Multi-Stream CNN for Spatial Resource Allocation: a Crop Management Application

Alexandre Barbosa, Thiago Marinho, Nicolas Martin, Naira Hovakimyan
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA
barbosa4@illinois.edu

Abstract

Modeling the spatial structure of crop inputs is of great importance for accurate yield prediction. It is a fundamental step towards optimizing the spatial allocation of resources such as seed and fertilizer. We propose two distinct architectures of Multi-Stream Convolutional Neural Network (MSCNN) - Late Fusion (LF) and Early Fusion (EF) - to model yield response to seed and nutrient management. A study presents a comparison between proposed models with conventional 2D and 3D CNN architectures, and existing agronomy methods. The dataset used to train and test the models is constructed using on-farm experiment data from nine corn fields across the US together with multi-spectral satellite images. Results show that the MSCNN-LF achieved a 20% reduction of the prediction's mean squared error value when compared to a 3D CNN, and a 26% reduction when compared to a 2D CNN. An optimization algorithm uses the MSCNN-LF model's gradient to change the manageable inputs variables in a way the expected profit is maximized subject to resource constraints. It is shown that an increase of up to 5.2% on expected crop yield return is obtained when compared to usual management practices.

1. Introduction

Improving crop nutrient management is an essential step towards solving the food security problem [12]. Traditional farm management practices have led to excessive fertilization of crops, generating a surplus nutrient flow that pollutes the water system [7]. Additionally, this outdated management does not necessarily result in the highest possible yield. Despite ongoing research in modeling and data generation, the improvement of decision tools is yet to reach its full potential [3]. New methods are needed to take full advantage of new field technologies (e.g., variable-rate applicators, and yield monitors) and create decision aid tools to help farmers increase production while accounting for environmental impacts [1].

In order to create decision tools, descriptive and predictive models of the process are necessary. Many models have been proposed that relate environmental and management variables to crop yield [14]. They can be separated into statistical models [23, 16] and analytic crop models [2]. While analytic crop models are dynamic system simulations based on variables that may not be measurable by farmers, statistical models are constrained by the representativeness of training dataset.

On-Farm Precision Experimentation (OFPE) is often used to improve statistical models [21, 28] by generating site-specific representative data. However, even at a field scale, the spatial structure of environmental and management variables may affect the yield through events such as nutrient and water transportation [34]. Therefore, the spatial structure of environmental and treatment variables plays an important role when trying to create a predictive model for yield. Moreover, the interaction among different explanatory variables may depend on such spatial structures in a nonlinear way. Many spatial econometric models were developed to account for data's spatial structure on an effort to better describe the relations between explanatory and response variables [4]. For example, Generalized Least Squares (GLS) models [27] account for spatial structure using a geostatistical semivariogram while performing linear regressions. This approach, however, uses a fixed kernel to model the influence of neighboring data in a particular sample based only on the distance between them rather than on the spatial structure of the data. Hence, a way to extract relevant spatial features from the data is needed to overcome this limitation.

A similar spatial feature extraction problem is also present in image recognition software, where Convolutional Neural Networks (CNNs) have demonstrated significantly higher performance over other methods [18]. Convolutional layers can be trained to encode relevant visual (and here we can also use the term "spatial") features of varying complexity. In agriculture, applications of CNNs [24] usually

focus on disease [13, 20] and plant [19] classification, and image-based estimation, such as soybean leaf defoliation level [8]. To the best knowledge of the authors, the usage of CNNs to learn relevant spatial features from different explanatory variables, and model the interactions among them as a regression problem has not been explored yet.

