Deep-Learning From Mistakes:

Automating Cloud Class Refinement for Sky Image Segmentation

Gemma Dianne Arnold Wiliem Brian C. Lovell

School of ITEE
University of Queensland

DICTA, 2019

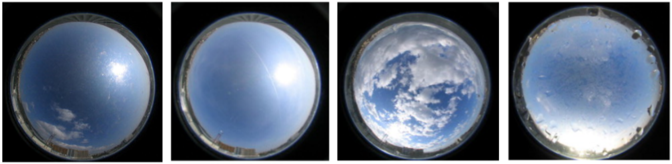
Outline

1. Whole-Sky-Image Cloud Detection
   - Challenges
   - Limitations in the Literature
2. Learning From Mistakes
   - Proposed Approach
   - Implementation
   - Validation
   - Results
3. Conclusion
   - Future Work

Why Sky Images?

Data

- Applications such as;
    - Air Traffic Control,
    - Unmanned Aircraft,
    - Cloud-track wind data monitoring,
    - Solar PV Power Predictions [1–3].
- Satellites images don't have temporal/spacial resolution necessary [2].
- Whole-Sky-Imager - most often a camera mounted on the ground, pointed at the sky, with a fish eye lense.

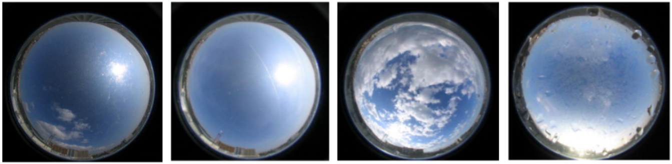


Examples of Whole-Sky Images [4]

What Are We Dealing With?

Challenges

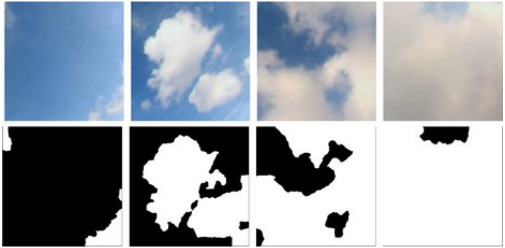
- The clouds to be detected in the resulting sky images vary in density, texture, shape, and height.
- Their relative position to the sun impacts their illumination.
- There can be multiple clouds obscuring/shading each other and/or blocking the sun.



Examples of Whole-Sky Images [4]

How Good Is Cloud Detection Now?
Limitations

- Open sky image data is limited, so appropriate comparison of methods is difficult.
- Majority of errors attributed to thin cloud
- Models produce only binary sky/cloud segmentations (excluding [5, 6])

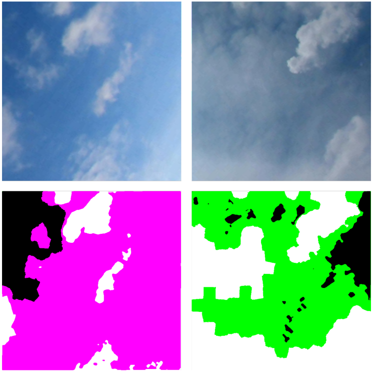


Examples of SWIMSEG images and corresponding ground truths

What Did We Do Differently?

Our Work

- Utilised errors to incorporate a new classes without the need for further hand labelling.
- Showed that additional cloud classes increases stability and speed of learning.
- Identified further avenues for improvements in classes and model training.

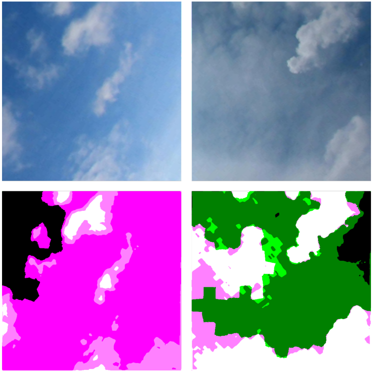


Example images where S-CSN-1 binary segmentations produced very high error

What Did We Do Differently?

Our Work

- Utilised errors to incorporate a new classes without the need for further hand labelling.
- Showed that additional cloud classes increases stability and speed of learning.
- Identified further avenues for improvements in classes and model training.

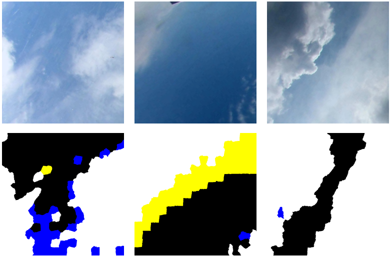


Example images where S-CSN-1 binary segmentations produced very high error

What Can The Mistakes Tell Us?

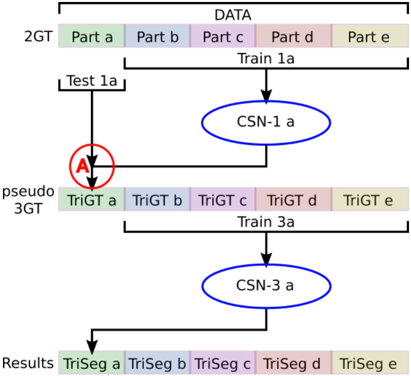
Proposal

- Humans even find it hard to agree on cloud boundaries.
- Consider; if a mistake is being made in a particularly ordered way, it may provide new insights.
- Network is 'splitting the difference'.



