CloudFCN: Cloud masking with Deep Learning for Landsat 8

A. M. Francis[1,2] *a.francis.16@ucl.ac.uk*
P. Sidiropoulos[1], J.-P. Muller[1]

[1] Imaging Group, Mullard Space Science Laboratory, Holmbury St Mary, UK [2] Cortexica Vision Systems Ltd., London, UK

UCL

ABSTRACT

Cloud masking is of central importance to the Earth Observation community. Recently, Machine Learning has offered promising solutions to the problem of cloud masking, allowing for more flexibility than traditional thresholding techniques. However, few approaches use multi-scale features (i.e. a combination of pixel-level and spatial) whilst also being fully differentiable. Differentiability is a desirable quality, as it allows for the entire model to be optimised in the training phase, based on a single loss function. Therefore, we introduce our algorithm, CloudFCN. The developed algorithm is based on a Fully Convolutional Network (FCN) architecture, which fuses the shallowest and deepest layers of the network, thus routing low-level visible content to its deepest layers. Here we present its state-of-the-art performance on Landsat 8 imagery.

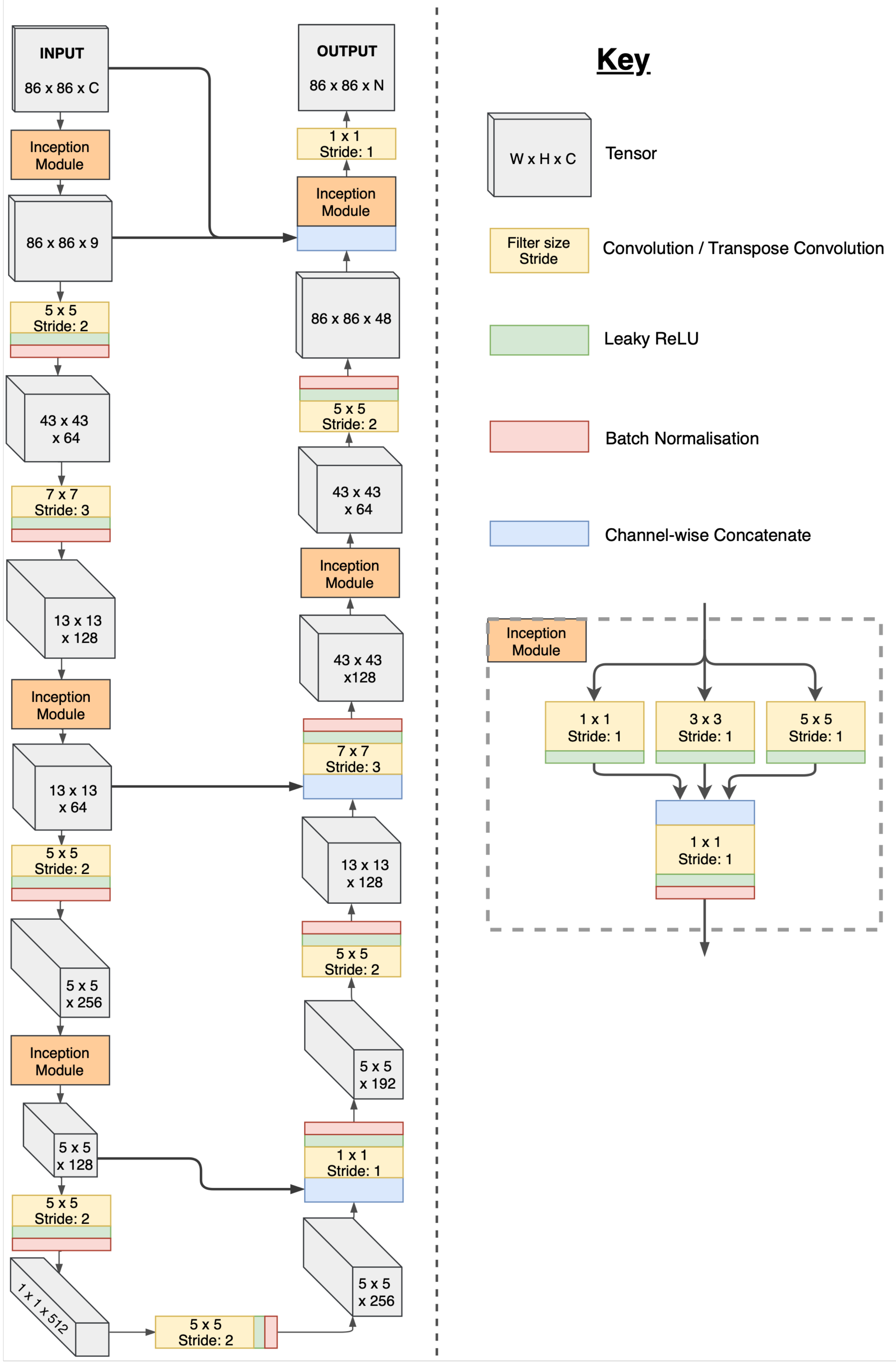

Figure 1: Flowchart of CloudFCN. Bold arrows represent residual connections, connecting different stages of the encoder and decoder. 'Inception module' denotes a set of parallel convolutions of stride 1, followed by dimensionality-reduction. Sizes of the input are the *minimum* and could be anything above 86 pixels across, with internal tensor dimensions changing accordingly. The first input layer is given C channels, as a placeholder for the number of spectral bands in the input.

