Understanding Satellite-Imagery-Based Crop Yield Predictions

Mark Sabini (msabini), Gili Rusak (gili), Brad Ross (bross35)

CS 231N (Convolutional Neural Networks for Visual Recognition), Stanford University

Introduction

- Crop yield prediction at local levels important for preventing food shortages
- Crop yield was predicted using crude and expensive censuses
- Remote-sensing data and technologies such as Convolutional Neural Networks (CNNs) makes localized predictions possible
- You et. al. attempted soybean yield prediction using CNNs and remote-sensing data
- We aim to investigate effectiveness of You et. al.'s model [1] and improve their results.

Problem Statement

- Let C be a set of agriculturally-important counties.
- Given year Y and county $c \in C$, predict the annual crop yield of c using satellite imagery of all counties in C from years $Y_0, Y_0 + 1, \cdots, Y - 2, Y - 1$
- **Evaluation:** Root mean squared error (RMSE) between predicted crop yield and ground-truth USDA survey results

Datasets

Raw Data: MODIS satellite imagery [2]

- Surface **Reflectance** 8-Day L3 Global 500m (Bands 1-7)
- Land Surface **Temperature** & Emissivity 8-Day L3 Global 1km (Bands 1 & 5)
- Land **Cover** Type Yearly L3 Global 500m (Band 1)

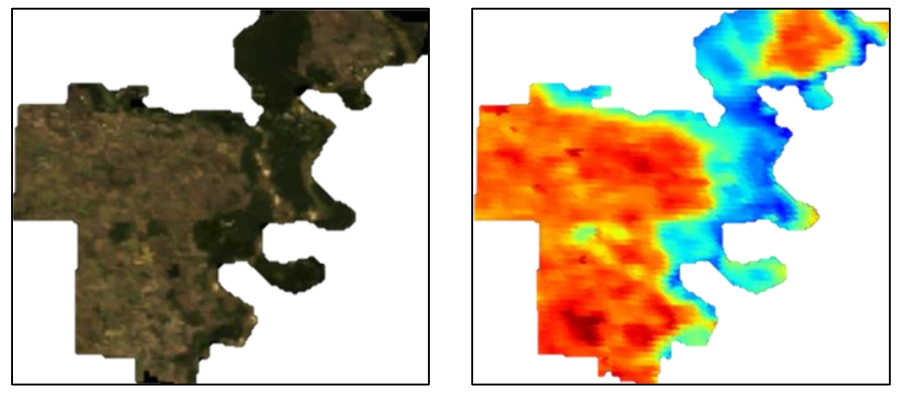

Figure 1: Marin County, CA (Left: RGB, Right: Temp.)

Time span: 2003-2013, sampled 46 times per year, of which 32 occur during the growing season

Ground Truth: USDA NASS Survey Data - Crop yields for soybean and corn [3]

Permutation Invariance

- **Key assumption:** position of pixels does not greatly affect average yield [1]
- Form of dimensionality reduction
- Given a time-band slice of a raw input image, form 32-bucket histogram
- **CNN input**: 32 buckets × 32 times × 9 bands

Methods

Training the Model

- Train on 2003-2012, validate on 2013
- **Training loss:** $L_2 = \frac{1}{2}\sum_{i=1}^{N}(\text{pred}_i - \text{real}_i)^2$
- **Validation error:** $\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\text{pred}_i - \text{real}_i)^2}$

	Ref.	Deep1	Deep2	Deep3	Deep4
CONV(128, 3, 1)	1	1	1	2	2
CONV(128, 3, 2)	1	1	1	1	1
CONV(256, 3, 1)	1	1	2	2	2
CONV(256, 3, 2)	1	1	1	1	1
CONV(512, 3, 1)	2	3	3	3	3
CONV(512, 3, 2)	1	1	1	1	1
CONV(1024, 3, 1)	0	0	0	0	1
FC(2048)	1	1	1	1	1

Figure 2: CNN Model Architectures

Each layer **CONV**$(\boldsymbol{c}, \boldsymbol{f}, \boldsymbol{s})$ represents a convolutional layer with $\boldsymbol{c}$ filters of size $\boldsymbol{f} \times \boldsymbol{f}$ with stride $\boldsymbol{s}$, followed by a ReLU nonlinearity, batch normalization layer, and dropout layer with keep probability $\boldsymbol{p}$.

