Efficient Convolutional Neural Networks for Cloud Detection in Satellite Images

John Merriman Sholar

CS 231N: Convolutional Neural Networks for Visual Recognition

Problem

Overview

- Goal: identify clouds in satellite images
- Motivation: removing ("masking out") clouds important for downstream analysis for satellite imagery.
- Task: image segmentation with classes including dense clouds, cirrus clouds, shadows, water, snow, and land.
- Evaluation: per-pixel classification accuracy, CE loss, qualitative analysis of output, classification speed.

Subtasks:

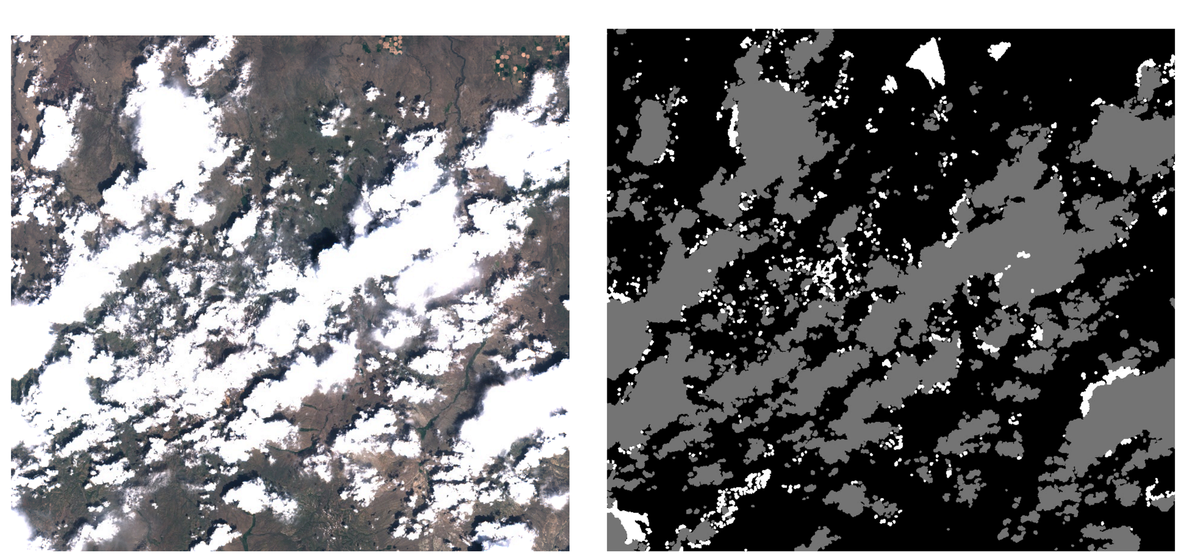
We consider several permutations of the problem definition.

- Given information: image segmentation using all 13 Sentienl-2 bands, or only using RGB.
- Output classes: we consider 2-class, 3-class, and 6-class classification problems, defined as follows.

2 Classes	CLOUD, CLEAR
3 Classes	DENSE, CIRRUS, CLEAR
6 Classes	DENSE, CIRRUS, SHADOW, WATER, SNOW, CLEAR

Data

- 60 Sentinel-2 satellite images ≈ 10,000 by 10,000 px. ea.
- Segmented into ≈ 120,000 tiles of size 224 by 224 px.
- Sentinel-2 images have 13 spectral bands, rather than 3 (RGB). RGB frequencies range from .665µm to .490µm, while S2 bands range from 2.190µm to .443µm, with more information (e.g. infrared, short-wave infrared).
- Satellite imagery is very different from "traditional" imagery (top- down, spatial covariance, lack of central focus, scale), making transfer learning or pretrained classifiers less useful.
- Sentinel-2 ships with a proprietary cloud mask (below), which provides noisy but useful training data.



