Descartes Labs

Lessons from industry

Manuel Weber
Applied Scientist (Machine Learning)
manuel@descarteslabs.com

DANCE ML Workshop 2020
07/08/2020

Overview

- Descartes Labs
- Common challenges
- Some solutions
  - Contrastive Sensor Fusion (CSF)
  - Neural Architecture Search (NAS)
- Conclusion / Questions

What we do at Descartes Labs

Descartes Labs has built a data refinery platform specifically for geospatial data to provide instant access to science-ready data.

High quality earth observations have been conducted continuously since the 1970s (first Landsat satellite) with a variety of sensors resulting in one of the largest data archive. Processing and extracting valuable information from each pixel is both technologically and computationally challenging.

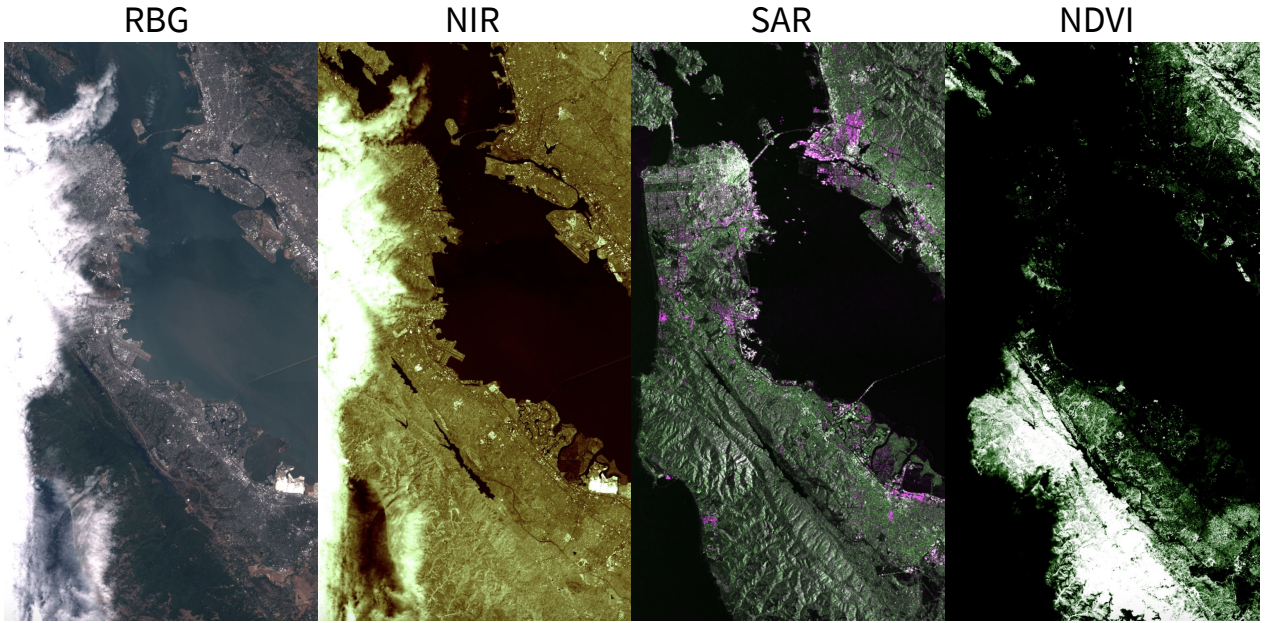
The DL platform gives researchers access to >15 PB of data to understand complex processes on earth by combining multiple sensors and multiband imagery.

Companies benefit from derived intelligence to understand the mechanisms that affect their business and are able to insightful decisions

What we do at Descartes Labs

As applied scientist we tackle customer and research projects in a variety of areas. In many cases Machine Learning algorithms are used to develop models that can be deployed at scale with the DL earth tiling system and cloud supercomputers.

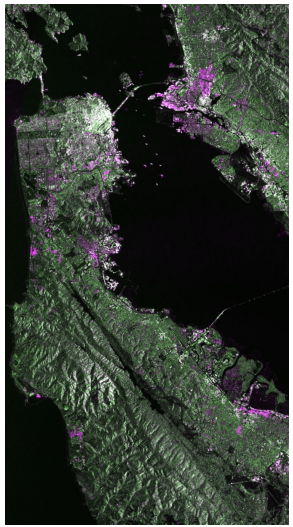
Agriculture
Oil & Gas
Mining & Metals
Power & Renewables
Shipping & Logistics
Financial Services & Insurance
Geospatial
Environmental Impact

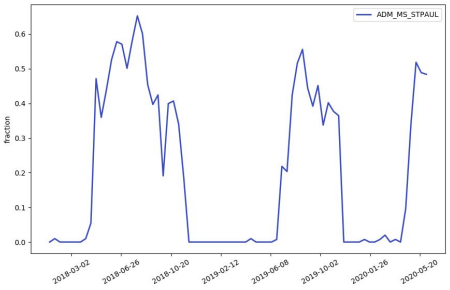


What we do at Descartes Labs

Many of the problems come down to understanding the value of pixels and their correlation among each other, across spectral bands and across sensors. Extracting this information then allows to figure out the complex processes underlying the problem.