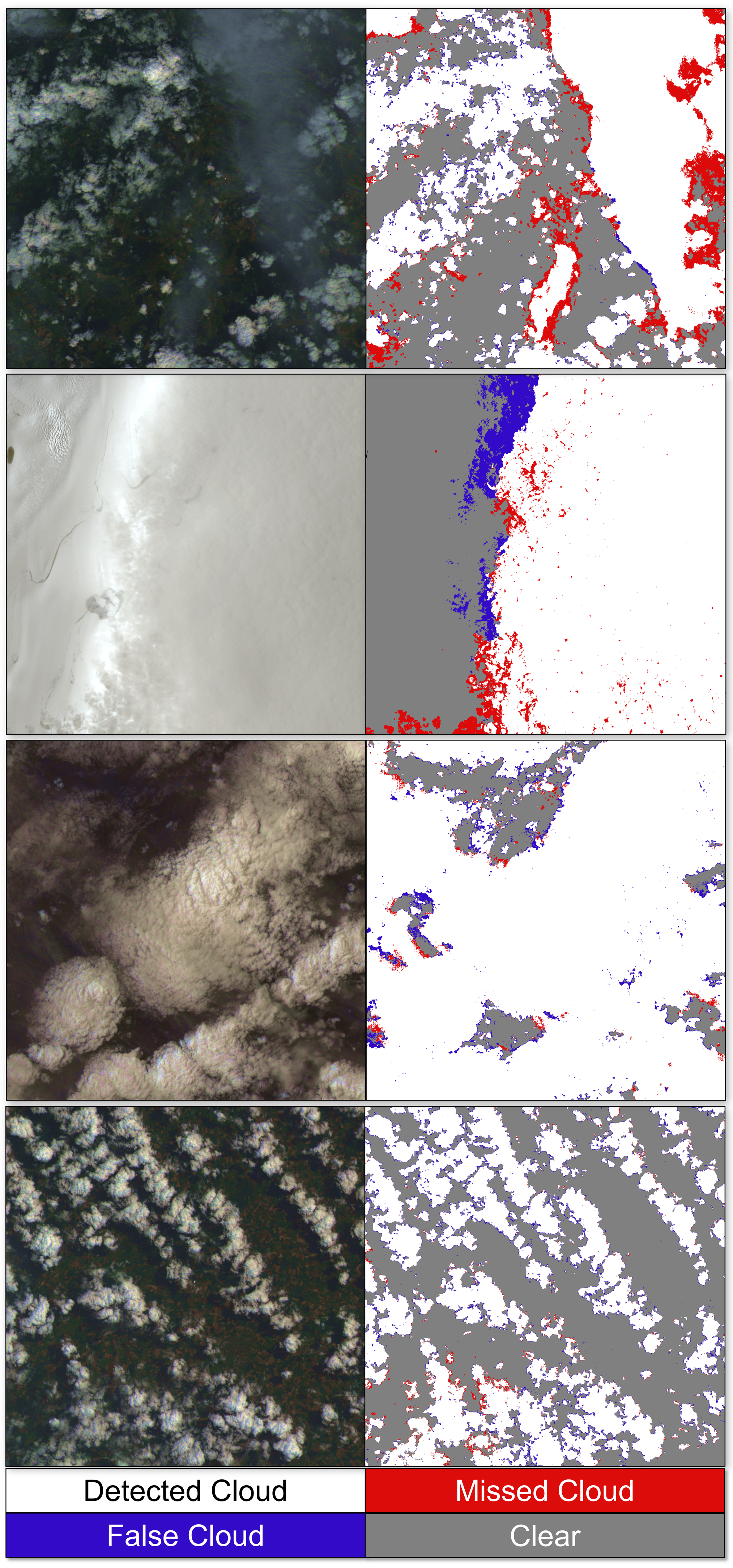

Figure 2: Examples of cloud masks generated by CloudFCN. The images are taken as 1k-by-1k pixel crops from scenes in the Biome dataset. Some errors of Omission can be seen for thin cloud in the top pane, whilst some errors of Commission are present over snow in the second pane.

METHOD

As input, the model takes scenes with arbitrary spatial dimensions. The output of the model is the same size as the input, wherein each pixel is classified as cloudy or clear. The model comprises two convolutional arms, the **encoder** and the **decoder** (see Fig. 1). The encoder serves to **extract spatial features** from the scene, reducing the spatial dimensions whilst increasing the number of channels. Meanwhile, the decoder takes these features and **reprojects them** to create the output mask. Residual connections link intermediate points in the encoder and decoder, to allow for **fusion between low- and high-level features**. This is in the style of models such as U-net [1].

