filters in the network. We take advantage of that, by tracing the gradients of contrail detection flowing into the final (8th)

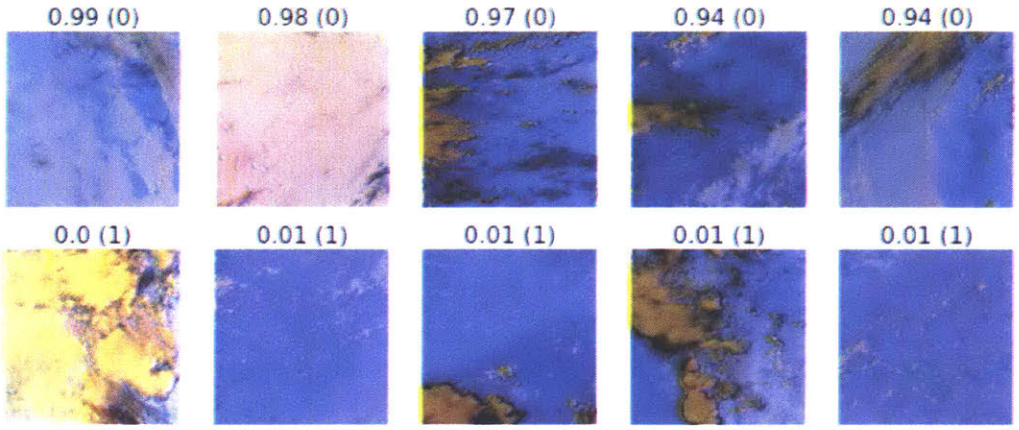

Figure 3-4: Incorrect classifications: first row shows false positives, second row false negatives. Number above the tile represents the estimated probability of a contrail detection, while the number in a bracket is the human-assigned label

convolutional layer of the ResNet network to highlight the relevant regions in the image. We use approaches and concepts described in [38, 9, 33] to produce coarse localization maps shown in Figure 3-5.

Figure 3-5: Gradient-based highlight of the pixels important for positive contrail classification, overlaid on a False-color IR image

Resulting localization maps largely overlap with contrails in the images. However, they are prone to spanning very large areas also including cirrus clouds. The coarse quality of the localizations combined with the diffuse nature of contrails (difficult

to extract sharp object edges) makes it insufficient for estimating metric contrail area coverage, necessitating a different approach if automated, pixel-level contrail identification is to be achieved.

3.2 Adaptive generative segmentation

To compute the high-resolution metric area coverage of contrails over the GOES domain, we re-formulate the problem as semantic segmentation and apply an encoder-decoder convolutional neural network architecture. Fully convolutional neural networks are capable of producing high accuracy segmentations for natural images and have been successfully applied in recent years to a wide variety of tasks [19] [35].

However, the formulation of the task as pixel-wise classification, while particularly suited for detecting large objects with minimum circumference, is ill-posed for other tasks such as road or lane detection [6]. Similarly, for slender, sparse contrails, an alternative is required to minimize blotches created by naive architectures, as shown in Figure 3-6.

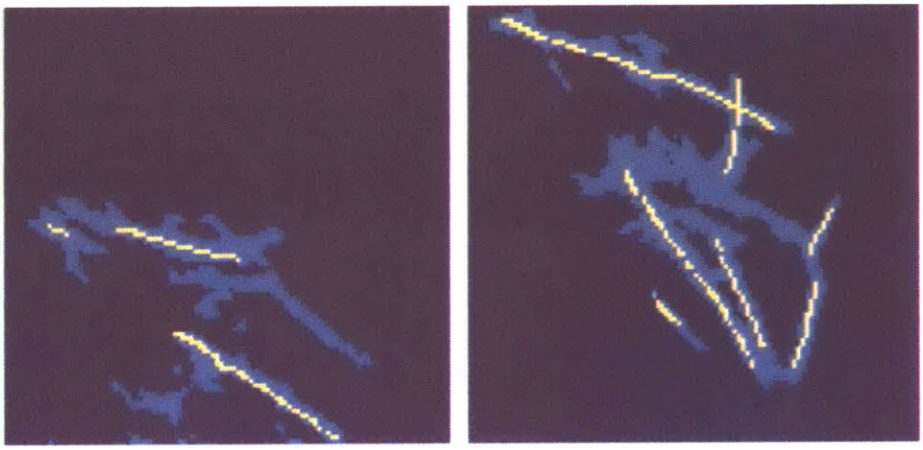

Figure 3-6: "Blotching" effect caused by neural network unable to learn the semantic structure of contrails through cross-entropy pixel loss. Human labels are shown in yellow, while positive detections of a naive U-Net architecture are shown in light blue.

Based on the ideas discussed in [6], the problem can be instead formulated as a generative adversarial network (GANs) to better preserve the high level structure of contrails, make the detections more realistic and eliminate post-processing. The

key idea is to use an "embedding loss" to add structure to the generator training, as shown in Figure 3-7. "Pixel-level loss cannot easily enforce high-level qualities such as thinness, straightness or the uniqueness of the detected line. The fundamental reason behind this is the way the training loss is formulated (e.g., per-pixel cross entropy), such that each output pixel in the segmentation map is evaluated independently of all others, i.e. no explicit inter-pixel consistency is enforced." [6] Intuitively, in this training setup, the discriminator becomes an adaptive content loss that's able to better capture a "human-like" semantic understanding of an image.

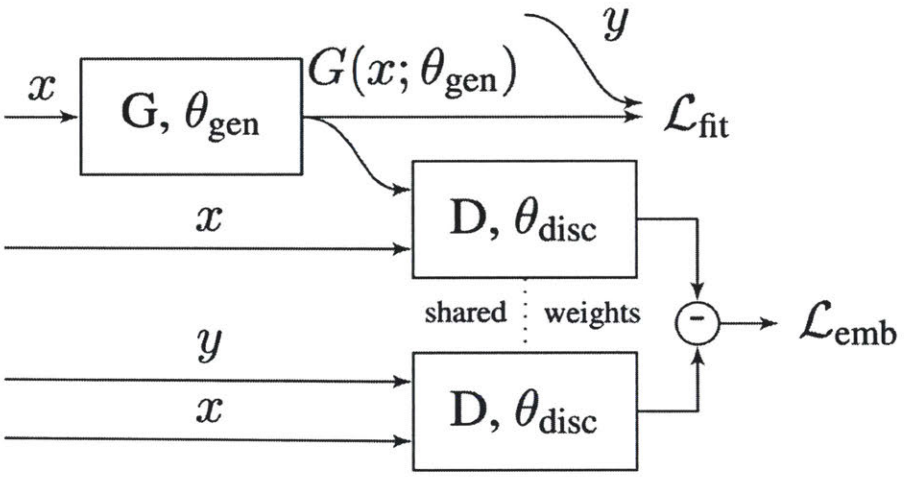

Figure 3-7: Generative adversarial network architecture

We use a two-stage training protocol, building on the network described in Section 3.1. We reuse the binary classifier as an encoder, tasked with generating a deep feature representation for each image. We then extend the network with a decoder designed to generate pixel-level predictions. The decoder upsamples the low-resolution, feature-rich representation using skip connections and trained weights. We train the complete model using pixel loss and adaptive perceptual loss [6] on 100 densely annotated images. During training, images are augmented by random cropping and flipping of 448×448 pixel patches. The generator is first pre-trained for 1000 epochs at a 0.04 learning rate with an exponential decay of 0.99, 0.9 momentum and 0.0001 weight decay. The discriminator is pre-trained for 400 epochs at 0.001 learning rate with an exponential decay of 0.99 and 0.0001 weight decay. The GAN is trained until

validation loss stops decreasing at 0.001 learning rate. We use 100:1 weighted cross entropy loss to balance the dataset.

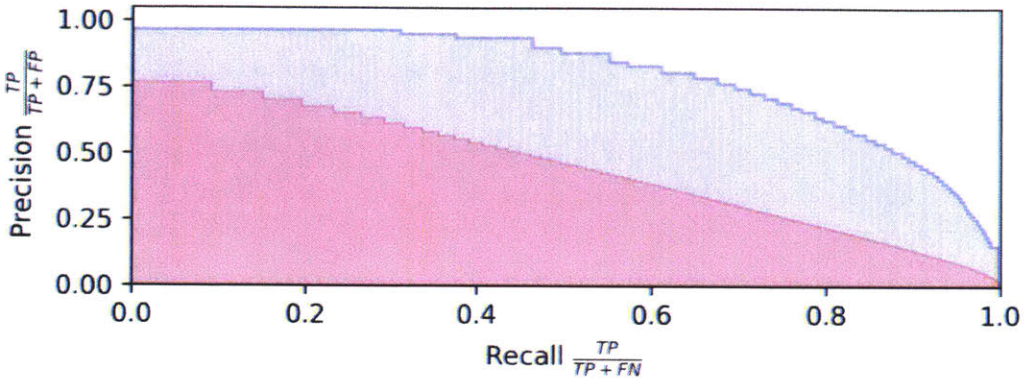

Figure 3-8: Precision-recall curve of the generator (red), relaxed precision-recall curve (blue)

We evaluate the performance of the trained generator on a validation set of 10 six-megapixel images using relaxed precision recall introduced by [26]. Relaxed precision recall introduces a buffer of ρ pixels within the range of pixels labeled as positives [45, 26]. Each pixel predicted as positive in this range is considered to be correctly classified. We set $\rho = 3$, which is consistent with existing literature [36, 26, 45].