SAR → Temporal information → Monitor





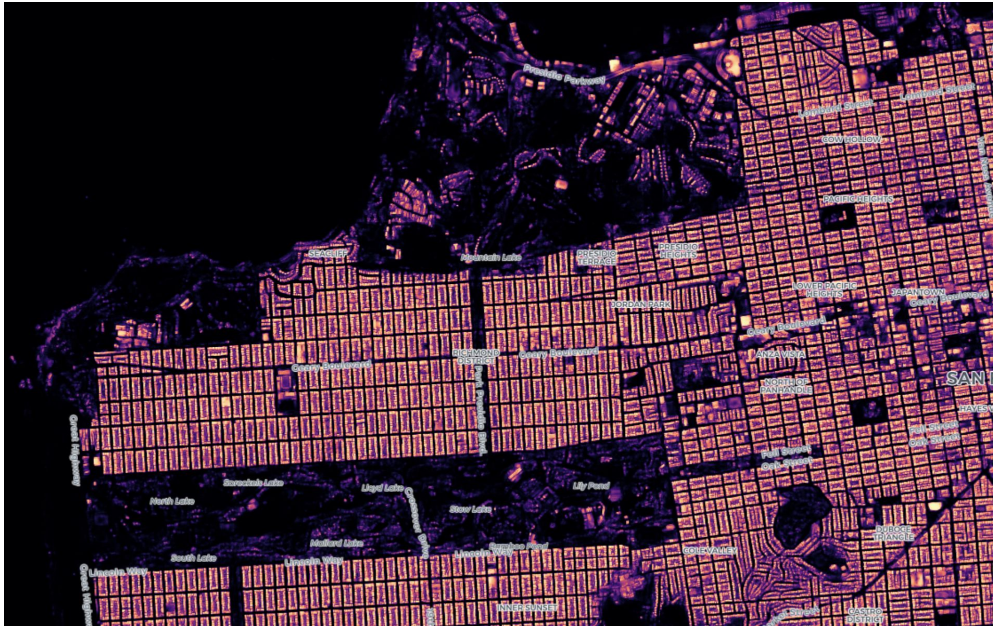
What we do at Descartes Labs

In most cases the data source are satellite images and scene understanding often involves object detection, image segmentation or instance segmentation. We use state of the art computer vision algorithms to accomplish this.

UNet building detection

What we do at Descartes Labs

In most cases the data source are satellite images and scene understanding often involves object detection, image segmentation or instance segmentation. We use state of the art computer vision algorithms to accomplish this.



UNet building detection

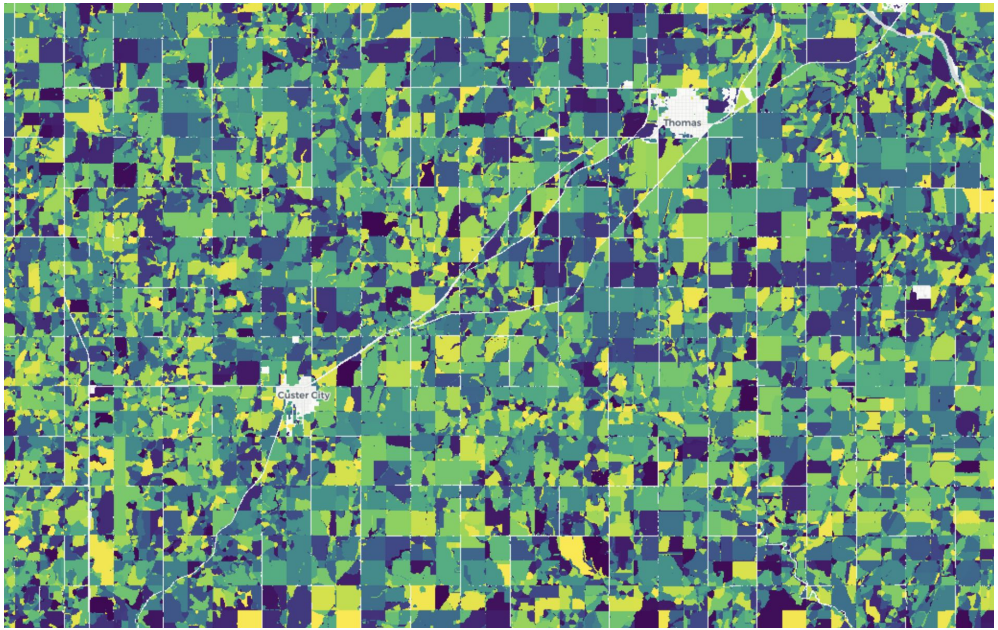
What we do at Descartes Labs

In most cases the data source are satellite images and scene understanding often involves object detection, image segmentation or instance segmentation. We use state of the art computer vision algorithms to accomplish this.

Field segmentation and crop identification, e.g. Mask-RCNN

What we do at Descartes Labs

In most cases the data source are satellite images and scene understanding often involves object detection, image segmentation or instance segmentation. We use state of the art computer vision algorithms to accomplish this.



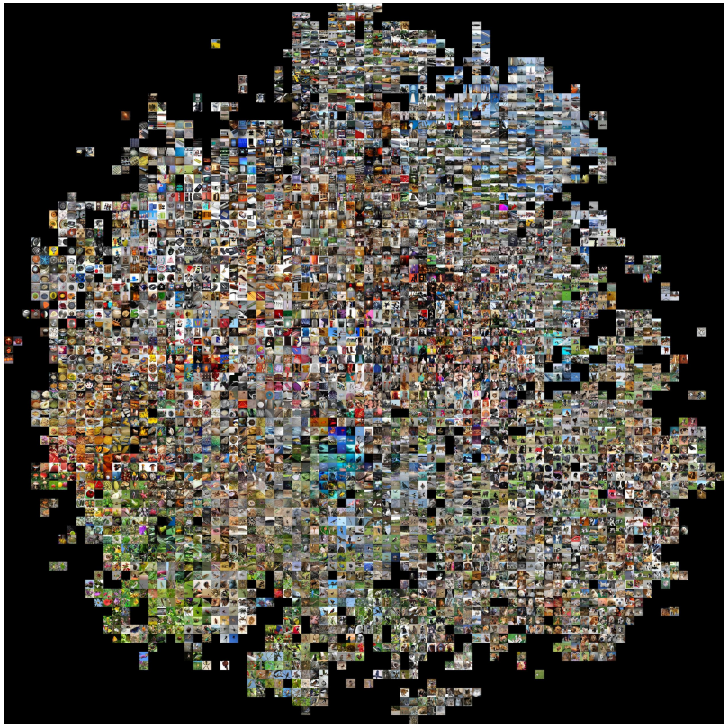
Field segmentation and crop identification, e.g. Mask-RCNN

Common issues that may apply to particle physics

Off the shelf computer vision deep learning models have O(M) parameters so you either need lots of labeled data or use pre-trained weights.