This work proposes a Multi-Stream CNN (MSCNN) architecture to learn relevant spatial structures in different explanatory variables and use them to model the yield response to nutrient and seed rate prescriptions. Unlike multi-channel CNNs, the MSCNN does not stack inputs as channels and allows nonlinear combinations among them, which is a fundamental property of the phenomena being modeled [11]. We present two realizations (LF and EF) of the MSCNN combining the independent 2D input variables at different layer depths in the network. Data from nine corn fields across the US were used to test and compare the proposed architectures. Such fields are part of an OFPE with randomized nitrogen and seed rates prescription. The models are trained in a supervised fashion, tested, and a thorough comparative study is performed. Inspired by the back-propagation algorithm, the gradient of the MSCNN is used to derive a gradient of the profit with respect to the manageable inputs, taking network weights as constants. Finally, a gradient ascent with momentum algorithm uses the derived gradient to find the nitrogen and seed rate maps that maximize the expected yield for the next season.

The key contributions can be highlighted as: (i) a novel spatial dependency modeling for yield prediction, (ii) a framework for optimization using CNNs applied to patches of the inputs, and (iii) a comparative study of the performance of multi-stream CNNs. In our application, item (ii) leads to the nitrogen and seed prescription maps that maximize crop yield return. The remaining sections in this paper are organized as follows: Section 2 details the construction of the dataset used for testing the models; the MSCNN architecture is proposed and analyzed in Section 3, while Section 4 describes the optimization algorithm. Finally, Section 5 presents the experiments, and conclusions are made in Section 6.

2. Dataset Construction

This paper uses data from the Data Intense Farm Management (DIFM) project [6], recorded during the 2017 season. This experiment is the first large-scale OFPE with high variability in the nitrogen and seed rates. Its database contains georeferenced management and environmental data, as well as the resulting yield map. Each field in the database has on average 40 ha and 200 experimental units, represented by 85 m long and 18 m wide polygons. Nine fields were selected, from which six are rain-fed fields in Illinois (Fields 2, 4, 7, 8, and 9) and Ohio (Field 5), and three are irrigated fields in Nebraska (Fields 3 and 6) and Kansas (Field 1). Nitrogen and seed rates were randomly assigned from four different levels to each experimental unit in a field, except for field 1, where the nitrogen rate is constant. The explanatory variables in the database are nitrogen and seed rates prescription maps, elevation map, and soil's shallow electroconductivity, which is a proxy of the organic matter in the soil. A cloud-free Planet Labs [26] PSSE4 multispectral image (converted to a single grayscale channel) from each field and taken during pre-season is also used as an additional variable. This image may highlight different soil's composition and accumulation of water in the soil before planting. Yield data is used as the response variable and was collected by yield monitors during harvesting. The data was standardized variable-wise, and samples more than 3 standard deviations away from the mean were removed. Notice that all explanatory variables are available at the beginning of the season, so one can use the model for future optimization of nitrogen and seed rates before applying them. As each field had its data recorded using different equipment and software, a way to put each variable from all fields into the same support is first needed. The smallest unit of analysis is chosen to be a square with a side of five meters and is represented by a single element in a larger raster spanning the field. Variables stored as polygon data (prescription maps) were sampled for each element in the raster (using the mean method), while the ones stored as point data (elevation map and soil's electroconductivity) were first interpolated using kriging [27] and then rasterized to the desired support (5×5m cell). The satellite picture was simply resampled to the desired support since it is already a raster.

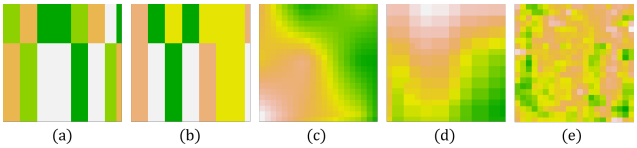


Figure 1. Rasters representing five different input variables at the same point: nitrogen rate (a), seed rate (b), elevation map (c), soil's electroconductivity (d), and satellite image (e).