The generator achieves 75% mean average precision and a smooth precision recall curve as shown in Figure 3-8. Our fully-convolutional neural network learns a multi-scale representation of contrails and does not impose constraints on input resolution. We also investigate the output contrail maps in Figure 3-9 and 3-10 - generator captures the majority of the area covered by contrails with high precision.

3.3 Inference and outlier filtering

We use Nvidia Quadro P6000 with 24 GB of vRAM for training and generating detections. We run neural network inference using two overlapping 1792 by 1792 pixel windows with a batch size of three. The inference windows are averaged and checked for discrepancies in the 584 pixel wide overlapping region. The deep learning model outputs probabilities that need to be thresholded to create a contrail map. We set the

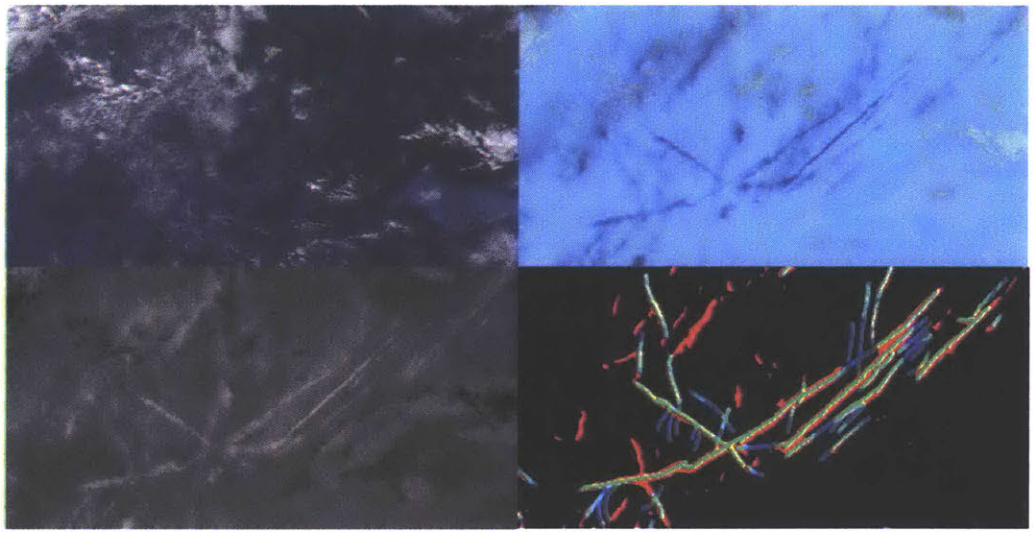

Figure 3-9: Sample automated detection on a validation set image #1. In a clockwise order: True Color, IR False Color, IR, automated contrail mask. Green indicates true positive, blue false negative and red false positive

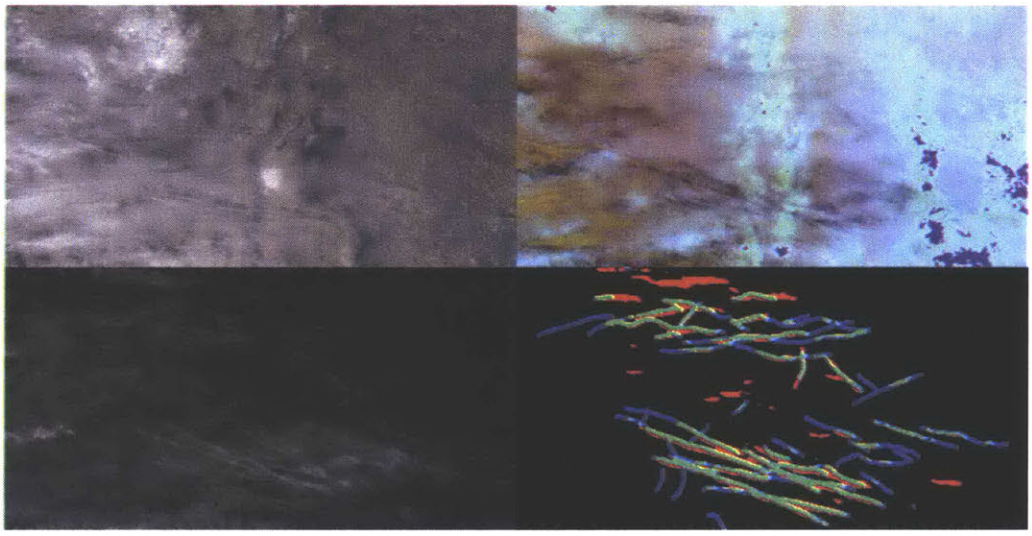

Figure 3-10: Sample automated detection on a validation set image #2. In a clockwise order: True Color, IR False Color, IR, automated contrail mask. Green indicates true positive, blue a false negative and red a false positive

threshold at 92% to minimize area coverage bias. We find the threshold by setting the relaxed precision equal to relaxed recall for all validation images, resulting in an approximately equal number of false positive and false negative pixels. We then apply an area-correction matrix to account for reprojection distortions and convert the pixel map to km^2/px coverage map. We split the resulting map in 11 channels representing

different geographical regions (Canada, United States, Mexico, Other Land, Ocean, Eastern Canada/Greenland/Iceland (SREX CGI), Western North America (SREX WNA), Central North America (SREX CNA), Eastern North America (SREX ENA), Central America/Mexico (SREX CAM) and aggregate their coverage in a timeseries.

For analysis, we remove images with artifacts based on GOES-East health-masks provided with the imagery - we reject a sample if any channel has over 0.005% bad pixels. Additionally, we remove 12 hours of data before and after all training samples to ensure no overfitted results. This represents respectively 4.5% and 10.4% of all samples.

Chapter 4

Results and discussion

We use the pipeline shown in Figure 4-1 to process 85,064 satellite images taken between March 1, 2017 and July 1, 2019. The total inference time is 30 hours, an estimated 2,800x improvement over human speed and 30x improvement over legacy algorithms (our implementation). A single human working continuously would need to spend 42 years on this task. We exclude all samples before July 1, 2017 because of high proportion of artifacts.

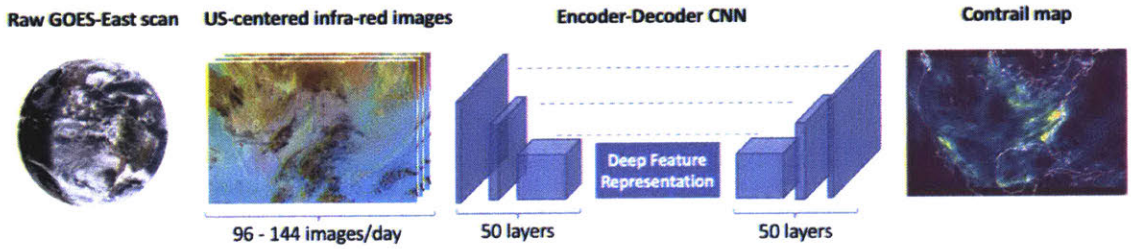

Figure 4-1: We use deep encoder-decoder CNN architecture for contrail classification. A raw GOES-East scan is retrieved online and reprojected to yield a 4-channel US-centered infra-red image (shown in false-color above). The infra-red channels are fed into the Encoder-Decoder Convolutional Neural Network. The CNN output is a 1:1 pixel-level map of contrail probabilities. By batching multiple images and thresholding the output, we obtain a prediction for the mean level of contrail coverage in km^2

Figure 4-1d shows the annual-average contrail coverage for the period July 2017 to July 2019, based on automatic analysis of 65,000 satellite images (390 billion pixels). Bulk metrics and trends in coverage are consistent with prior studies (0.5-1%

coverage over similar domain; winter maximum; summer minimum) [24, 31]. We find an average coverage of 0.55% of the area of the United States (excluding Hawaii and Alaska in all cases in this paper), equivalent to a quarter of the area of Florida. Over the entire domain, we also find that 35% of all contrail coverage occurs at night, while 38% is over ocean - conditions under which climate damages are expected to be maximized [40, 37].