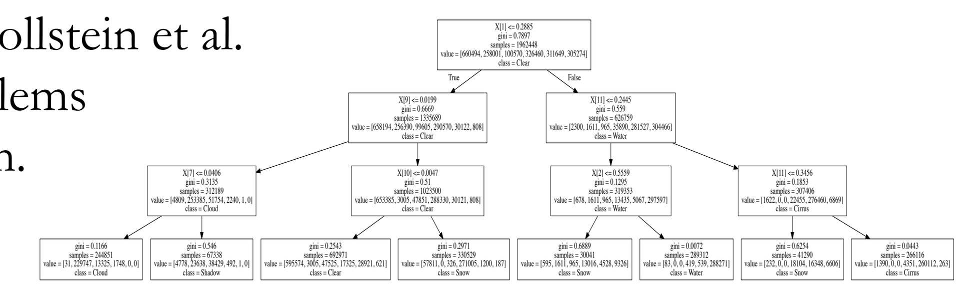
Baseline Methods

Previous Approaches:

- Previous approaches are decision trees and linear classifiers, emphasizing speed.

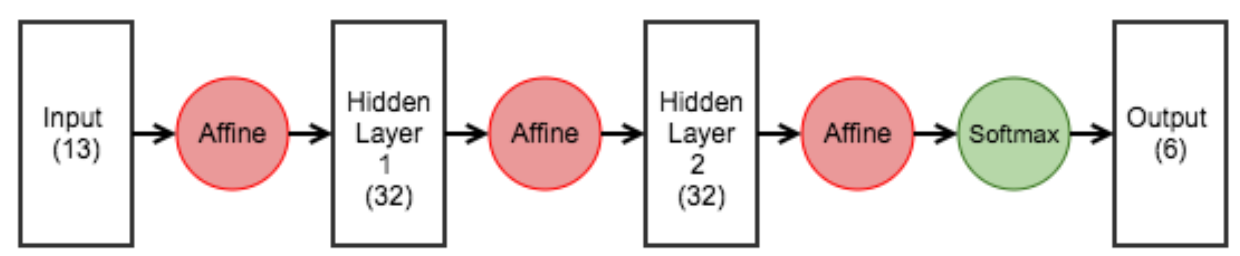
Pixel-Level Decision Tree

- A recreation of previous research by Hollstein et al.
- Extremely fast, but innacurate for problems more advanced than the binary problem.
- Useful for a form of transfer learning: we generate training data for CNN's by applying decision trees, using the Sentinel-2 cloud mask as an extremely noisy label.



Pixel-Level MLP

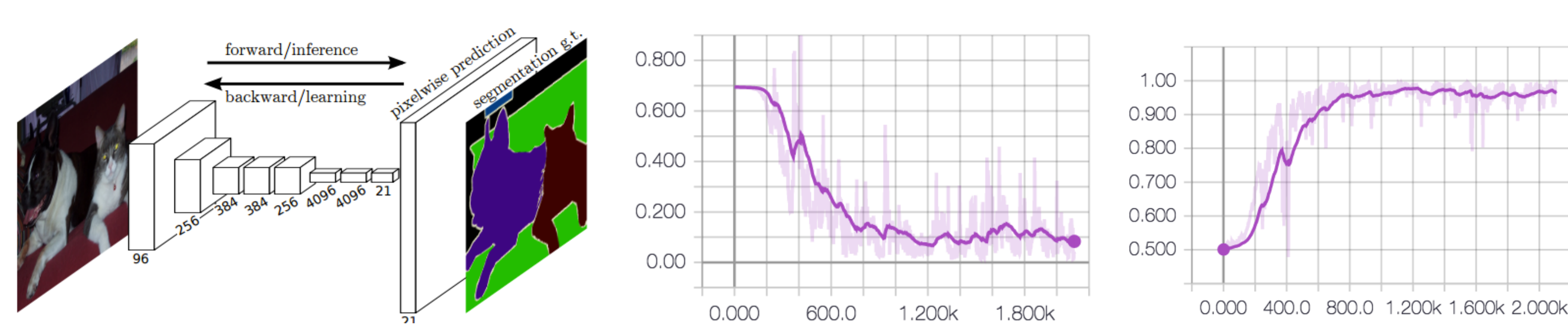
- An expansion on the pixel-level classification using decision trees. Ultimately no more accurate than decision trees, likely due to the same inability to take advantage of spatial covariance, with significantly increased inference time.