In remote sensing there is a lot of data but a very small fraction of it is labeled and it can be very expensive to get labeled data. Similar challenges exist in particle physics.

Over the years massive collections of natural imagery (ImageNet, COCO, etc.) have been created and algorithms with increasing accuracy are being developed. But most pre-trained architectures use 3 band (RGB) inputs.



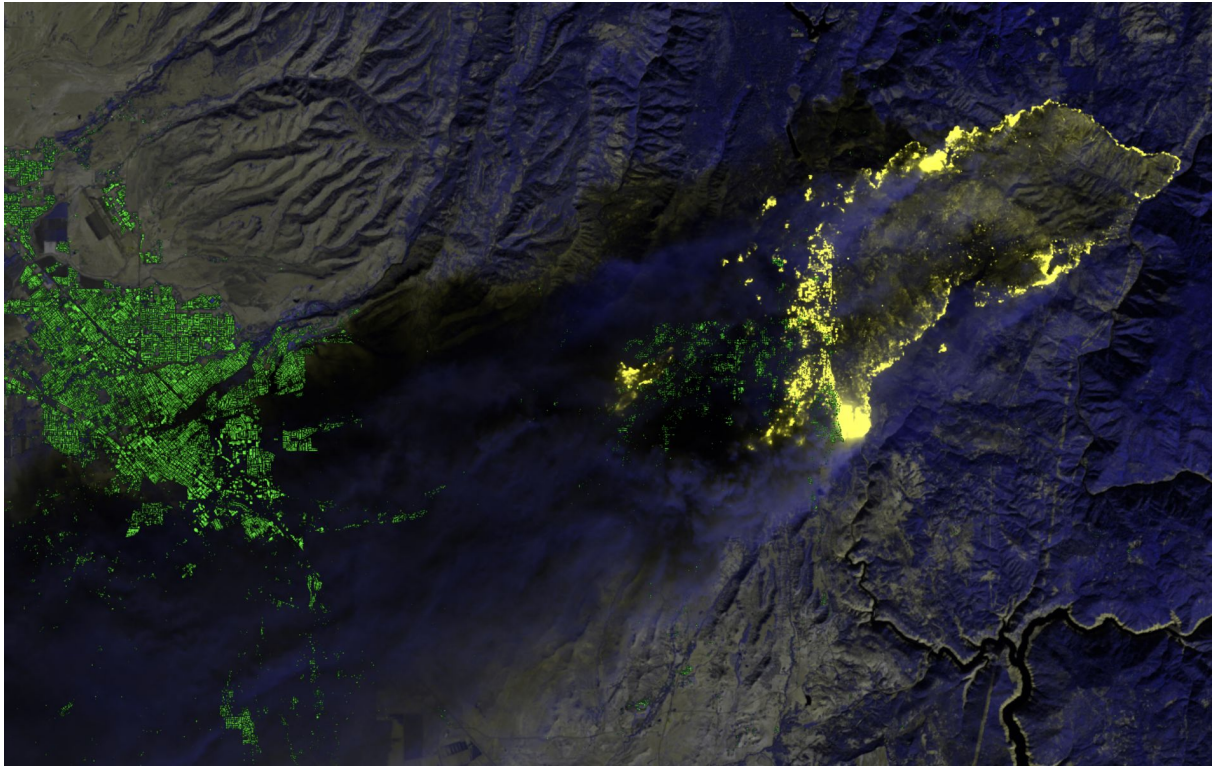
Common issues that may apply to particle physics

Want to use all information available. But are encoder networks as good with multi band and different type of imagery as natural images?

Camp Fire, CA (8 Nov 2018)
Landsat-8 (RGB)

Common issues that may apply to particle physics

Want to use all information available. But are encoder networks as good with multi band and different type of imagery as natural images?



Camp Fire, CA (8 Nov 2018)
Landsat-8 (thermal) + buildings layer

Some solutions

Contrastive Sensor Fusion (CSF)

Learning feature representations that capture valuable information from multi-band, multi-sensor images that supersede pre-trained models for downstream tasks.