Saliency Map Visualization

- Image i for crop c: $W_{ci} \in \mathbb{R}^{32\times32\times9}$
- Saliency map for W_{ci}: $S_{ci} = \frac{\partial \text{RMSE}}{W_{ci}}$
- Normalized map for W_{ci}: $N_{ci} = S_{ci} / \max_{j,k,\ell}\left|W_{cijk\ell}\right|$
- L_2 diff. between crops 1, 2: $\text{sqrt}\left(\sum_{i=1}^{K}\sum_{j,k,\ell}\left(W_{1ijk\ell} - W_{2ijk\ell}\right)^2 / K\right)$
- L_1 diff. between crops 1, 2: $\sum_{i=1}^{K}\sum_{j,k,\ell}\left|W_{1ijk\ell} - W_{2ijk\ell}\right| / K$

Differing Crops

- Rescale crop 1 pred. to crop 2 pred.: $\tilde{y}_{2i} = \frac{\hat{y}_{1i} - \mu_1}{\sigma_1}\cdot\sigma_2 + \mu_2$

Results

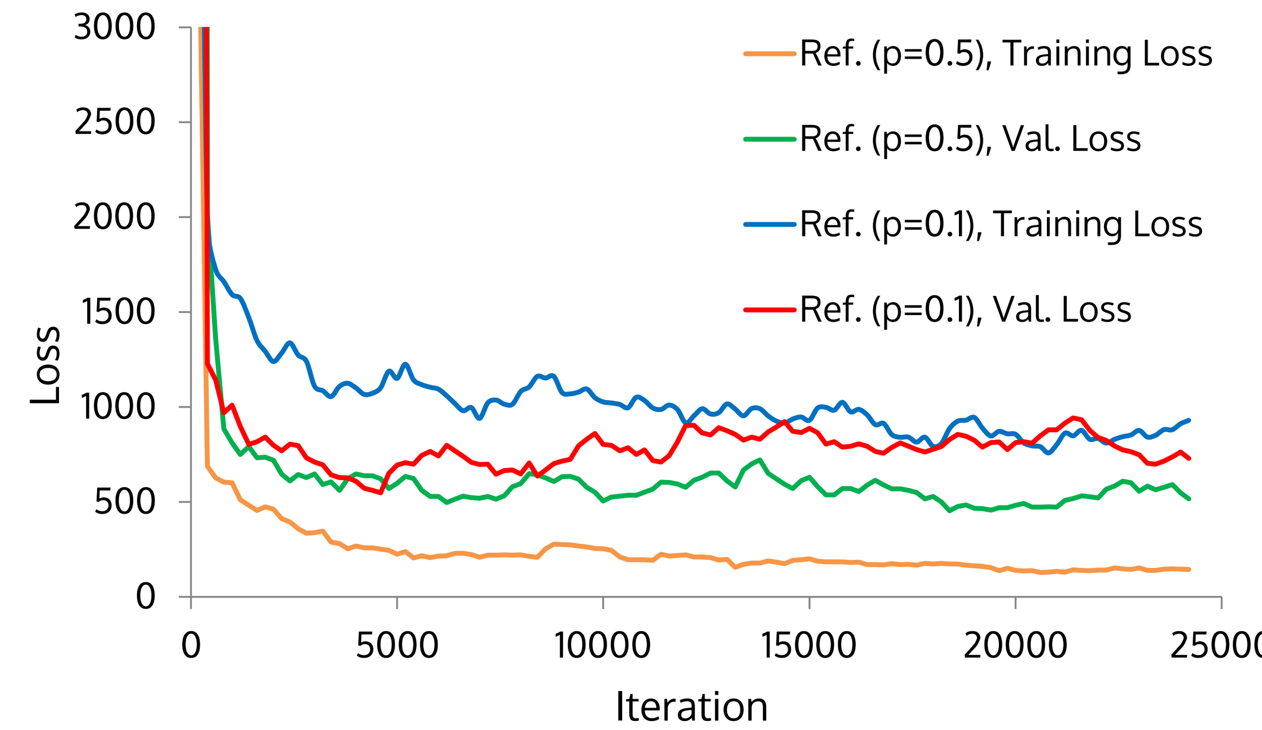

Figure 3: CNN Model Training & Validation Loss

Complex model ($p = 0.5$) causes overfitting; simple model ($p = 0.1$) doesn't overfit, but doesn't train well.