Methods

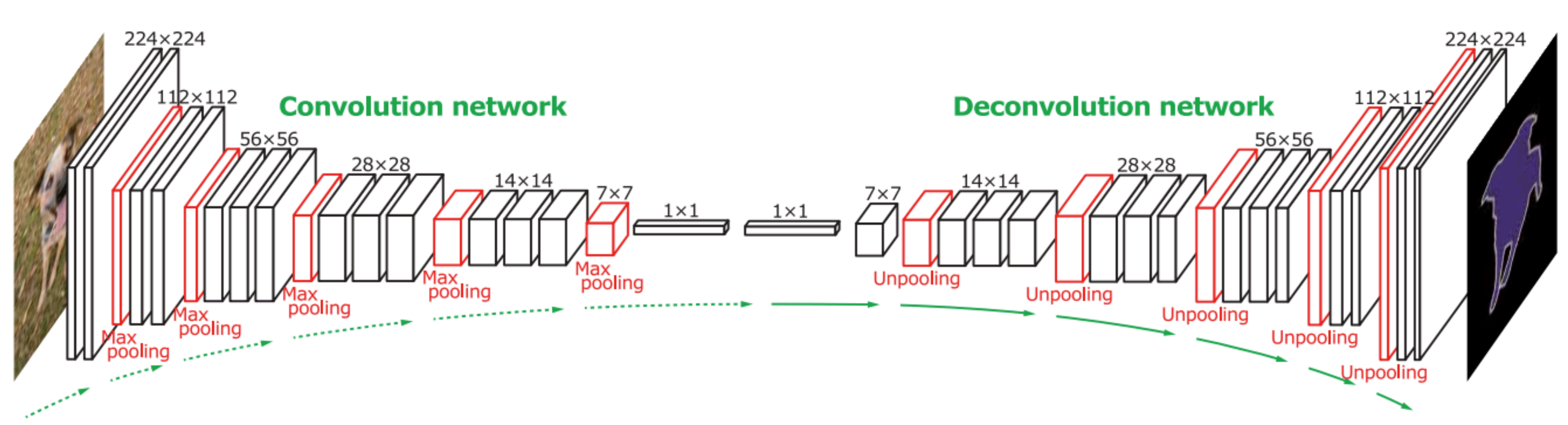
Fully Convolutional Networks Networks (Long et al.)

- A modification of the Alexnet (Krizhevsky et al)
- architecture which removes the fully connected layers in favor of fully convolutional layers, and adds a singular deconvolution. (Below are architecture, loss, and accuracy)



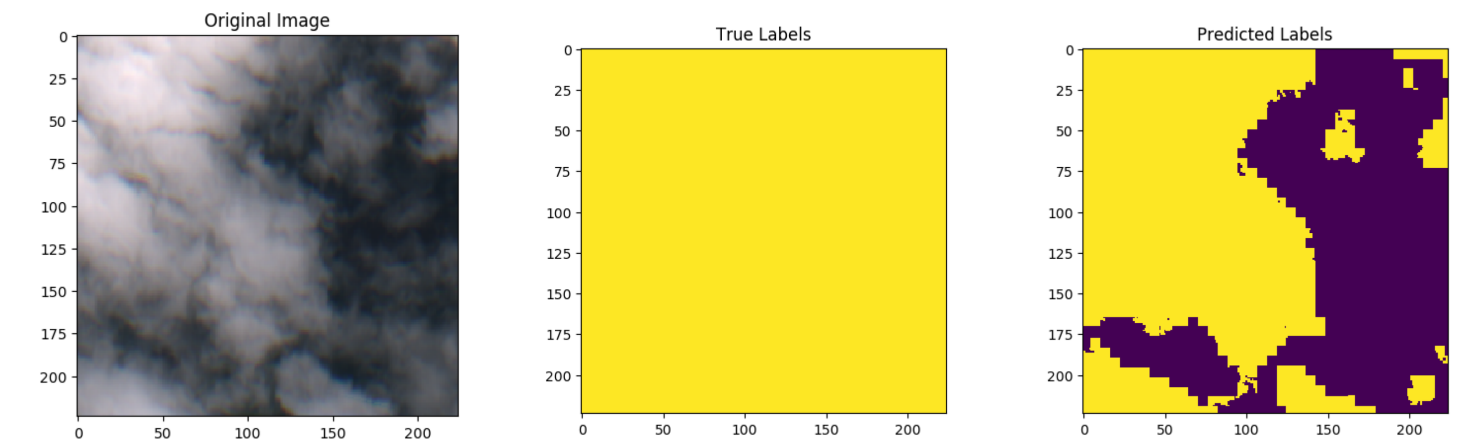
Deconvolutional Neural Networks (Noh et al.)