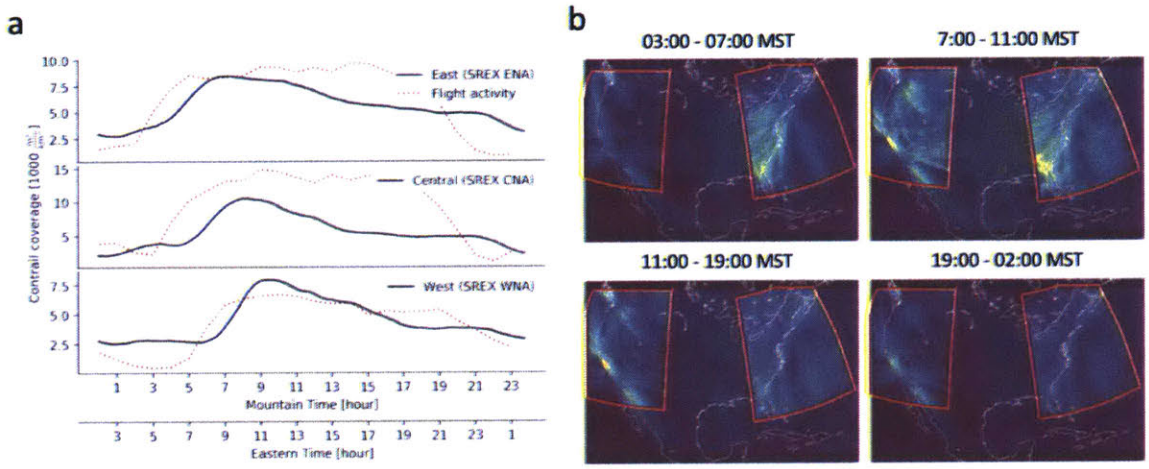

Figure 4-2: Mean diurnal contrail patterns averaged over two years. Left panel (a) shows contrail coverage in three regions - East, Central and West North America. The red dotted line represents flight activity scaled to 100,000 $\mathrm{m\,km^{-2}}$. Diurnal flight patterns are averaged using 2018 traffic data. Gray shading indicates the 95% confidence interval of the mean (1.96 standard deviations). Right panel (b) shows four averaging periods selected to highlight independent patterns of coverage at night, morning, afternoon, and late evening. The red outlines highlight the East and West regions. Contrail intensity is normalized on a scale from 0 (dark blue) to 2.5%. (bright yellow)

Thanks to the 10-15 minute temporal resolution of the GOES-East satellite data, we are also able to quantify sub-hourly trends in contrail coverage. Correlating these trends with air traffic volume data allows us to empirically quantify for the first time the relationship between contrail coverage and traffic volume. Figure 4-2 shows daily trends in coverage (averaged over the full observational period) using regions defined by the Intergovernmental Panel on Climate Change [28]. Contrail coverage follows a diurnal pattern, with the maximum coverage exceeding the minimum by 210% as shown in Figure 4-2a. Coverage peaks around 8-10 AM local time, reaching 8500,

10500 and 7900 $m^2 km^{-2}$ for East, Central and West regions respectively. This is followed by a gradual reduction in coverage, reaching a minimum at around 3-5 AM local time. Prior estimates based on low Earth orbit satellites made observations at 7:30 AM and 2:30 PM local time, meaning that they likely captured the maximum but did not resolve the daily minimum in coverage and may have overestimated total contrail impact as a consequence [31].

Contrail coverage starts increasing rapidly in all regions approximately two hours after the morning rise in traffic. This delay can be attributed to contrail formation and growth until they are large enough to appear in the imagery. Flight activity and rise in contrail coverage are offset by approximately three hours between East and West regions, corresponding to the time zone difference between Eastern and Pacific time. Average contrail coverage is comparable between East and Central at 5600 $m^2 km^{-2}$, with West slightly lower at 4400 $m^2 km^{-2}$. Central region reaches highest average peak coverage at 10500 $m^2 km^{-2}$, with maxima over Chicago and Dallas-Houston-Austin area (see Figure 4-2b).

While the initial morning contrail peak appears driven by air traffic, subsequent tail off in the afternoon is not correlated with flight activity, which remains relatively steady throughout the day. The drop in traffic in the evening is followed by an increase in the slope of contrail decay approximately 1-2 hours later. East and Central regions reach a distinct minimum in the early morning hours, while West region coverage stays relatively constant between 2500 - 2800 $m^2 km^{-2}$ throughout the night, despite dropping traffic. During the day contrail coverage averages respectively 60,000, 80,000 and 100,000 m^2 per air-kilometer for Central, East and West regions respectively. Apparent contrail persistence at night makes this ratio peak to 0.4, 0.6 and 0.6 $km^2 km^{-2}$ for the same regions.

Figure 4-2b shows contrail clusters visible over airports, with maxima over Seattle, Los Angeles, San Francisco, Chicago, Dallas, New York City, and Atlanta. High traffic routes are also visible, both along the original paths and as the contrails are advected. The pattern of the early morning rise up to a noon peak and gradual drop-off appears to be upheld in all regions. East and West regions coverage is driven by maxima over

major hubs, while Central coverage is more distributed. Central region also supports less contrail coverage, despite high overpass traffic.

Finally, we analyze the dependence of contrail coverage on season. Figure 4-3 shows total contrail coverage split by surface type throughout the year and by meteorological season. Overall seasonal variation in coverage is consistent with prior satellite- and ground-based studies [20, 31]. We observe a different seasonal trend for contrail coverage over ocean than over land. Contrail coverage over land is maximized in late summer and minimized in winter (80% decrease in coverage between January 2018 and August 2018). Peaks in over-ocean coverage correspond to peaks over-land, however the late summer/fall minimum lags that over land by approximately two months. Coverage over the Ocean, primarily Atlantic dominates the domain coverage in July 2017 - August 2017 and June 2018 - August 2018, which is also noticeable in the third panel of Figure 4-3b. These trends are not reflected in the traffic data, which exhibits the opposite trend (summer peak) and is characterized by low intra-annual variability of 25%.

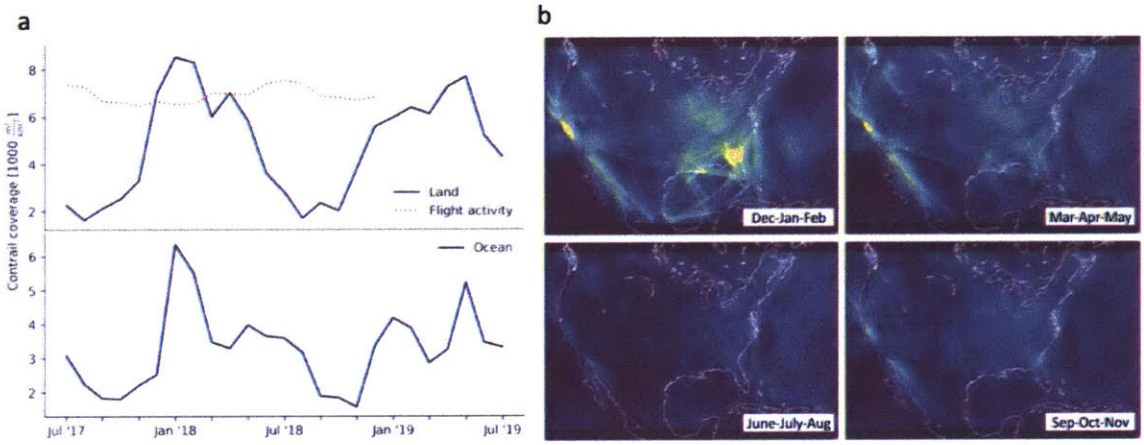

Figure 4-3: Mean contrail coverage throughout the year. Left panel (a) shows average monthly contrail coverage split by surface type. The red dotted line represents flight activity scaled to 100,000 m km^{-2}. Gray shading indicates the 95% confidence interval of the mean (1.96 standard deviations). Right panel (b) shows averaged contrail coverage maps by meteorological season. Contrail intensity is normalized on a scale from 0 (dark blue) to 2.5% (bright yellow).

We also observe significant interannual variability. Peak coverage in the first 12 month period occurs in January for both land and ocean. However, in the second 12

month period, coverage during January is 25% lower than in 2018, and peak coverage instead occurs during May 2019. This is not correlated with traffic, which is expected to be consistent between both periods.