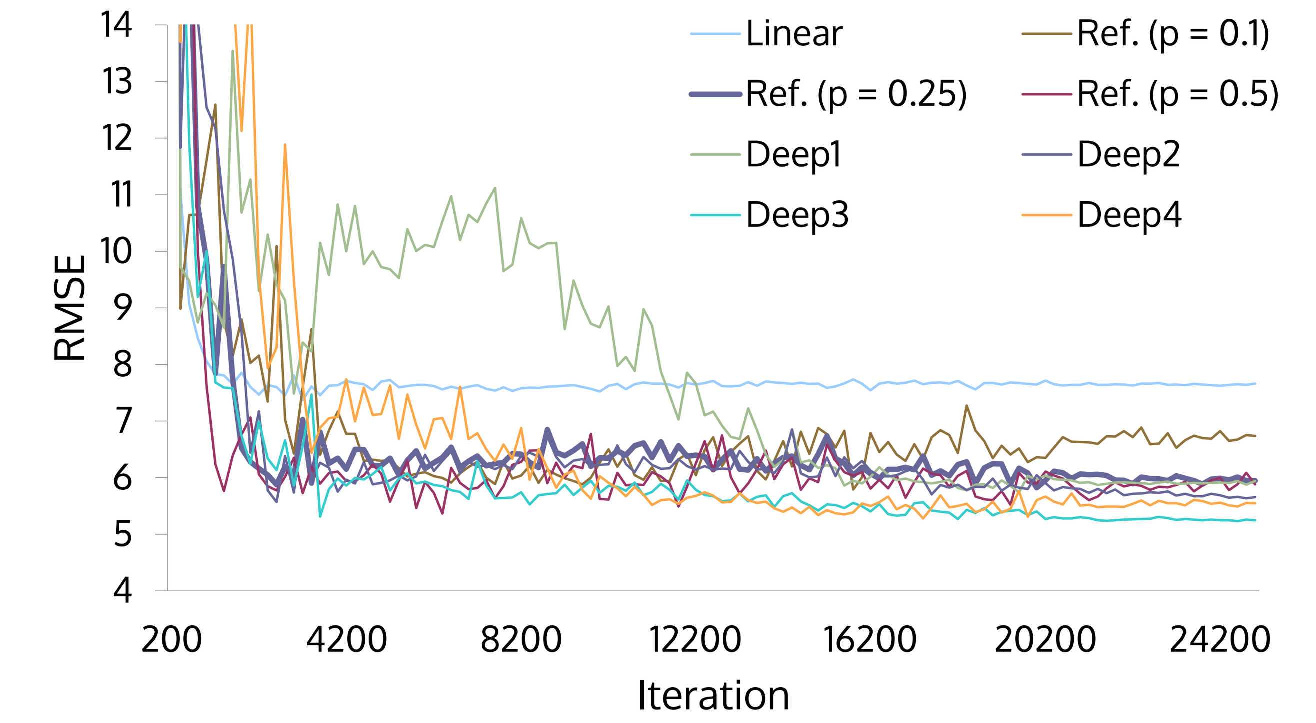

Figure 4: CNN Model Minimum RMSE

Validation set RMSE of various architectures over the course of training.