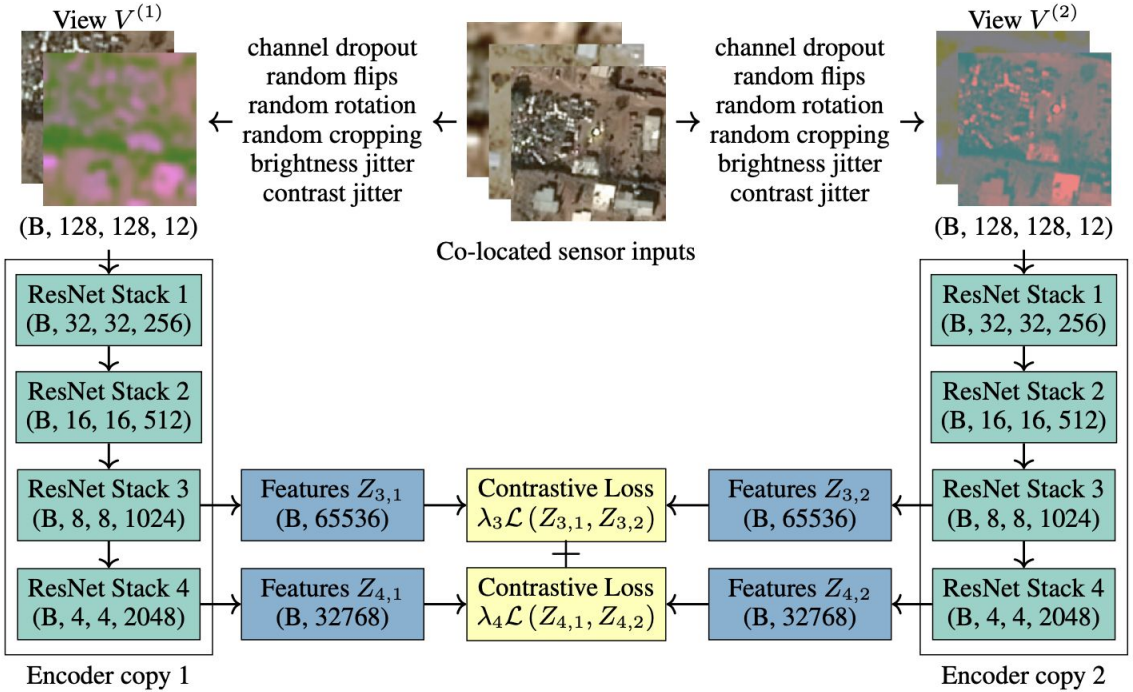
Take advantage of co-registered images from different sources using the DL platform for unsupervised training approach.



Distributional hypothesis does not apply for satellite imagery, however, the same layout measured by different sensors should have a similar representation.

Some solutions

Contrastive Sensor Fusion (CSF) - Setup



Loss function

$$\mathcal{L}_L^{\text{forward}}(X) = -\log \frac{\exp(\phi(Z_L^{(1)}, Z_L^{(2)}))}{\exp(\phi(Z_L^{(1)}, Z_L^{(2)})) + \sum_{\tilde{Z} \in \tilde{Z}_{L,\text{noise}}} \exp(\phi(Z_L^{(1)}, \tilde{Z}))}$$

where

$$\phi(Z_L^{(1)}, Z_L^{(2)}) = {Z_L^{(1)}}^T Z_L^{(2)}$$

similarity between view 1 and view 2

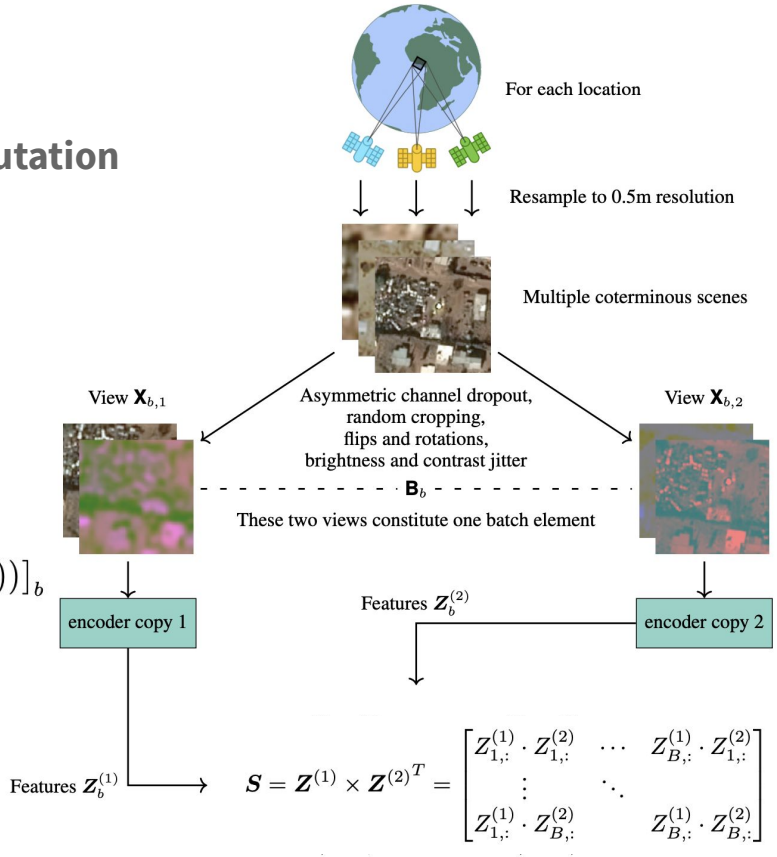
Some solutions

Contrastive Sensor Fusion (CSF) - Efficient loss computation

47 million image triplets, 4 bands in each image

Training on single TPU with batch size of 2048

$$\mathcal{L}_{\text{contrastive}}(\boldsymbol{Z}^{(1)}, \boldsymbol{Z}^{(2)}) = \frac{1}{B}\sum_{b}^{B}\left[\text{diag}\left(\text{log-softmax}(\boldsymbol{S})\right) + \text{diag}\left(\text{log-softmax}\left(\boldsymbol{S}^{T}\right)\right)\right]_b$$