The key assumption of this work is that the yield at a single unit of analysis depends on the spatial structure of observed explanatory variables around it. It is fair to assume however that the influence of neighbor data over a unit is limited to a certain range. So, by finding such range, models' complexity can be reduced without losing valuable data, which makes the model easier to train and more data-efficient. Therefore, a variogram containing the averaged variability from all explanatory variables is constructed to compute the range that better describes the data. For the selected fields the computed range is approximately 100m, indicating that no significant higher variability is obtained when considering a greater distance between two points.

Then, a sample used as input for the proposed MSCNN

model described in the next section is defined as a set of 21×21 elements rasters (one for each of the five input variables), spanning a square of 100×100m around the unit where the yield value is being considered (Figure 1). Then, to construct a dataset, rasters (with dimension equal to 21×21) are cropped from the original raster variables around each non-zero cell in the yield map. Such procedure results in a set of samples containing five rasters each and the value of yield at their respective center cell.

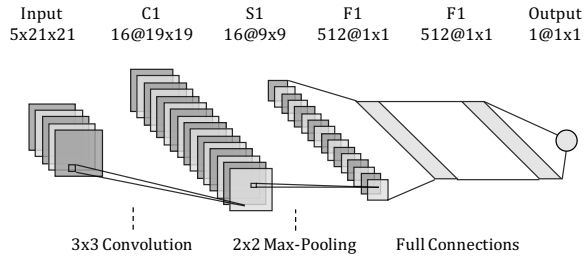
3. MSCNN Model

Information from five different fields' attributes is used to model yield response to nitrogen and seed rates management. Defining how such different sources of information are combined through the network has a direct impact on model's performance [10]. Previous papers have addressed this problem in different research areas. A CNN was proposed in [30] for flower grading, where three images are necessary for full flower's description, and authors merged the result of independent convolutions from each image before performing sub sequential convolutions. Different architectures were also proposed to identify human actions in video data. A multi-stream architecture in [29], where inputs are combined late in the network have demonstrated better performance over models that used stacked frames as inputs [15]. A 3D convolutional neural network architecture was proposed by [31], where 3D kernels were convolved with adjacent input frames, demonstrating results comparable to the state of the art techniques. In this section we introduce the Early Fusion and Late Fusion realizations of the MSCNN and compare them to standard CNNs.

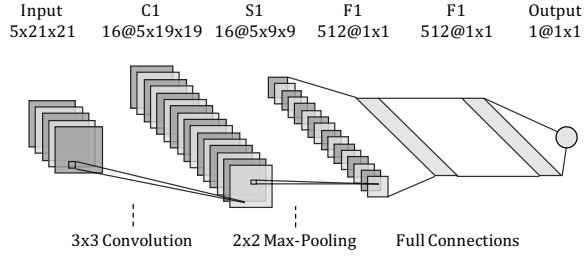
3.1. Architectures

In this work four different CNN architectures have their performance tested and compared: (i) a stacked model (ST) that combines inputs by stacking them as a multi-channel image for further 2D convolutions (Figure 2(a)), (ii) a 3D CNN using the same stack as input (Figure 2(b)), (iii) a multi-stream early fusion (MSCNN-EF) network that concatenates flattened max-pooling layers from independent convolutional layers before the fully connected layers (Figure 2(c)), and (iv) a multi-stream late fusion (MSCNN-LF) network that concatenates the outputs from single neurons coming from independent convolutional layers (Figure 2(d)).

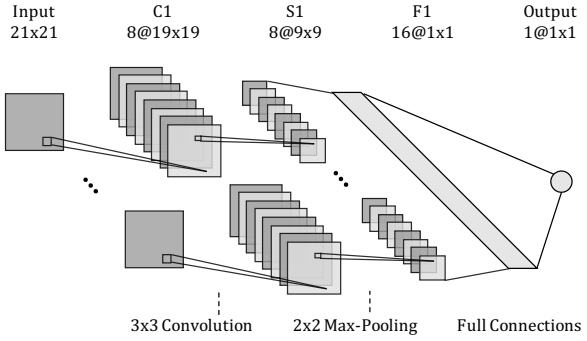
To create a baseline, most commonly used models [25] for yield prediction were also tested together with all four architectures described. Such models are a multiple linear regression model (MLR), a fully connected artificial neural network (FC), which has received significant attention from agronomy community over the last decade [9][32][5], a support vector machine (SVM), and a random forest (RF).