- Parallel convolutional and deconvolutional structure.
- Convolutional first half initialized using ILSVRC-pretrained VGG-16

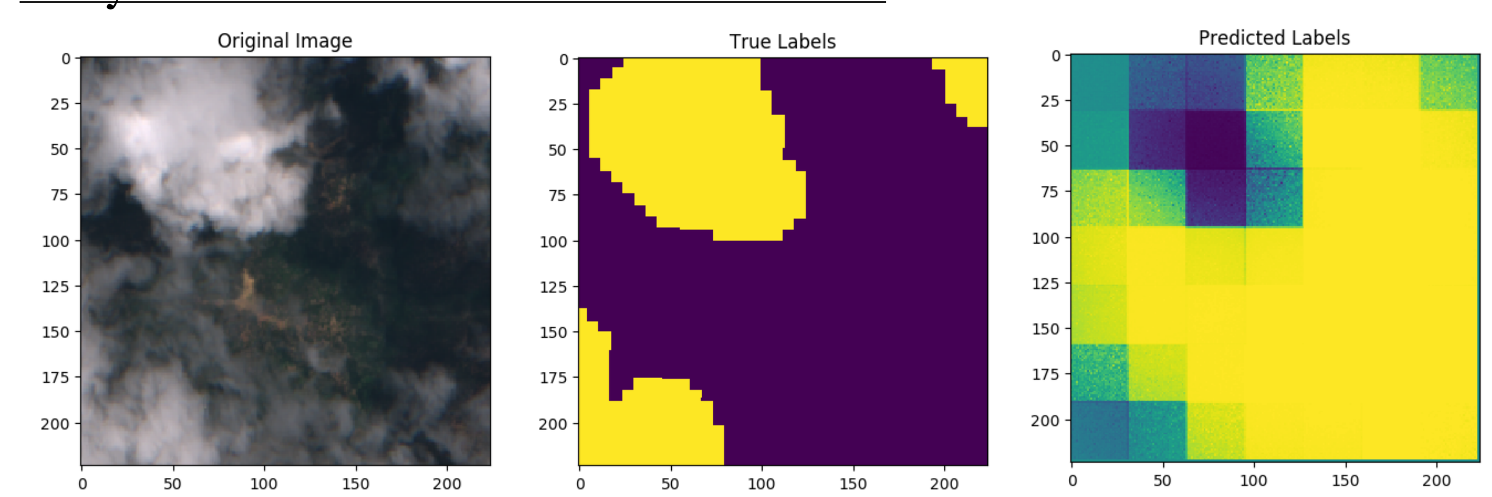


Results

Pixel-Level Decision Trees



Fully Convolutional Networks Networks



Deconvolutional Neural Networks (Noh et al.)

- We are unsuccessful in training deconvolutional neural networks. This is likely a result of several factors. First, for the RGB-only task, the VGG-16 ConvNet is pretrained on ILSVRC data, which does not parallel Sentinel-2 data. For the 13-band task, training may simply require more computational resources.

Inference Speed and Accuracy

	Inference Accuracy (F1 Score)	Inference Speed
Decision Trees	0.653	16 million pixels/second
Fully Conv. Networks	0.822	1 million pixels/second
Deconv. Networks	0.516	40,000 pixels/second

Discussion and Future Work

- Pixel-level decision trees achieve visually satisfactory results and high numerical accuracy for the binary problem, but fail on more difficult classes, such as shadows.
- Fully convolutional neural networks produce coarse output, but could be refined by using a series of deconvolutional layers – likely the most viable next step.
- Deconvolutional networks have strong potential for this task, but can't be pretrained on ILSVRC data, which is significantly different from satellite imagery. Future work might include training a DNN end-to-end on Sentinel-2 data.
- Satellite images are extremely large; processing time is key. Future work might include application of SqueezeNet (Iandola et al.) architectures to Sentinel-2.