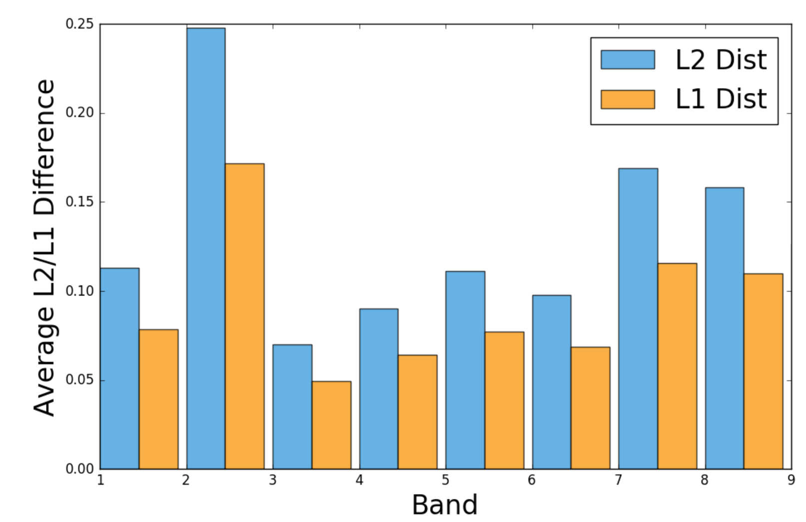

Figure 5: Relative Importance of Bands

Average distances between saliency maps computed for each band; the first two bands are key for crop discrimination

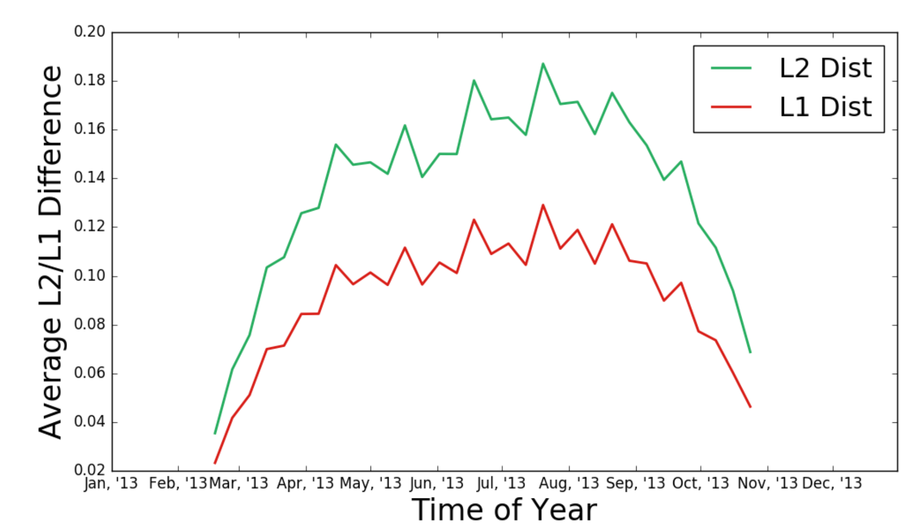

Figure 6: Relative Importance of Times

Average distances between saliency maps for each time slice; photos from May through Sept. are key for crop discrimination

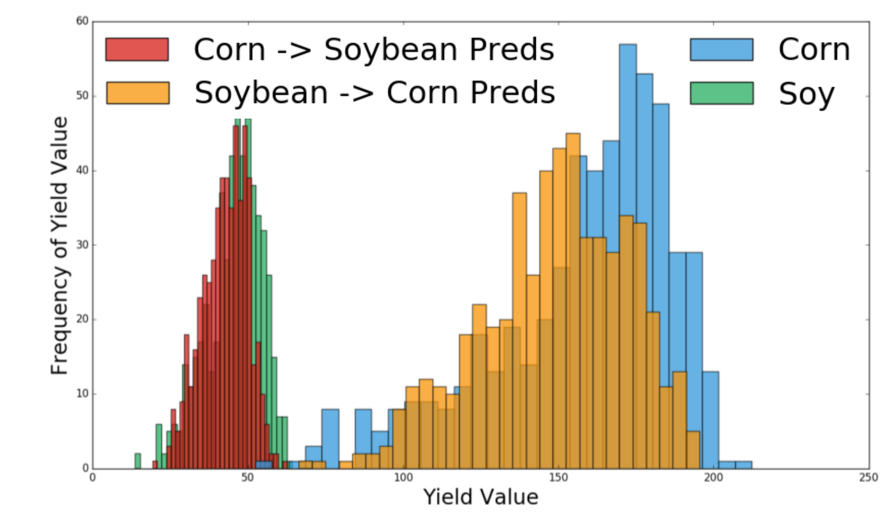

Figure 7: Dists. of Original Yields vs. Rescaled Predicted Yields

Actual dists. of crop yields compared with pred. yield dists. computed by rescaling predictions for one crop to predictions for the other crop