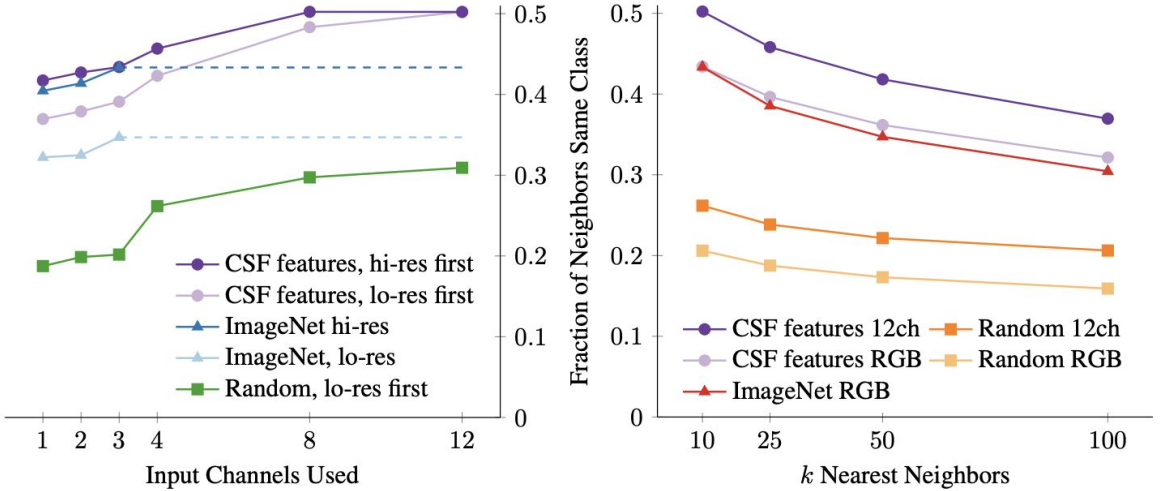


Some solutions

Contrastive Sensor Fusion (CSF) - Results

Evaluate the learned representations on OSM dataset with 8400 samples and compare to network pre-trained on ImageNet and random weights.

Reduce features to 2048 dimensional representation space (PCA) and measure the quality of clustering by determining the fraction of k nearest neighbors that belong to the same OSM class.

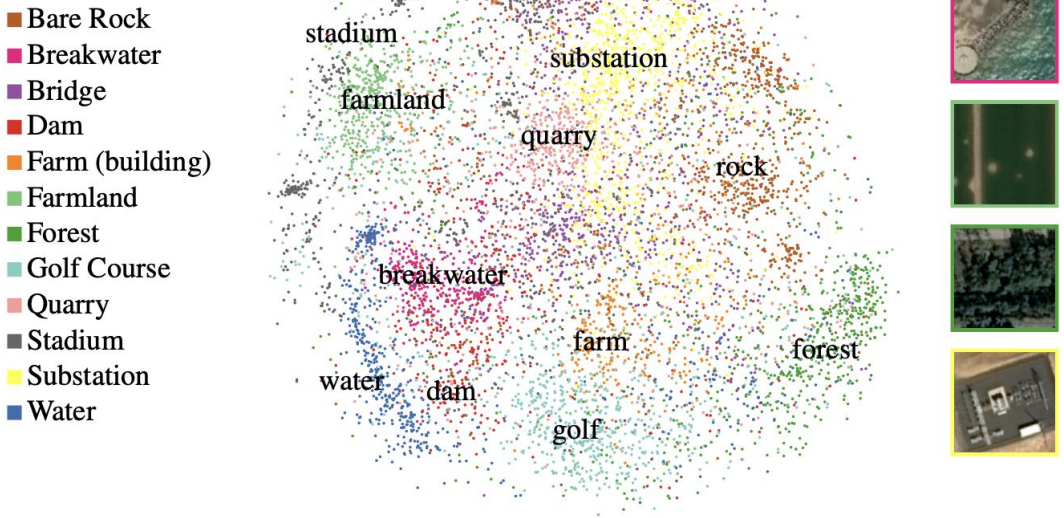


Some solutions

Contrastive Sensor Fusion (CSF) - Results

Representations learned by CSF are consistently better than pre-trained on ImageNet (supervised) and successfully disentangle semantically meaningful information.

Similar representations for Different sensors which allows fusing sensor for better transfer learning tasks



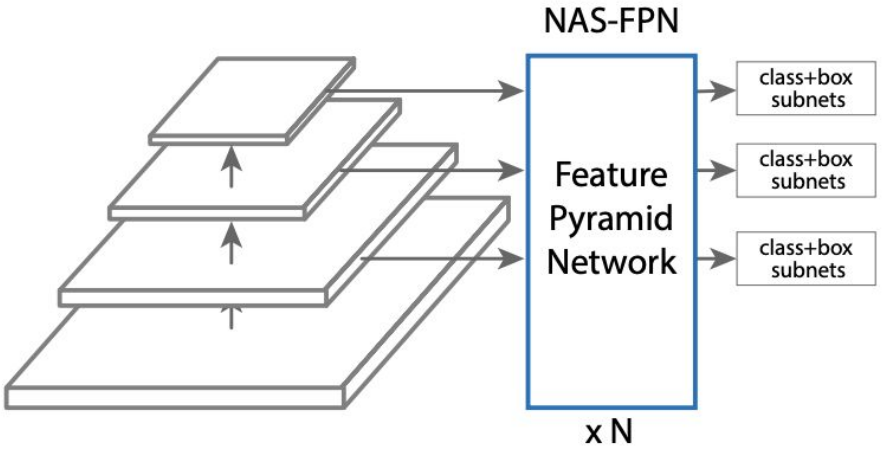
Some solutions

Neural Architecture Search (NAS)

The accuracy of semantic segmentation tasks varies with the choice of neural network architecture.

Generally well suited networks such as the UNet perform well but is there a better architecture specifically for satellite imagery?

Recent research has shown that the classical FPN may not yield the best possible results.