Chapter 5

Limitations and future work

We discuss the limitations of our work and proposed next steps below.

5.1 Limitations

Key limitations of our approach revolve around assumptions of how various physical effects and dataset characteristics impact the predictions of the contrails detector. We make the following assumptions:

- The detector is not sensitive to area distortions, shape distortions, viewing angle, surface illumination angle and satellite distance effects
- Contrails in the initial stages of formation and spreading (before detection) do not meaningfully contribute to contrail coverage
- Contrails in the final stages of spatial dilution, when they lose contrast to the background and are not detected do not meaningfully contribute to contrail coverage
- Contrails are counted equally whether they do or do not overlap with other clouds
- Trained human labelers are able to correctly annotate all contrails visible in an image

- The detector's estimated probability of a contrail presence is independent of position and time of day, month or year.

5.2 Future work

Going forward, we see opportunities for expanding the work described in this thesis both in the context of contrail detection and other applications. In this section we present actionable next steps for contrail detection to address the issues outlined above, in addition to new research directions outside of contrail detection which are enabled by this work.

5.2.1 Contrail detection

Following from this work, we see three major areas for improvement in contrail detection - data quality and scope, science objectives and data availability. First, the current model can be expanded in terms of detection fidelity (quality of output) and scope by increasing the domain of detection, processing data from other satellites, and tracking individual contrails over time. Specific recommendations are listed below:

- Expand the domain to GOES-East full disk for coverage between 52° N and 52° S, enabling contrail detection for both North and South America
- Expand coverage to Himawari-8 and GOES-17 (GOES-West) for a maximum theoretical coverage of 292 ° (8° W to 60° E)
- Improve spatial resolution by combining GOES-16 high-frequency imagery with NOAA-20 high-resolution (0.4 km) imagery
- Expand the classifier to recognize individual contrail instances and track them over time
- Refine the classifier through finding incorrect or weak (low probability) predictions, re-annotating them and re-training the model (hard example mining)

These improvements will enable new scientific objectives to be reached, reducing the uncertainty in contrail lifetime, overlap with lower clouds, radiative forcing and impact of individual routes:

- Estimate the contrail lifetime by tracking individual contrail instances
- Estimate contrail radiative forcing by detecting contrail edges more accurately and using other satellite instruments for assistance
- Compare contrail modelling and observational results for cross-validation
- Run case studies of RF impact of individual aircraft or routes

Finally, improved data availability will make it easier for other researchers to build on top of this work without encountering excessive storage and compute requirements. The proposed improvements in this space are:

- Simplify access to contrail estimates: expose an API to request contrail detections for a particular date and time (or an aggregate) online
- Reach broader audience: collaborate with online visualization tools [23] to display real-time contrail detections
- Reduce the storage overhead of contrail detections: store contrail labels in a sparse format (e.g., polylines instead of pixel-level annotations)

5.2.2 Other applications

Machine learning approaches open a number of new avenues for Earth science. In addition to further work on contrail impact estimation, we have identified two major research directions which are natural extensions of this thesis.

Empirical estimation of ship track impacts

Commercial shipping is estimated to cause climate impacts which are comparable to aviation in magnitude, and expected to increase over coming decades. These impacts are in part offset by the low-altitude marine clouds which form in ship exhausts.

Unlike contrails, these clouds are thought to provide an overall global cooling effect. However, lack of coverage data makes these impacts uncertain, while some work has suggested that these ship tracks may actually have a smaller or even net warming impact due to a suppressive effect on natural clouds. This hypothesis will likely be tested in the near future when new rules from the UN International Maritime Organization significantly reduce the mean sulfur content of shipping fuel, from 3.5% sulfur to 0.5% sulfur. However, without direct observations, the actual impacts of this change will be difficult to assess.

We propose an extension to our contrail work, allowing us to automatically identify ship tracks from visual-spectrum satellite imagery. This will provide a direct, empirical estimate of low-altitude cloudiness resulting from ships. It will also allow us to directly quantify the effect of changes in shipping sulfur. By comparing estimated ship-attributable cloudiness before and after the implementation of new fuel sulfur requirements, we will be able to assess the direct impact on the production of line-shaped marine clouds. Finally, we will combine this information with existing data on ship movements. By calculating observations of total marine cloud coverage along the routes taken by ships, comparing this before and after the implementation of ship sulfur regulation, and quantifying the contribution of machine-identified line-shaped ship tracks, we will be able to empirically evaluate whether ship sulfur increases or decreases the surface albedo of the ocean.

Tracking of intercontinental pollution plumes

Recent work on air quality has begun to acknowledge the role that remote sources have in dictating local air quality. Elevated ozone episodes in California have been attributed to emissions from Asia, while European air quality is adversely impacted by outflow from North America. However, atmospheric modeling techniques are unable to accurately reproduce these flows, preventing accurate, automated attribution of sources and complicating regulatory efforts. Automated identification and tracking of these plumes would permit two major advances. Firstly, it would provide an empirical mechanism to quantify the rates of atmospheric diffusion experienced by

these plumes, critical to improving the aforementioned atmospheric models. Secondly, it would enable the routine, quantitative attribution of pollution events to remote sources. Our proposed approach is to use a convolutional neural network to identify new pollution plumes, and to track the migration and evolution of these plumes over the days and weeks which follow. The neural network would utilize information from the observation time series to identify physically realistic movement of pollution. This will provide an ongoing, global, empirical estimate of the impact of remote pollution sources on remote air quality. This will provide actionable information on pollution impacts. Furthermore, by directly observing the rate of increase of polluted area, the effective diffusion rate of plumes in different locations and seasons can be determined, providing physical information which can be used to improve global modeling techniques.

Chapter 6

Conclusion

In this thesis, we demonstrate a novel approach for detecting contrails using deep learning. We show that the previously posed challenges in data collection and training time are surmountable and provide a near real-time contrail coverage over the US with minimal manual parameter tuning and post-processing. Our results empirically demonstrate for the first time how air traffic affects cloud cover on a sub-hourly basis. This analysis provides new insight into contrail coverage patterns across the US and the relationship between contrail formation and aircraft traffic.

Using our automatic detection algorithm on GOES-East data, we exceed the temporal resolution of the LEO satellite [31] and scalability and spatial resolution of the ground observations [20]. Our fully-convolutional neural network learns a multi-scale representation of contrails and does not impose constraints on input resolution. We expect our approach to improve further once sub-km resolution geostationary infrared imagers become available.

We find that contrails cover an average of 0.55% of the contiguous United States, an area equivalent to a quarter of the size of Florida. Over the entire domain, we also find that 35% of all contrail coverage occurs at night, while 38% is over ocean - conditions under which climate damages are expected to be maximized [40, 37]. Average contrail coverage is comparable between US East and Central regions at 5600 $m^2\,km^{-2}$, with West region slightly lower at 4400 $m^2\,km^{-2}$.

Coverage peaks around 8-10 AM local time, reaching 8500, 10500 and 7900 $m^2\,km^{-2}$

for East, Central and West regions respectively. This is followed by a gradual reduction in coverage, reaching a minimum at around 3-5 AM local time. Contrail coverage over land is maximized in late summer and minimized in winter (80% decrease in coverage between January 2018 and August 2018).

The dataset produced here is freely available on request, constituting the first open-source, temporally- and spatially-resolved, publicly available dataset of contrail coverage. We have also made the network architecture and weights required to process future GOES-East data available for download on GitHub, allowing researchers to directly apply this technique without the need for additional tuning.

To date, studies estimating contrail coverage over the US relied on low earth orbit satellites with low temporal resolution (1-2 measurements per day) and retrieval and processing latencies measured in days [21, 39, 31]. The processed data and algorithms used were siloed and transferring them to new satellite datasets required manual tuning. With our algorithm, we demonstrate for the first time that contrail detection algorithms can be fast, generalizable, and adaptive. We demonstrate a 30x improvement in contrail detection speed over legacy algorithm [21], allowing to process a full year of satellite observations in under 15 hours on a single consumer GPU. Combined with the satellite downlink latency below 5 min, this method can deliver a near real-time view of the contrail-covered sky over the US.

We expect on-demand, cloud-based deep learning detectors to be the key to making the petabytes of environmental data more accessible to atmospheric scientists. The findings of this study, which resolve the relationship between individual flight tracks and contrail formation at an unprecedented spatial and temporal scale, will directly enable reductions in aviation's climate impacts. The model presented here could be the first step to optimize flight routes and flight times for minimum climate impact, and analyze the effectiveness of aircraft convoys aiming to avoid forming contrail cirrus.