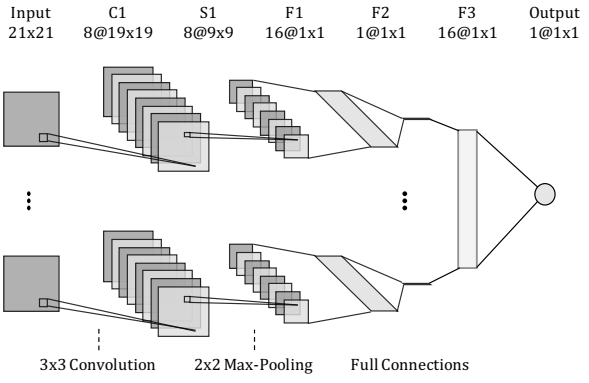
(a) CNN-ST architecture.



(b) CNN-3D architecture.



(c) MSCNN-EF architecture.



(d) MSCNN-LF architecture.

Figure 2. Comparison between architectures

Hyperparameters of each baseline model were fine tuned to obtain the best performance for comparison.

Table 1. Crossvalidation averaged Mean Squared Error (MSE) over test dataset, yield standard deviation, and yield mean.

Field	MSE								Yield [Kg/ha]	
	MLR	FC	LF	EF	3D	ST	SVM	RF	Stdv.	Mean
1	1.53	0.97	**0.66**	0.73	0.75	0.69	0.86	0.92	3240	12500
2	1.29	0.88	**0.83**	0.86	0.87	0.88	0.86	0.90	2290	10700
3	1.90	0.63	**0.58**	0.58	0.59	0.60	0.86	0.59	1230	14400
4	1.03	0.76	**0.75**	0.78	0.77	0.77	0.77	0.76	900	12200
5	0.74	0.72	**0.70**	0.72	0.75	0.73	0.76	0.70	1360	14500
6	1.09	0.51	**0.48**	0.51	0.53	0.56	0.58	0.52	1140	14700
7	0.75	0.72	**0.69**	0.70	0.71	0.73	0.76	0.73	1150	15700
8	1.11	0.94	0.94	0.94	0.94	0.94	**0.89**	0.96	2267	14100
9	1.10	0.69	**0.63**	0.65	0.66	0.66	0.67	0.63	1140	12600
Avg.	1.17	0.76	**0.70**	0.72	0.73	0.73	0.78	0.75	1635	13489

3.2. Evaluation

Yield response to nutrient management also depends on other environmental factors and management practices with whole field impact (e.g., solar radiation, planting date, seed genetics, among others). These factors vary from one field to another while being constant within the same field. Using site-specific data is a way to reduce the input space and increase the representativeness of the data aiming better recommendations to each farmer individually. So, we train and test a different model for each of the nine fields (described in Section 2) rather than creating a single model to work in all fields. However, data within a single field is often spatially autocorrelated and may lead to model over-fitting depending on the way the test set is chosen. For instance, a random partition of the data results in training and test sets with very similar samples, which overcasts the generalization power of the model. So, we spatially partition the data to account for this problem as proposed by [33].

Data from each field was spatially partitioned in five stripes perpendicularly to the longest dimension of the field to maximize the distance between samples from two different partitions. Three stripes (60% of the data) were used for training the model, one (20% of the data) for validation, and one (20% of the data) for tests. A grid search was performed to define hyperparameters such as the number of filters in the convolutional layers, the number of layers and neurons in the fully connected layers, and the dropout regularization probability for all models. Each architecture was trained over the same database using Adam optimizer [17], and the validation set was used online as an early stop criterion (allowing at most eight consecutive iterations of increase in the validation loss) to avoid over-fitting the data. The loss function is given by the Mean Squared Error (MSE) between yield predictions and true yield values at every raster's position in the batch.