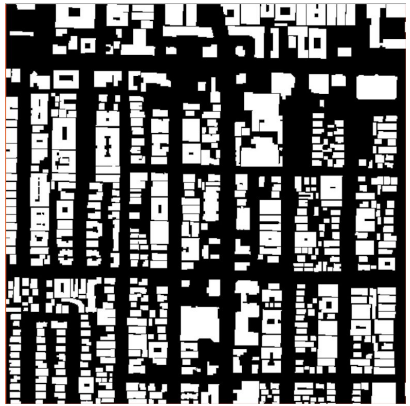


Some solutions

Neural Architecture Search (NAS)

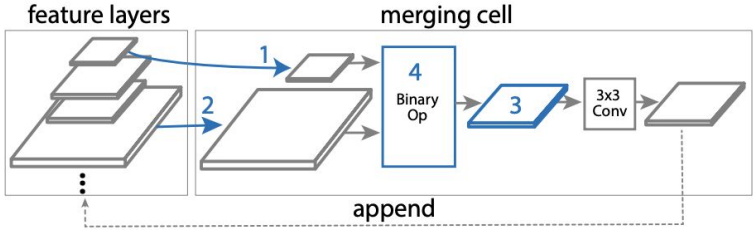
NAS trial:

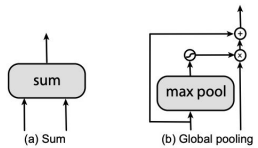
- Selected dataset: SPOT buildings segmentation, 4k/60k samples
- Backbone: UNet encoder, EfficientNet-B5
- Search space: FPN



NAS search space:

Choose any two feature layers from backbone, re-sample, binary op, append to list of features





Possible binary ops

Some solutions

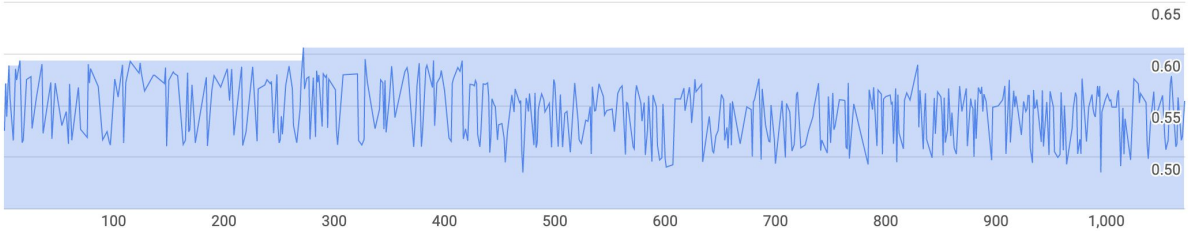
Neural Architecture Search (NAS) - Training

Stage 1

- Train multiple networks in parallel (50 TPUs) for a few epochs
- Determine metric (IoU) and inform RL agent

Stage 2

- Select top 5 networks and fine tune on full dataset

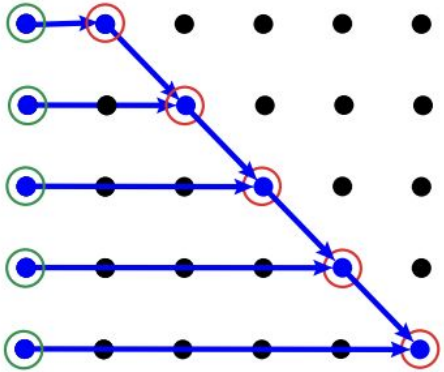


Trial ID	miou ↓	Training step	Elapsed time
272	0.60613	1,500	28 min 32 sec
328	0.59491	1,500	28 min 43 sec
390	0.59364	1,500	29 min 24 sec
15	0.59357	1,500	45 min 49 sec
416	0.59349	1,500	28 min 45 sec