Appendix A

Distortion maps

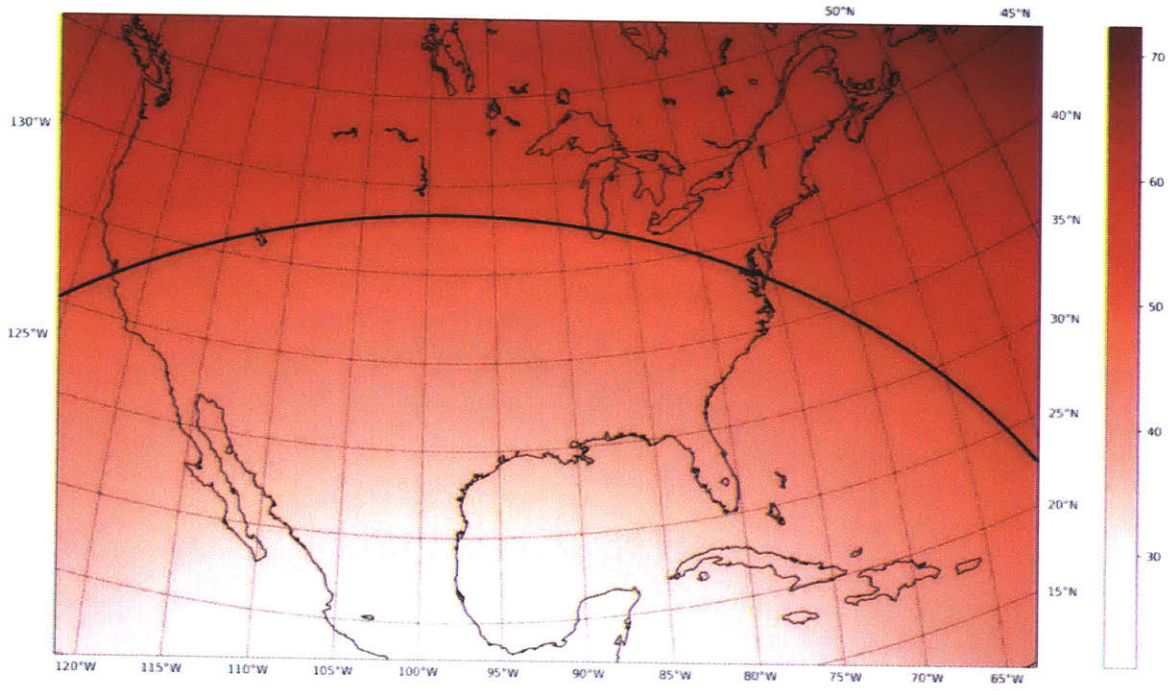

Figure A-1: Viewing angle in degrees for the contiguous US domain. Black contour line delineates 50° viewing angle boundary for reference. Angle is calculated using spherical approximations.

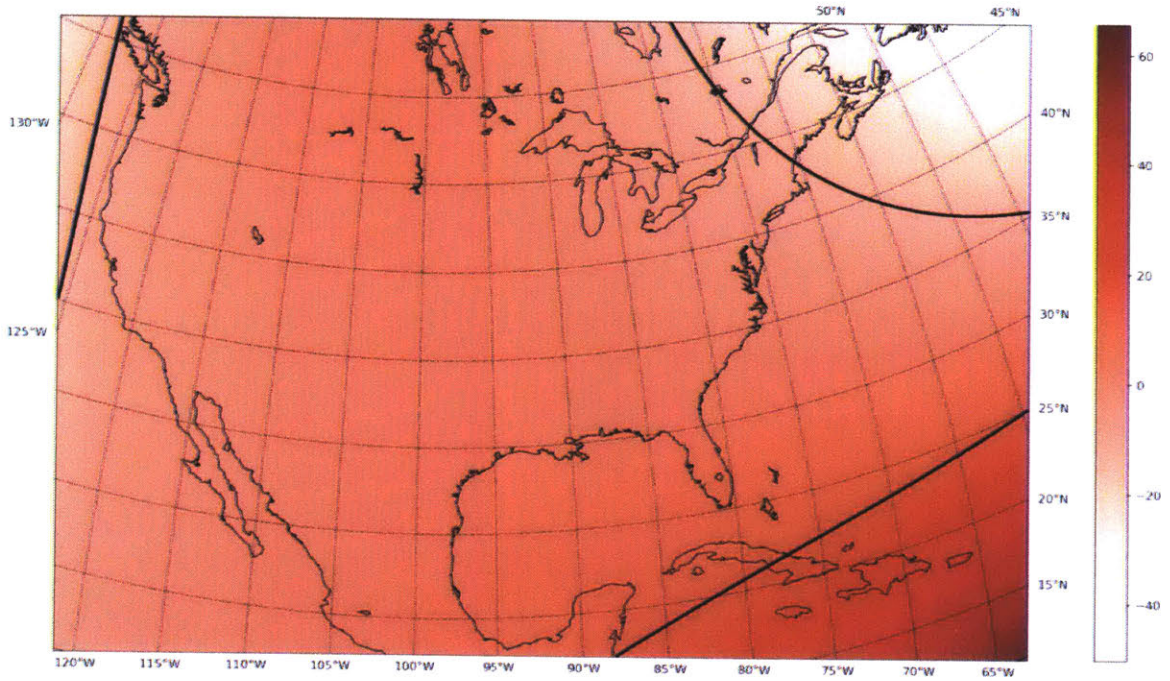

Figure A-2: Area distortion in % for the contiguous US domain. Black contour line delineates 15% distortion boundary for reference. Base area is 4 km^2

Appendix B

Labelling guidelines

The following information was shared with all data labellers, along with examples of correctly annotated images.

Introduction

Condensation trails (contrails) are line-shaped clouds produced by aircraft engine exhaust. They are primarily composed of condensed water, which freezes after exiting the engine. You can sometimes observe them from the ground forming behind an aircraft as shown in Figure B-1. Contrails are usually short-lived (less than 30 minutes), but in suitable atmospheric conditions, they can persist for up to 12 hours. We are interested in studying their formation, coverage and evolution because of their impact on Earth's climate.

More specifically, the purpose of labeling contrails on satellite images is to produce the first-ever dataset of labelled contrail images captured using a geostationary satellite. This dataset will then be used to train an algorithm, which will be able to detect contrails in near-real time. It is critical to collect very high-quality data to allow the machine learning algorithm to learn the correct image features. Your job will be to help annotate the satellite images - this document outlines how.

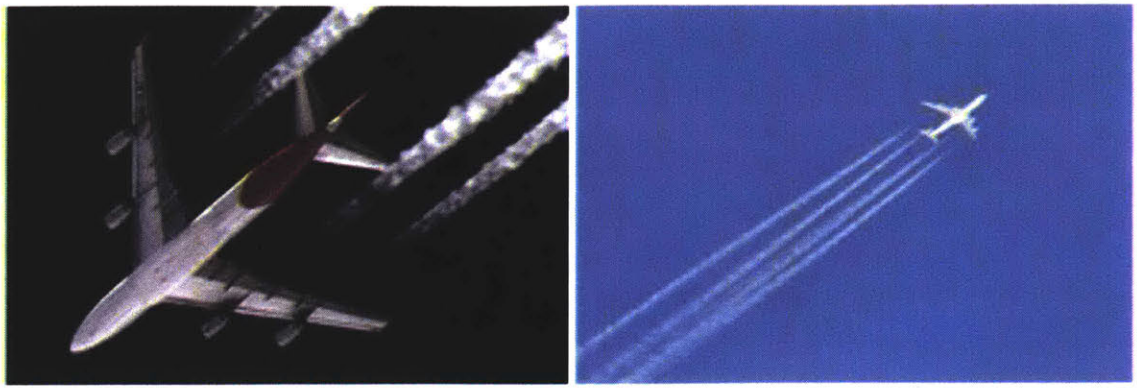

Figure B-1: Contrails forming behind aircraft as seen from the ground

Background

A geostationary satellite orbits the Earth at an altitude of over 35,000 km always pointing at the same location on the ground. It collects light in 16 frequency bands - some of the examples are human-visible "RGB" spectrum (Figure B-2, left) or infrared spectrum (Figure B-2, right). These bands capture different physical phenomena ranging from vegetation to cloud particle size.