Cross-validation was used to evaluate the model using five folds according to the spatial partitioning. Table 1 shows the averaged MSE over the five test sets for each model in each field. Notice that since yield data is standardized, the MSE value is a fraction of the standard deviation from the original yield values.

3.3. Model Analysis

The MSCNN-LF model is the one with lowest MSE value over the test dataset in eight out of nine tested fields. Field three is the exception, showing a slightly higher MSE value when compared to the other CNN models. The better performance of MSCNN-LF may be explained by the way each architecture combines the input variables in the network as described next.

The CNN-ST model linearly combines the inputs element-wise at the very first convolutional layer, suppressing any possible nonlinear interaction between different attributes in the field. To make this clear, suppose a $n \times n$ kernel convolved with an L-channel input, and let w_{ij} be the kernel's weights, y_{uv} the convolution output at position (u, v), and x_{kij} be the input value at channel k that is aligned with (i, j) kernel's position. Then we have that y_{uv} is given by (1)

$$y_{uv} = \sum_{k=1}^{L} \sum_{i=1}^{n} \sum_{j=1}^{n} (w_{ij} x_{kij}). \quad (1)$$

It is evident the output of the convolutional layer only contains a linear combination of the values from each input, and that the information from each separate input is not preserved. As a consequence, any nonlinear interaction between the input variables will not be modeled at the fully connected layers. One possible solution to overcome this limitation is to convolve the filter input-wise so the information from each input is carried separately through the convolutional layer. In fact, the proposed CNN-3D architecture considers the stacked different inputs as a third dimension to convolve the filter. However, even if this architecture solves the problem of carrying the information from each input and combining it in the fully connected layers,