The proposed model has several desirable traits. First, our design allows for **flexible input formats**, with no specific spectral requirements, or size restrictions. Second, we combine the simplest brightness and colour information with more complex spatial descriptors, creating a **complementary set of features** that exhibit the strengths of both pixel-level techniques and convolutional ones. Lastly, the model is **fully differentiable**, allowing for it to be optimised and adapted for other datasets easily.

RESULTS

CloudFCN achieves **state-of-the-art performance on Landsat 8**. To validate our algorithm, we used the 'biome' dataset provided by USGS [2], which separates 96 scenes into 8 different terrain categories. This allows us to measure performance over different terrains, and compare our results with previous algorithms. We use the Quality Metric, defined as *%Quality = %Correct - %Omission - %Commission*. Two modes were tested, 'RGB' (bands 4,3 and 2) and 'Multispectral' (all Landsat 8 bands). In each experiment, a model is trained and validated on half the data, which is then reversed and the results combined (see Table 1). The RGB performance suffers in high albedo terrains (e.g. Snow-Ice) because these terrains are similar to clouds in visible wavelengths. **Multispectral performance was generally higher**, with performance higher than all results reported in [3] in **5 of the 8 biomes**.

		Barren	Forest	Grass-Crops	Shrubland	Snow-Ice	Urban	Water	Wetlands	Average
CloudFCN (RGB)	%Correct	78.92	82.69	94.69	93.41	49.65	93.41	92.34	77.36	82.81
	%Omission	12.24	23.20	4.08	8.82	7.25	2.88	6.34	24.32	11.14
	%Commission	29.89	11.00	7.36	4.47	76.96	8.64	8.18	19.90	20.80
	%Quality	36.79	48.48	83.25	**80.11**	-34.56	81.89	77.82	33.14	50.87
CloudFCN (Multispectral)	%Correct	92.95	95.12	96.12	88.68	72.93	95.56	95.43	91.24	91.00
	%Omission	4.70	7.21	6.27	19.66	17.60	2.21	4.62	13.16	9.43
	%Commission	8.95	1.92	1.79	3.23	27.53	5.75	4.50	3.89	7.19
	%Quality	**79.30**	**85.99**	**88.06**	65.79	27.80	**87.61**	**86.31**	74.19	**74.38**
ACCA	%Quality	63.02	68.69	62	60.47	**36.25**	68.33	71.43	62.48	61.56
AT-ACCA	%Quality	66.67	73.83	74.09	70.65	35.86	74.06	70.51	**76.25**	67.72
cfmask	%Quality	77.1	67.27	85.74	75.53	26.37	74.72	50.98	65.97	65.69
cfmask_conf	%Quality	66.78	66.72	83.59	72.3	20.75	76.54	51.11	67.45	63.63
cfmask_nt_cirrus	%Quality	54.23	57.2	70.71	71.58	-15.87	74.37	50.23	47.16	51.62
cfmask_nt_cirrus_conf	%Quality	54.44	38.79	60.01	66.38	-43.68	73.2	49.04	35.14	41.66
cfmask_t_cirrus	%Quality	69.82	64.78	77.98	72.75	-24.1	72.42	57.21	53.27	49.01
cfmask_t_cirrus_conf	%Quality	69.37	43.99	77.76	72.34	-52.76	74.72	57.24	52.14	49.63
See5	%Quality	54.19	51.88	42.15	42.46	35.48	57.4	39.35	68.17	49.17

Table 1: Cloud detection results for Biome dataset. Comparison with 9 other algorithms validated in [3] are given at the bottom, with Quality values derived from the values given in the study. For each biome, the best-performing algorithm is in bold. The multispectral performs best overall, with an average Quality 6.7% greater than the next highest, AT-ACCA.

CONCLUSIONS

Our experiments have shown that our Deep Learning CloudFCN model performs well in a range of settings, and is immediately applicable to Landsat 8 data as a state-of-the-art solution to the problem of cloud masking. This work demonstrates the large potential for FCN's in solving many segmentation tasks in Remote Sensing. Re-purposing the algorithm for new use cases is straightforward, assuming the existence of sufficient labelled data. A direct extension to this work will be the masking of cloud shadows, and more tangential possibilities also exist. For example---land-use segmentation, sea ice classification and vegetation indexing.

ACKNOWLEDGEMENTS

We would like to thank Cortexica Vision Systems, Antoine Broyelle, Agneta Rüsy, Earth-i, and the USGS Cloud Cover Assessment Validation team for their help in producing this work.

For further information and upcoming source code release:

github.com/aliFrancis/cloudFCN-news

[1] Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," may 2015. Available: http://arxiv.org/abs/1505.04597
[2] U.S. Geological Survey, 2016. L8 Biome Cloud Validation Masks. U.S. Geological Survey, data release. doi:10.5066/F7251GDH
[3] S. Foga et al. "Cloud detection algorithm comparison and validation for operational Landsat data products," Remote Sensing of Environment, 2017

CORTEXICA
AI SOLUTIONS FOR BUSINESS