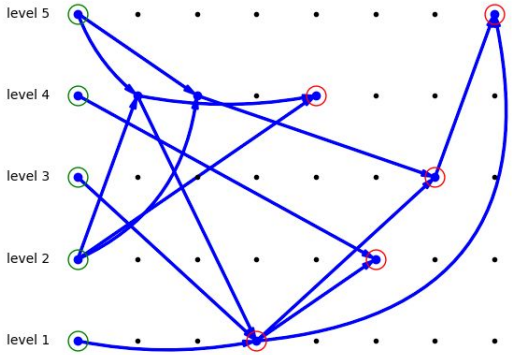
Some solutions

Neural Architecture Search (NAS) - Results

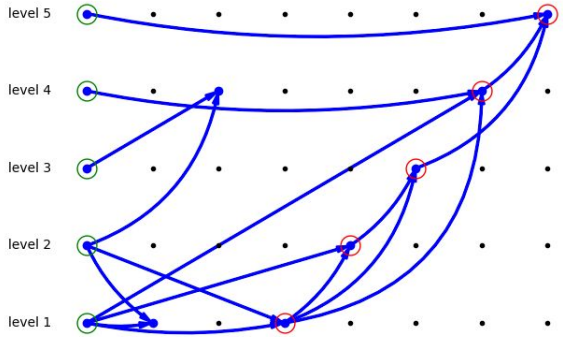
Optimization metric: IoU



FPN



NAS-FPN UNet



NAS-FPN EfficientNet-B5

Some solutions

Neural Architecture Search (NAS) - Results

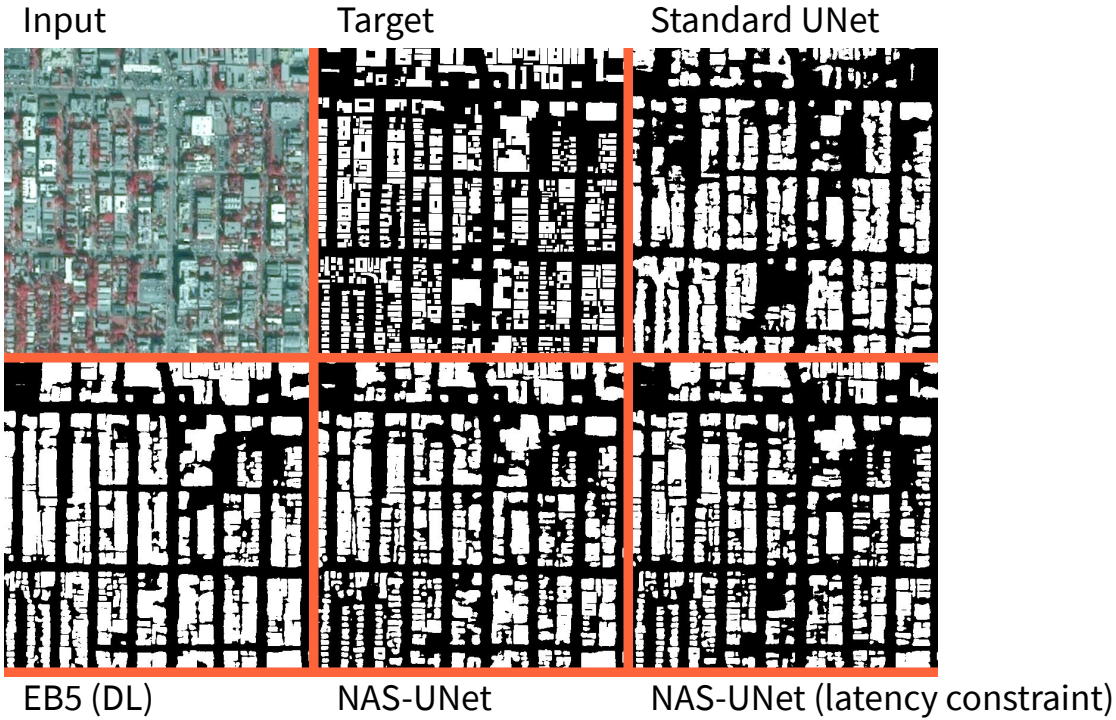
Evaluation of searched architecture on test set and comparison with UNet and EfficientNet-FPN

- NAS-FPN outperforms UNet in both accuracy and latency
- Accuracy is not better than manually constructed EfficientNet-FPN
- Latency is important for large scale deployments

Network	NAS	Latency (s)	IoU (buildings)	IoU (background)	mIoU	
UNet	No	19.1	0.512	0.928	0.715	
EfficientNet-b0	No	6.2	0.505	0.931	0.719	
EfficientNet-b5	No	17.2	0.545	0.938	0.742	
UNet	Yes	10.5	0.536	0.947	0.741	
UNet + latency contraints	Yes	9.87	0.610	0.955	0.782	(trained on larger dataset)