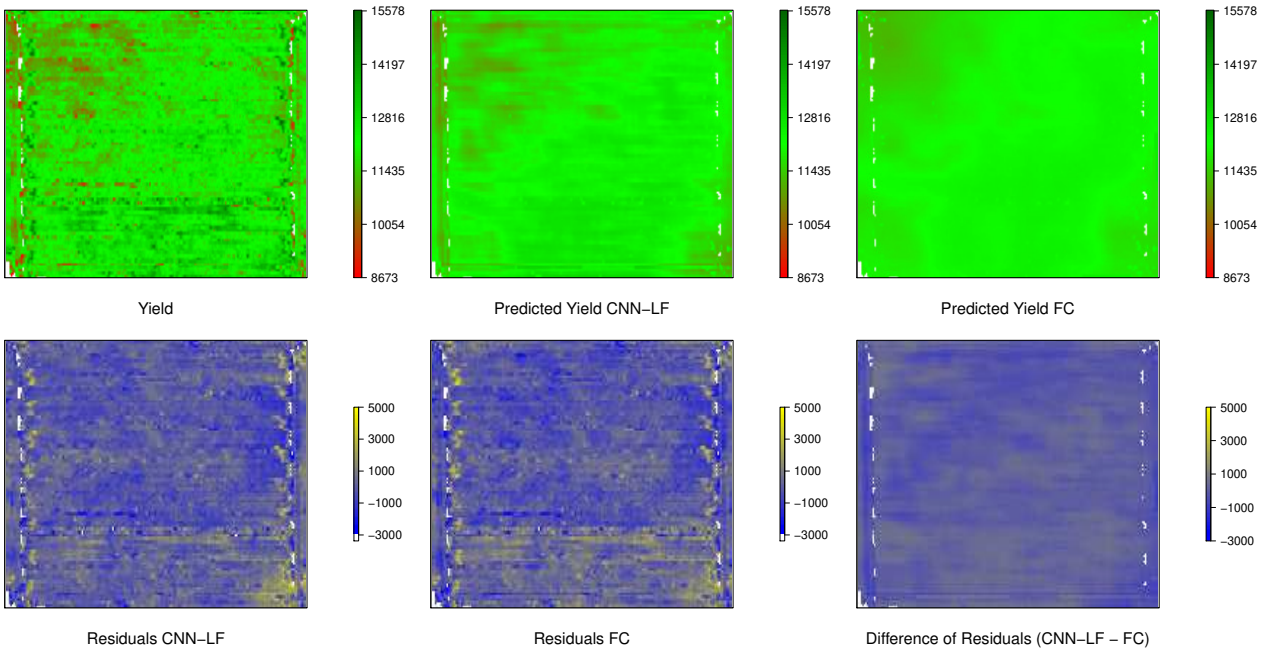


Figure 3. Groundtruth Yield map, predicted yield map using MSCNN-LF, predicted yield map using FC, residuals for both models, and difference of residuals for field 4

it introduces a new undesired feature. By using this architecture, the output of the convolutional layer increases with the number of inputs, since the filters are convolved with all inputs independently. This may lead to an unnecessary increase in the number of parameters in the network, if we assume that the same filter may not be relevant (as feature extractors) for all input variables, resulting in a model with lower data efficiency.

The multi-stream architecture addresses both problems discussed above. It performs independent convolutions with the inputs while using a different set of filters for each one. When comparing the LF and EF models it is reasonable to say that LF focuses on better feature extraction from inputs while EF can model more complex interaction between variables. In addition, by reducing the dimension of each input before combining them, the LF model becomes easier to train, leading to higher data efficiency.

To construct a predicted yield map, the yield value at each position in the output raster is obtained using the MSCNN-LF model and the input maps. Figure 3 shows a qualitative comparison between predicted yield maps using the MSCNN-LF and FC model in field four. It is possible to observe that the spatial structure of the data is better captured by the MSCNN-LF model, while FC results in a "blurrier" map. The difference between the residual maps from both models shows a spatial structure that resembles the spatial distribution observed on the true yield map, indicating that the main difference between both predictions lies indeed on the spatial structure of the data. Moreover, one can observe that most of the higher values in the residuals maps are located closer to field's borders. The data in such regions are expected to have more noise due to yield monitoring system errors caused by the change in harvester speed and direction, and cannot be explained by the selected explanatory variables.

4. Optimization Algorithm

Choosing the optimal rates of nitrogen and seeds is one of the main reasons to create a good yield prediction model. The complexity of optimizing these rates depends on the chosen model, and the MSCNN-LF introduces a much larger input search space since it maps patches of the input maps to a predicted yield value at a single point in the field. Also, such patches overlap between themselves, making the sequential optimization of each point in the field not possible. Then, we present an optimization framework for the MSCNN-LF model to find the spatial allocation of nitrogen and seeds that result in the maximum expected profit to the farmer. We propose a gradient ascent with momentum algorithm under the assumption that this is a non-convex optimization problem. The first step is to derive the gradient of yield return (total yield value discounting nitrogen and seeds costs) with respect to nitrogen and seed rate maps.

Let $A, B \in \mathbb{R}^{m \times n}$ be the matrices representing the applied nitrogen and seed rates maps respectively. Now, let $\mathcal{N}, \mathcal{S} \in \mathbb{R}^{l \times l}$ be sets of matrices cropped from matrices A and B such that $l < n < m$. We define a function $\tilde{f}(N, S) : \mathbb{R}^{l \times l} \times \mathbb{R}^{l \times l} \longrightarrow \mathbb{R}$, such that $N \in \mathcal{N}$, and