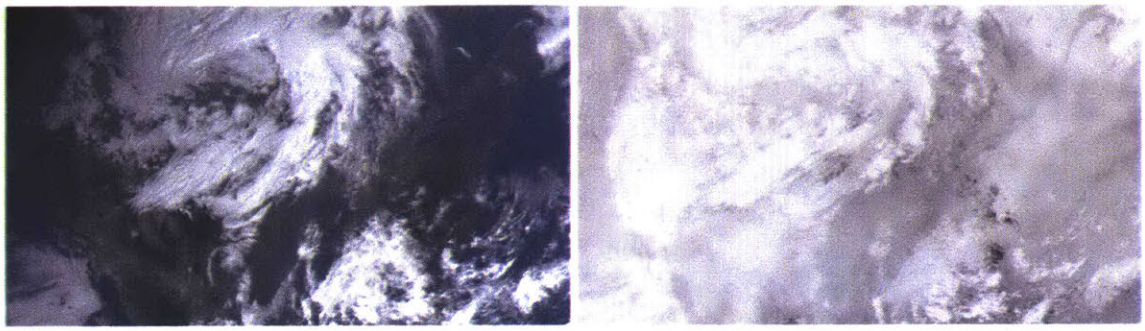

Figure B-2: North America as seen from a geostationary satellite in RGB (left) and infrared (right)

When contrails form behind an aircraft, their initial span is on the order of the wingspan of the aircraft (approx. 50 m). However, if weather conditions let the contrail persist, it will diffuse and grow in width. After reaching a critical width of a few hundred meters, it is possible to spot it on a satellite image. Figure 3C shows an area above Los Angeles in RGB with color-coded contrails as seen from the ground (Figure B-3B).

Figure B-3: Area above Los Angeles as seen from above (A) and below (B) the atmosphere with color coded contrails (B and C) [contrailscience.com]

We note that contrails only appear in region, where the weather conditions are favorable for their formation and persistence. As such, you cannot expect a contrail to form throughout the entire flight path of an aircraft, but only in selected spots.

A particular feature of contrails is that they absorb infrared light emitted from Earth's surface, causing them to appear darker on an infrared image (Fig. B-4). Most of the objects at room temperature radiate their heat away as infrared light. The Earth also "glows" in infrared light, invisible to the human eye, because of its temperature.

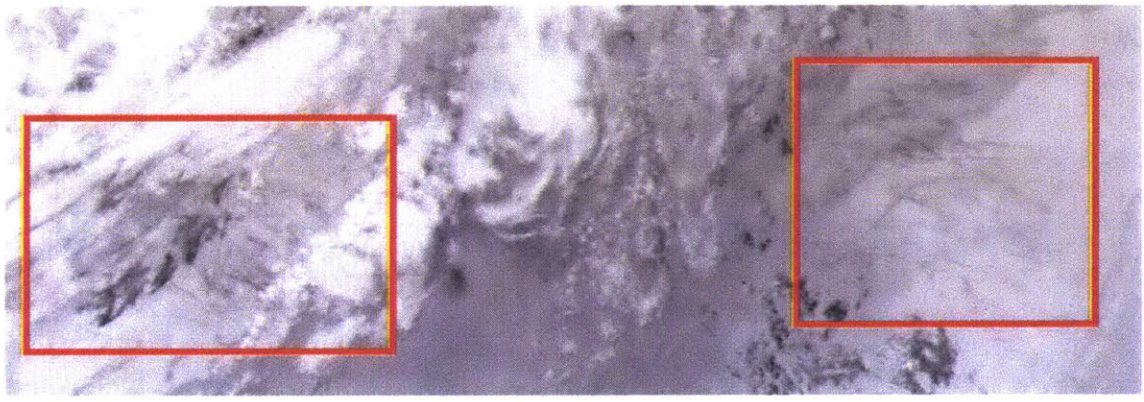

Figure B-4: Formation of condensation trails on an infrared image, note the slender dark grey streaks as highlighted in red boxes, which can be considered contrails

To make it easier for humans to see contrails, we can combine three infrared channels into a single false-color image, using the so-called "ash transformation" (Fig. B-5). In an ash-transformed image, contrails and their natural counterparts - cirrus clouds - appear in dark blue/indigo color. The ash transformation will most often

show thick, cold ice clouds in dark red and thin cirrus clouds in dark blue to black. Mid-level water clouds are yellow or brown.

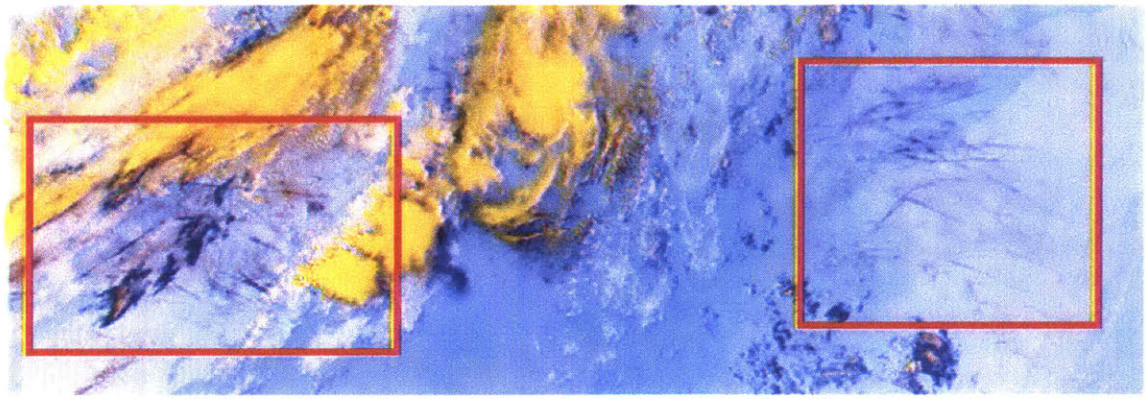

Figure B-5: Formation of condensation trails on a composite image consisting of three infrared channels, note how contrails appear as slender dark blue streaks as highlighted in red boxes

There are three key difficulties in detecting contrails on ash-transformed satellite images:

- Contrails appear almost identical in color to natural cirrus clouds (examples to follow)
- Contrails lose their linear shape over time, diffused by the movement of air at high altitudes
- Contrails typical width barely exceeds the detection threshold of best-in-class geostationary satellites
- Static land features can sometimes be mistaken for contrails

To mitigate these difficulties, we will restrict our annotations to contrails which maintain approximately linear shape and are wide enough to be detected by our geostationary satellite. A number of example snippets with and without contrails under this definition are shown in Figure B-6.

Distinguishing between contrails and natural cirrus clouds is a bit more difficult and requires some practice and judgement. A helpful heuristic is to remember that

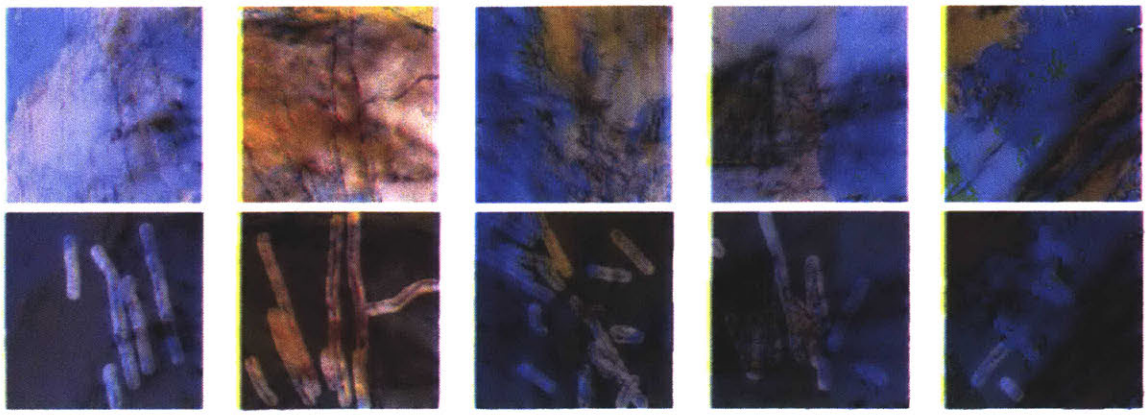

Figure B-6: Ash-transformed satellite images snippets with contrail formations (top). Highlighted contrail features (bottom)

contrails are created by an artificial process. The speed of an aircraft is orders of magnitude larger than the rate of atmospheric diffusion, resulting in "unnatural-looking" extremely slender clouds. Often, because of overlapping flightpaths, multiple contrails will be approximately aligned in the same direction or perpendicular to each other in a given air patch - see Figure B-7.