Inconsistencies in SWIMSEG ground truths for repeated images.

- Using a two step learning procedure.
- Data partitioned to avoid advantage.
- A - defining three classes for the next stage of training;
    - Thick: Both 2GT and CSN-1 classify as Cloud
    - Thin: 2GT and CSN-1 disagree on classification.
    - Sky: Both 2GT and CSN-1 classify as Sky

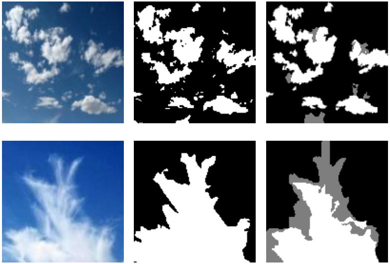


Proposed methodology overview

How Do We Know If It Works?

Validation

- Utilised the ternary labels of the HYTA set for validation.
- Also trained models on the HYTA dataset in order to understand the benefits of a hand-labelled ternary trained model.



HYTA images with both binary and ternary ground truths [5, 7]

	HYTA -Binary				
Net	**Precision**	**Recall**	**IoU**	**F-Score**	**Error**
H-CSN-1	**86.3**	86.2	**75.8**	**86.2**	**14.3**
H-CSN-2	84.9	86.0	74.5	85.4	15.3
H-CSN-3	74.1	91.8	69.5	82.0	21.0
H-CSN-3GT	75.4	84.2	66.0	79.5	22.5
S-CSN-1	$\boxed{80.6}$	90.0	$\boxed{73.9}$	$\boxed{85.0}$	$\boxed{16.6}$
S-CSN-2	78.6	90.4	72.5	84.0	18.0
S-CSN-3	70.5	$\boxed{\mathbf{96.7}}$	68.8	81.5	22.8

HYTA Ternary Ground Truths

Results

	Sky		Thin		Thick	
Method	**Error**	**IoU**	**Error**	**IoU**	**Error**	**IoU**
H-CSN-3	**20.6**	60.6	21.9	11.9	10.2	79.1
H-CSN-3GT	22.0	**60.9**	**21.6**	8.3	10.8	77.1
S-CSN-3	23.0	53.9	22.9	**15.7**	**6.5**	**86.8**

- Note that there seems very little advantage to training on Ternary Ground Truths
- The Swimseg Trained model still performs very well, arguably the best on thin cloud

HYTA Model Comparisons

Qualitative Results

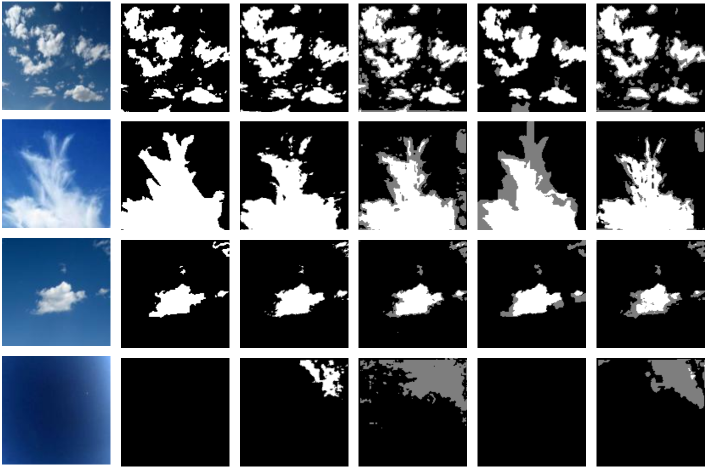
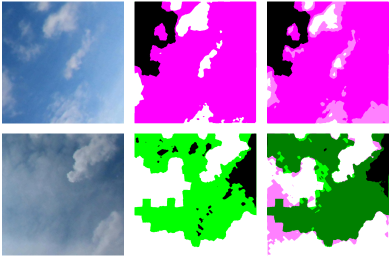


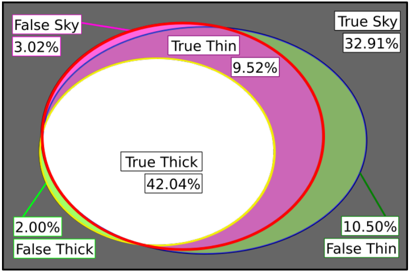
Image 2GT H-CSN-1 H-CSN-3 3GT CSN-3GT

SWIMSEG Binary Ground Truths
Results

	SWIMSEG				
Net	**Precision**	**Recall**	**IoU**	**F-Score**	**Error**
H-CSN-1	89.6	83.6	$\boxed{76.2}$	$\boxed{86.5}$	$\boxed{14.3}$
H-CSN-2	$\boxed{90.0}$	82.7	75.7	86.2	14.5
H-CSN-3	81.2	$\boxed{91.4}$	75.4	86.0	16.3
H-CSN-3GT	85.9	85.7	75.0	85.7	15.6
S-CSN-1	86.7	90.8	**79.7**	**88.7**	**12.6**
S-CSN-2	**91.7**	84.7	78.6	88.0	12.6
S-CSN-3	80.5	**94.5**	76.9	86.9	15.5