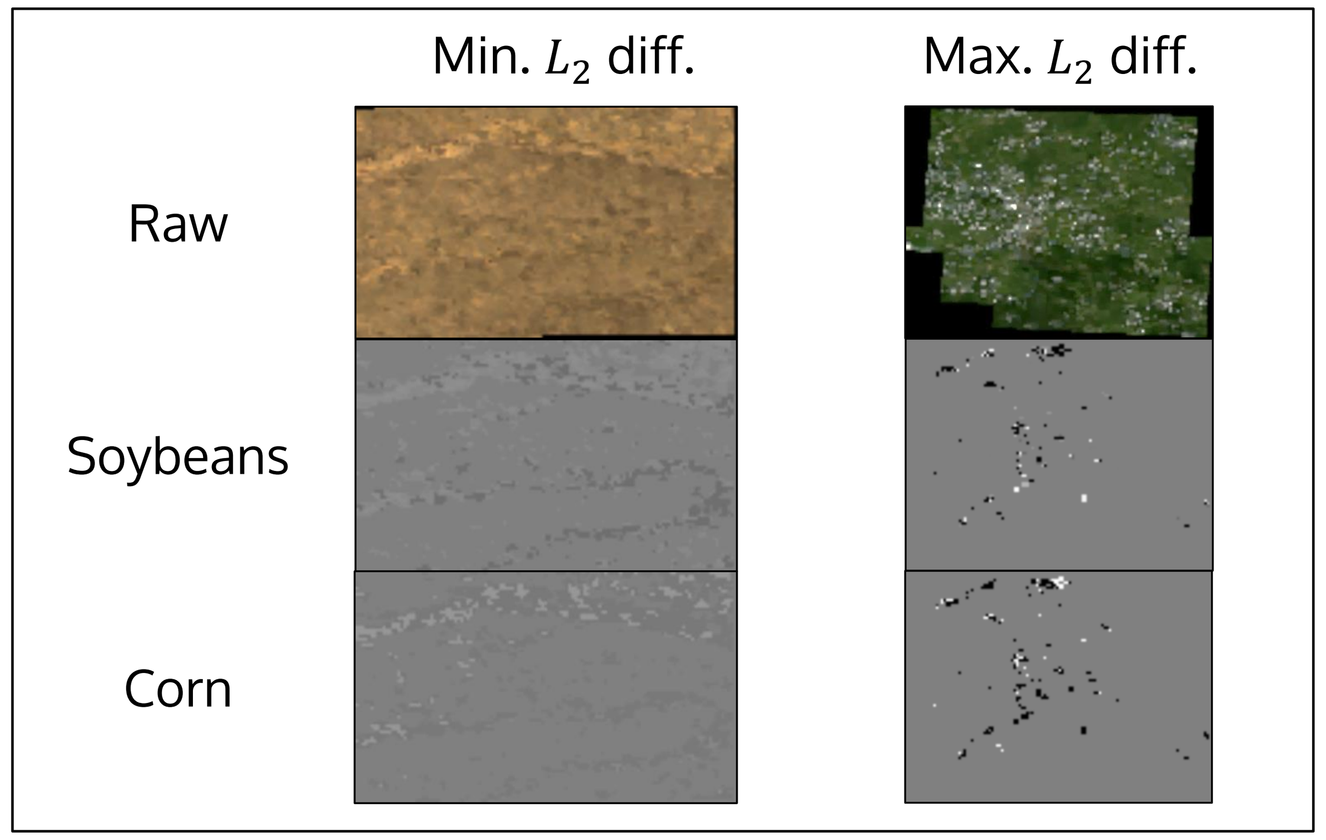

Figure 8: Most Similar & Dissimilar Saliency Maps

Brighter pixels positively impact prediction accuracy, darker pixels negatively impact prediction accuracy

Conclusions

- The model determines difference between corn and soybean farms, at least to some extent
- There is still signal to extract from the data since deeper models perform better

Future Work

- Test permutation invariance assumption by attempting to build a better model based on raw images