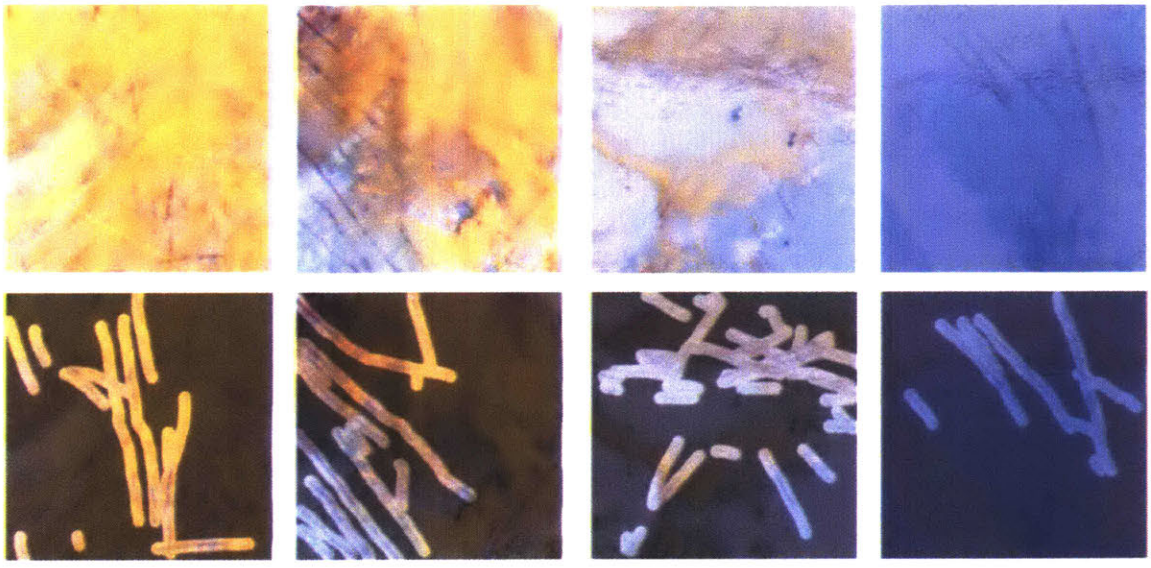

Figure B-7: Additional examples of ash-transformed images of contrail formations - from left to right: occluded parallel contrails; partially occluded parallel contrails; partially occluded cross-hatched contrails; lightly diffused approximately parallel contrails. Original images (top); Highlighted contrail features (bottom)

Because the conditions required for formation of cirrus clouds and contrails are

very similar, both phenomena commonly occur together. This further increases the difficulties for distinguishing between contrail and those natural clouds. Their shapes and surrounding cloudiness provide some guidance though. A number of cases showing cirrus clouds (not contrails) are shown in Figure B-8 for reference, alongside an example which shows overlapping cirrus and contrails (Figure B-8C).

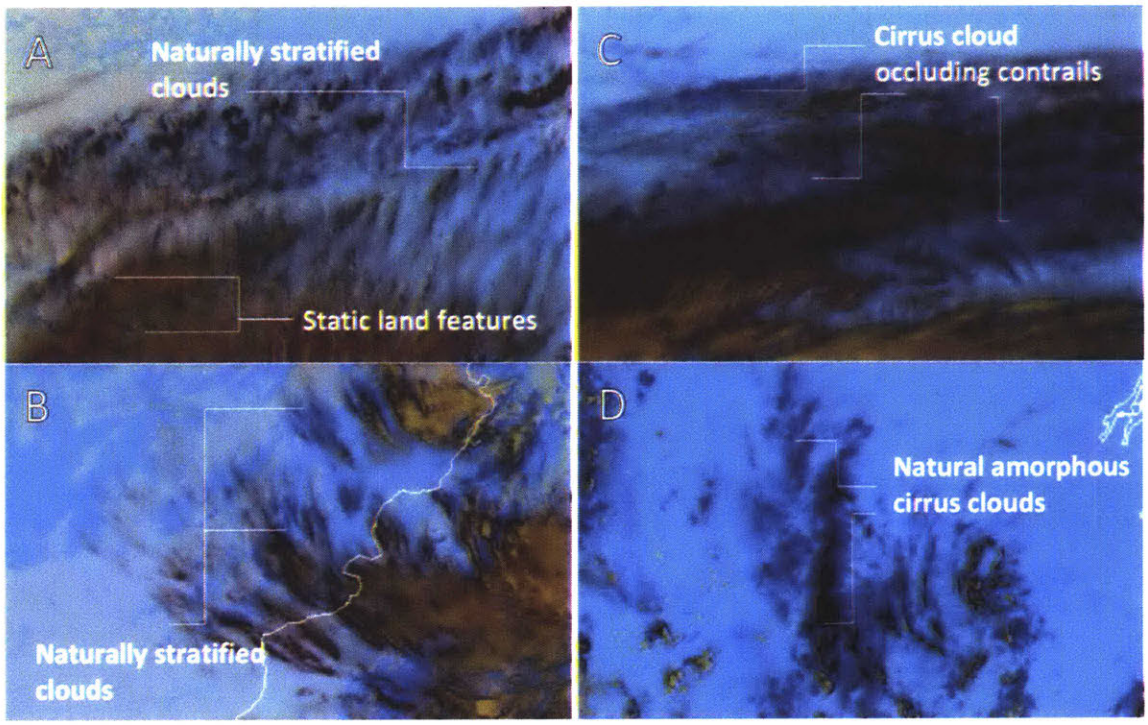

Figure B-8: Examples of cirrus clouds which have similar IR signature (color) to contrails - A and B: natural linear-like formations; C: contrails covered in a diffused cirrus cloud; D: large cirrus cloud

Finally, to help reject static land features you will be provided with two "before" and "after" satellite images that will show the movement of clouds.

Labelling process

Your task will be to take preprocessed ash-transformed satellite images and carefully annotate them on a pixel-by-pixel basis. The expected end goal is a "contrail mask" accompanying each image. It is expected that each image will take from 45 to 75 minutes to densely annotate (depending on contrail coverage). Figure B-9 contains a

number of examples of correct contrail annotations.

Please see additional attachments for high definition examples of complete contrail masks.

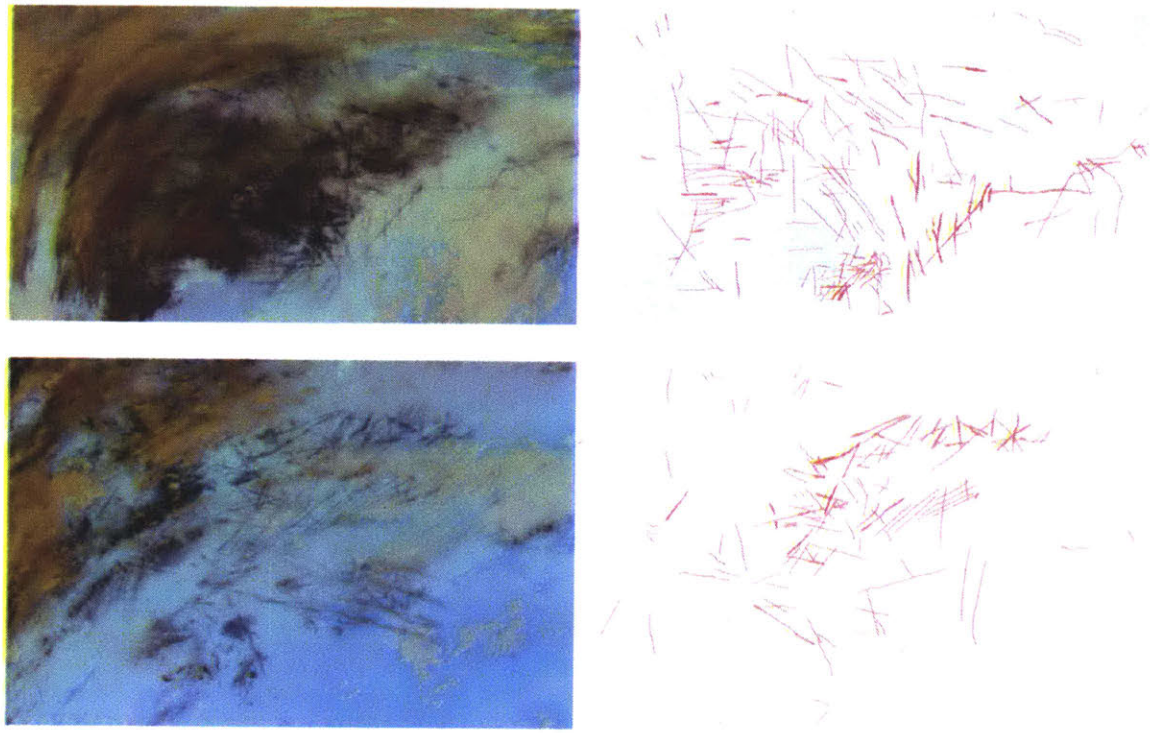

Figure B-9: Examples of densely annotated satellite images

Bibliography

[1] Seyed Majid Azimi, Peter Fischer, Marco Körner, and Peter Reinartz. Aerial LaneNet: Lane marking semantic segmentation in aerial imagery using Wavelet-Enhanced cost-sensitive symmetric fully convolutional neural networks. March 2018.