- Now we seek to understand what these ternary labels mean in relation to SWIMSEG binary ground truths





Venn Diagram of S-CSN-3 segmentation results

HYTA Images -The Good, The Bad, and The Ugly

Comparisons

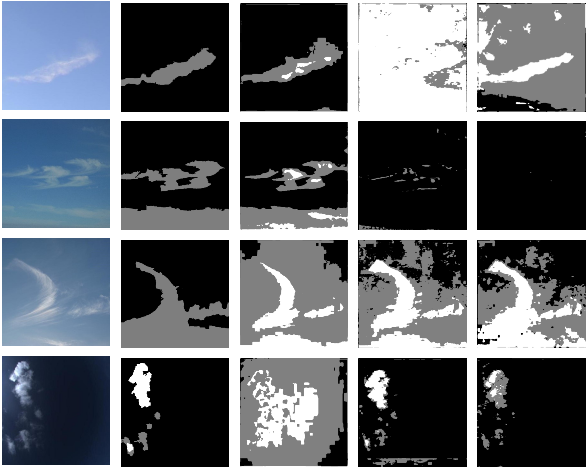


Image 3GT S-CSN-3 H-CSN-3 CSN-3GT

SWIMSEG Images -The Good, The Bad, and The Ugly

Comparisons

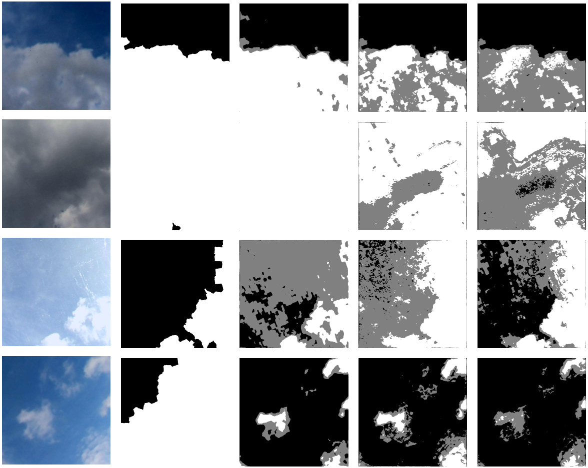


Image | 2GT | S-CSN-3 | H-CSN-3 | CSN-3GT

Final Points

Conclusion

- SWIMSEG trained models had a better grasp of classes, though struggled with previously unseen illumination conditions (colour and contrast)
- While HYTA could handle a wider variety of image types, it struggled with dark thick/thin cloud.
- Analysis of models identified further inconsistencies/ambiguities in both dataset's ground truths.

Transfer	Cost	Precision	Recall	IoU	F-Score	Error
SWIMSEG ↪HYTA	CSN-1/ CSN-3	5.7/ 3.6	**-3.8/ -4.9**	**1.9/ 0.7**	**1.2/ 0.5**	2.3/ 1.8
HYTA ↪SWIMSEG	CSN-1/ CSN-3	**-2.9/ -0.7**	7.2/ 3.1	3.5/ 1.5	2.2/ 0.9	**1.7/ 0.8**

What is left to be done?

Future Work

- Increase cloud classes further
- Improve datasets, to include a wider variety of images and sufficient quantity
- extend to whole sky images rather than limiting to image patches.

For Further Reading I

[1] S. Dev, B. Wen, Y. H. Lee, and S. Winkler, "Ground-based image analysis: A tutorial on machine-learning techniques and applications," *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, pp. 79–93, June 2016.

[2] S. Dev, Y. H. Lee, and S. Winkler, "Color-based segmentation of sky/cloud images from ground-based cameras," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 1, pp. 231–242, Jan 2017.

[3] D. Tulpan, C. Bouchard, K. Ellis, and C. Minwalla, "Detection of clouds in sky/cloud and aerial images using moment based texture segmentation," in *2017 International Conference on Unmanned Aircraft Systems (ICUAS)*, June 2017, pp. 1124–1133.

[4] A. Kazantzidis, P. Tzoumanikas, A. Bais, S. Fotopoulos, and G. Economou, "Cloud detection and classification with the use of whole-sky ground-based images," *Atmospheric Research*, vol. 113, pp. 80–88, 2012.

[5] S. Dev, Y. H. Lee, and S. Winkler, "Multi-level semantic labeling of sky/cloud images," in *2015 IEEE International Conference on Image Processing (ICIP)*, Sept 2015, pp. 636–640.

[6] S. Dev, S. Manandhar, Y. H. Lee, and S. Winkler, "Multi-label cloud segmentation using a deep network," in *Proc. IEEE AP-S Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting*, 2019.

[7] Q. Li, W. Lu, and J. Yang, "A hybrid thresholding algorithm for cloud detection on ground-based color images," *Journal of atmospheric and oceanic technology*, vol. 28, no. 10, pp. 1286–1296, 2011.