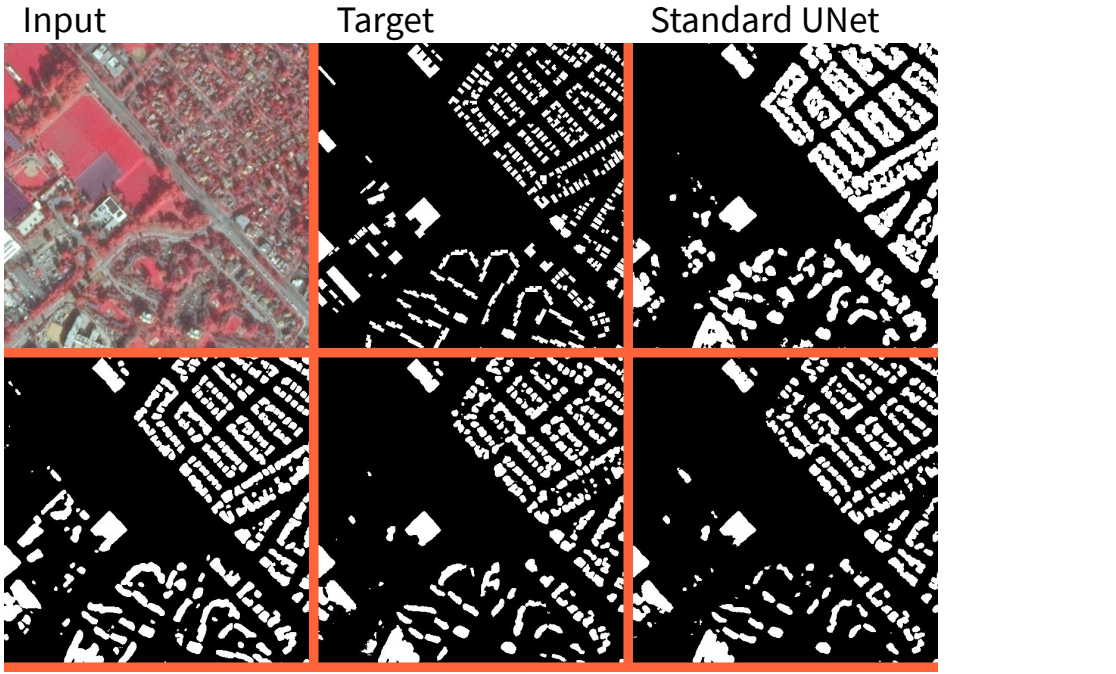
Some solutions

Neural Architecture Search (NAS) - Results



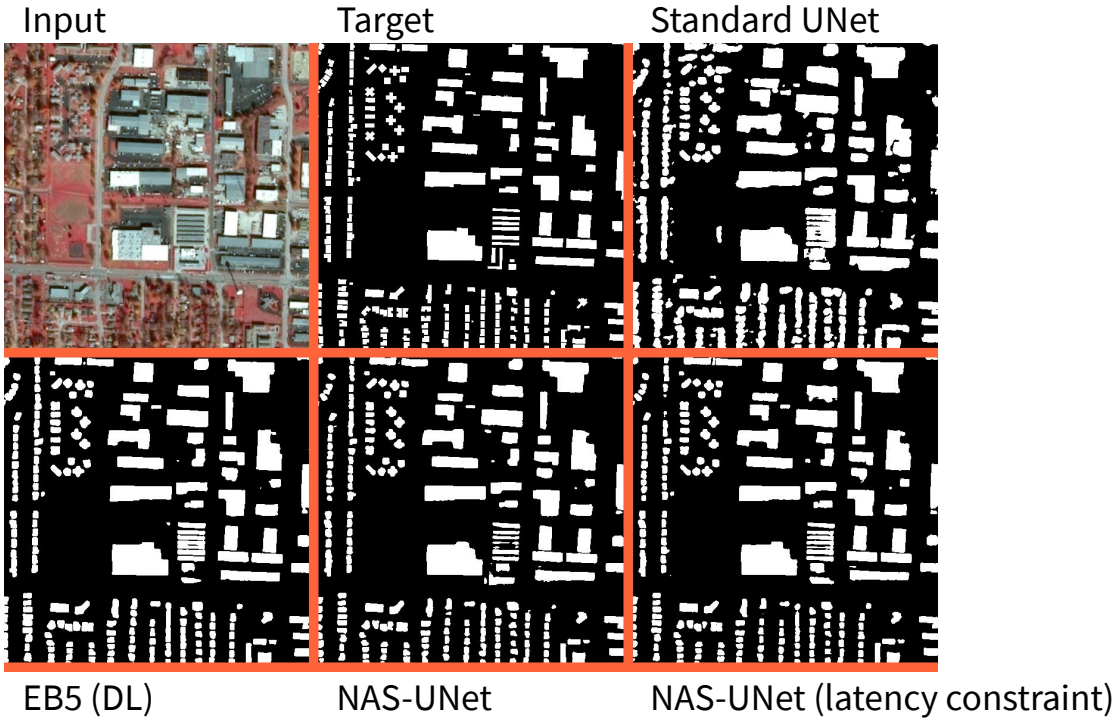
Some solutions

Neural Architecture Search (NAS) - Results



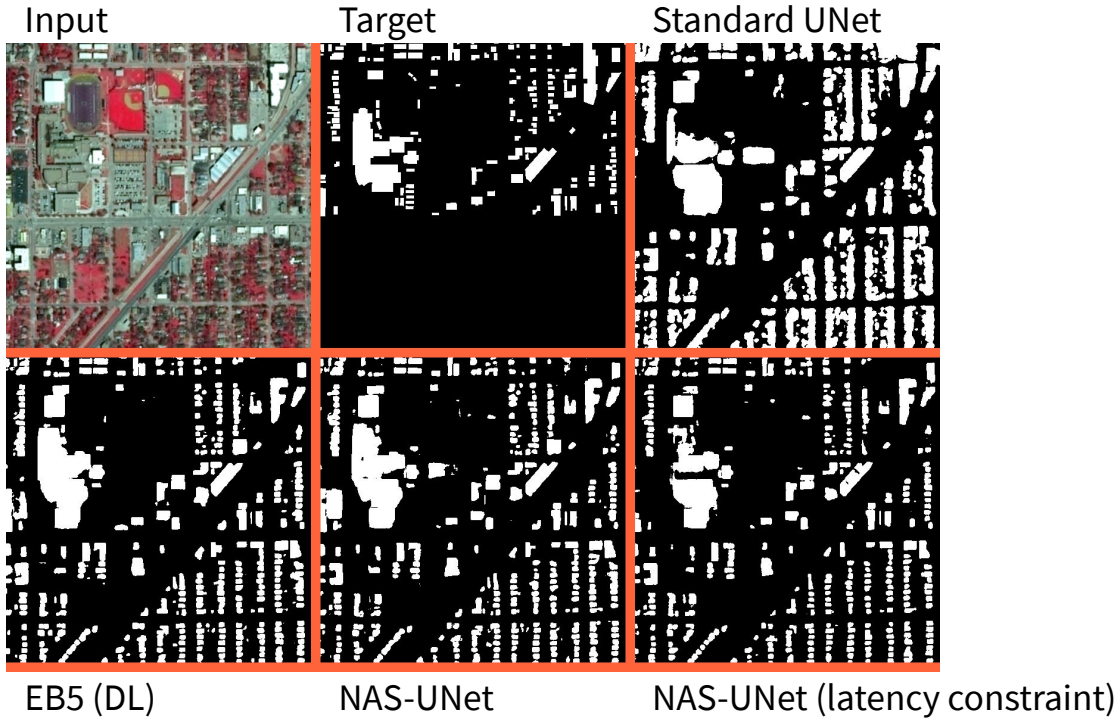
Some solutions

Neural Architecture Search (NAS) - Results



Some solutions

Neural Architecture Search (NAS) - Results



Conclusion

- State of the art deep learning architectures are pushing the accuracy for computer vision tasks
- Large datasets allow the development of complex architectures

- Many of the architectures perform well on different type of data
- In order to push task specific performance, better representations can be learned through unsupervised learning on large datasets
- ML informed architecture searches yield better performances for accuracy and latency
- Such approaches are applicable to areas where ML research does not particularly focus on

Thank you!

Learn more about Descartes Labs here:

- descarteslabs.com
- medium.com/@DescartesLabs
- Recommending this episode of Age of AI: https://www.youtube.com/watch?v=0wy4u34fii4&t=2168s

Manuel Weber
Applied Scientist (Machine Learning)
manuel@descarteslabs.com