[2] Ulrike Burkhardt and Bernd Kärcher. Global radiative forcing from contrail cirrus. *Nat. Clim. Chang.*, 1:54, March 2011.

[3] Ulrike Burkhardt, Bernd Kärcher, and Ulrich Schumann. Global modeling of the contrail and contrail cirrus climate impact. *Bull. Am. Meteorol. Soc.*, 91(4):479–484, April 2010.

[4] Ni-Bin Chang, Kaixu Bai, and Chi-Farn Chen. Integrating multisensor satellite data merging and image reconstruction in support of machine learning for better water quality management. *J. Environ. Manage.*, 201:227–240, October 2017.

[5] Kevin Fuell, Brian Guyer, Deirdre Kann, Andrew Molthan, and Nicholas Elmer. Next generation satellite RGB dust imagery leads to op-erational changes at NWS alburquerque. *J. Operational Meteor.*, 04(06):75–91, March 2016.

[6] Mohsen Ghafoorian, Cedric Nugteren, Nóra Baka, Olaf Booij, and Michael Hofmann. EL-GAN: Embedding loss driven generative adversarial networks for lane detection. June 2018.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *Computer Vision – ECCV 2016*, pages 630–645. Springer International Publishing, 2016.

[8] Andrew Heymsfield, Darrel Baumgardner, Paul DeMott, Piers Forster, Klaus Gierens, and Bernd Kärcher. Contrail microphysics. *Bull. Am. Meteorol. Soc.*, 91(4):465–472, April 2010.

[9] Seunghoon Hong, Junhyuk Oh, Bohyung Han, and Honglak Lee. Learning transferrable knowledge for semantic segmentation with deep convolutional neural network. December 2015.

[10] IATA. IATA forecast predicts 8.2 billion air travelers in 2037. `https://www.iata.org/pressroom/pr/pages/2018-10-24-02.aspx`. Accessed: 2019-6-10.

[11] Neal Jean, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell, and Stefano Ermon. Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301):790–794, August 2016.

[12] Bernd Kärcher. Formation and radiative forcing of contrail cirrus. *Nat. Commun.*, 9(1):1824, May 2018.

[13] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet classification with deep convolutional neural networks. In F Pereira, C J C Burges, L Bottou, and K Q Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[14] David S Lee, David W Fahey, Piers M Forster, Peter J Newton, Ron C N Wit, Ling L Lim, Bethan Owen, and Robert Sausen. Aviation and global climate change in the 21st century. *Atmos. Environ.*, 43(22):3520–3537, July 2009.

[15] J Lee, R C Weger, S K Sengupta, and R M Welch. A neural network approach to cloud classification. *IEEE Trans. Geosci. Remote Sens.*, 28(5):846–855, September 1990.

[16] Sanggyun Lee, Jungho Im, Jinwoo Kim, Miae Kim, Minso Shin, Hyun-Cheol Kim, and Lindi J Quackenbush. Arctic sea ice thickness estimation from CryoSat-2 satellite data using machine Learning-Based lead detection. *Remote Sensing*, 8(9):698, August 2016.

[17] X Li, Weitao Chen, X Cheng, and L Wang. A comparison of machine learning algorithms for mapping of complex surface-mined and agricultural landscapes using ZiYuan-3 stereo satellite imagery. *Remote sensing*, 8(6), 2016.

[18] David Lister, David J Griggs, Mack McFarland, and David J Dokken. *Aviation and the Global Atmosphere: A Special Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, June 1999.

[19] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[20] H Mannstein, A Brömser, and L Bugliaro. Ground-based observations for the validation of contrails and cirrus detection in satellite imagery. *Atmospheric Measurement Techniques*, 3(3):655–669, June 2010.

[21] Hermann Mannstein, Richard Meyer, and Peter Wendling. Operational detection of contrails from NOAA-AVHRR-data. *Int. J. Remote Sens.*, 20(8):1641–1660, January 1999.

[22] Dieter Meinert. Training a neural network to detect jet contrails in satellite images. *Impact of emissions from aircraft and spacecraft upon the atmosphere*, pages 401–406, 1992.

[23] Kevin Micke. Every pixel of goes-17 imagery at your fingertips. *Bulletin of the American Meteorological Society*, 99(11):2217–2219, 2018.

[24] P Minnis, L Nguyen, D P Garber, D P Duda, and others. Simulation of contrail coverage over the USA missed during the air traffic shutdown. *European*, 2003.

[25] Patrick Minnis, Rabindra Palikonda, Bryan J Walter, J Kirk Ayers, and Hermann Mannstein. Contrail properties over the eastern north pacific from AVHRR data. *Meteorologische Zeitschrift*, 14(4):515–523, September 2005.

[26] Volodymyr Mnih and Geoffrey E Hinton. Learning to detect roads in High-Resolution aerial images, 2010.

[27] K J Murphy. NASA's earth science data systems overview and evolution. volume 2018, December 2018.

[28] V Murray and K L Ebi. IPCC special report on managing the risks of extreme events and disasters to advance climate change adaptation (SREX). 2012.

[29] D. Shindell F.-M. Breon W. Collins J. Fuglestvedt J. Huang D. Koch J.-F. Lamarque D. Lee B. Mendoza T. Nakajima A. Robock G. Stephens T. Takemura Myhre, G. and H. Zhang. *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change: Anthropogenic and Natural Radiative Forcing.* Cambridge University Press, 2014.

[30] Rabindra Palikonda, Patrick Minnis, P K Costulis, and David P Duda. Contrail climatology over the USA from MODIS and AVHRR data. In *Proc. 10th Conf. on Aviation, Range, and Aerospace Meteorology*, pages 13–16, 2002.

[31] Rabindra Palikonda, Patrick Minnis, David P Duda, and Hermann Mannstein. Contrail coverage derived from 2001 AVHRR data over the continental united states of america and surrounding areas. *Meteorologische Zeitschrift*, 14(4):525–536, September 2005.

[32] Douglas J Parker and Mariane Diop-Kane. *Meteorology of Tropical West Africa: The Forecasters' Handbook.* John Wiley & Sons, February 2017.

[33] Pedro O Pinheiro and Ronan Collobert. From image-level to pixel-level labeling with convolutional networks. November 2014.

[34] Martin Raspaud, David Hoese, and Panu Lahtinen et al. pytroll/satpy: Version 0.16.0, June 2019.

[35] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241. Springer International Publishing, 2015.

[36] Shunta Saito, Takayoshi Yamashita, and Yoshimitsu Aoki. Multiple object extraction from aerial imagery with convolutional neural networks. *Electronic Imaging*, 2016(10):1–9, 2016.

[37] U Schumann, B Mayer, K Graf, and H Mannstein. A parametric radiative forcing model for contrail cirrus. *J. Appl. Meteorol. Climatol.*, 51(7):1391–1406, July 2012.

[38] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-CAM: Visual explanations from deep networks via gradient-based localization. October 2016.

[39] Douglas A Spangenberg, Patrick Minnis, Sarah T Bedka, Rabindra Palikonda, David P Duda, and Fred G Rose. Contrail radiative forcing over the northern hemisphere from 2006 aqua MODIS data. *Geophys. Res. Lett.*, 40(3):595–600, February 2013.

[40] Nicola Stuber, Piers Forster, Gaby Rädel, and Keith Shine. The importance of the diurnal and annual cycle of air traffic for contrail radiative forcing. *Nature*, 441(7095):864–867, June 2006.

[41] M Vázquez-Navarro, H Mannstein, and S Kox. Contrail life cycle and properties from 1 year of MSG/SEVIRI rapid-scan images. *Atmos. Chem. Phys.*, 15(15):8739–8749, August 2015.

[42] M Vazquez-Navarro, H Mannstein, and B Mayer. An automatic contrail tracking algorithm. *Atmospheric Measurement Techniques*, 3(4):1089–1101, August 2010.

[43] Margarita R Vazquez Navarro. *Life cycle of contrails from a time series of geostationary satellite images.* PhD thesis, November 2009.

[44] Jinglin Zhang, Pu Liu, Feng Zhang, and Qianqian Song. CloudNet: Ground-Based cloud classification with deep convolutional neural network. *Geophys. Res. Lett.*, 33:909, August 2018.

[45] Z Zhang, Q Liu, and Y Wang. Road extraction by deep residual U-Net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, May 2018.