$S \in \mathcal{S}$, that represents the MSCNN-LF model mapping a patch from nitrogen and seed rate maps to a yield value, taking environmental attributes as constants. The total crop yield is then given by $f(A,B) = \sum^i \tilde{f}(N_i, S_i)$, where i is the index of each element in the yield map. Then, we want to derive $\nabla f(A,B)$ as in (2)

$$\nabla f(A,B) = \left(\frac{\partial \sum^i \tilde{f}(N_i,S_i)}{\partial A}, \frac{\partial \sum^i \tilde{f}(N_i,S_i)}{\partial B} \right). \quad (2)$$

Now, define a cropping function $g(C,i) : \mathbb{R}^{m\times m} \times \mathbb{N} \longrightarrow \mathbb{R}^{l\times l}$ that crops a $l \times l$ matrix from C around the element indexed by i. We also define a zero padding function $z(c,i) : \mathbb{R}^{l\times l} \times \mathbb{N} \longrightarrow \mathbb{R}^{m\times n}$ that centers the original $l \times l$ matrix at element i in a $m \times n$ matrix, and completes the remaining elements with zero. Then, we have that:

$$\begin{aligned} \frac{\partial \sum^i \tilde{f}(N_i,S_i)}{\partial A} &= \sum^i z\left(\frac{\partial \tilde{f}(N_i,S_i)}{\partial N_i}, i \right) \\ &= \sum^i z\left(\begin{bmatrix} \frac{\partial \tilde{f}(N_i,S_i)}{\partial N_{i_{11}}} & \cdots & \frac{\partial \tilde{f}(N_i,S_i)}{\partial N_{i_{1l}}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \tilde{f}(N_i,S_i)}{\partial N_{i_{l1}}} & \cdots & \frac{\partial \tilde{f}(N_i,S_i)}{\partial N_{i_{ll}}} \end{bmatrix}, i \right). \end{aligned}$$

Similarly,

$$\frac{\partial \sum^i \tilde{f}(N_i,S_i)}{\partial B} = \sum^i z\left(\begin{bmatrix} \frac{\partial \tilde{f}(N_i,S_i)}{\partial S_{i_{11}}} & \cdots & \frac{\partial \tilde{f}(N_i,S_i)}{\partial S_{i_{1l}}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \tilde{f}(N_i,S_i)}{\partial S_{i_{l1}}} & \cdots & \frac{\partial \tilde{f}(N_i,S_i)}{\partial S_{i_{ll}}} \end{bmatrix}, i \right).$$

Fortunately, all elements of $\frac{\partial \tilde{f}(N_i,S_i)}{\partial N_i}$ and $\frac{\partial \tilde{f}(N_i,S_i)}{\partial S_i}$ are easily obtained from the MSCNN-LF model's gradients computed for the backpropagation algorithm. With (2) derived, our final goal is to maximize the crop yield return discounting the costs from nitrogen and seeds, which can be formulated as:

$$\max_{A,B} \left(p_V f(A,B) - p_N \sum^{u,v} A_{(u,v)} - p_S \sum^{u,v} B_{(u,v)} \right) \quad (3)$$

$$\begin{aligned} \text{subject to: } & N_{min} \leq A_{uv} \leq N_{max}, \forall\,(u,v) \\ & S_{min} \leq B_{uv} \leq S_{max}, \forall\,(u,v), \end{aligned}$$

where p_V, p_N, and p_S are the prices per smallest unit area of corn, nitrogen, and seed, respectively, and the rates of nitrogen and seed are bounded by minimum and maximum values. The boundary values come from the rates applied during the on-farm experiment, since the MSCNN-LF model is not trained with values above or below them. Finally, let P_N and P_S be $m \times n$ matrices containing all their elements equal to p_N and p_S respectively. Then we write the gradient of (3) as:

$$\begin{aligned} \nabla Y(A,B) = \Bigg(& p_V \sum^i z\left(\frac{\partial \tilde{f}(N_i,S_i)}{\partial N_i}, i \right) - P_N\,, \\ & p_V \sum^i z\left(\frac{\partial \tilde{f}(N_i,S_i)}{\partial S_i}, i \right) - P_S \Bigg). \end{aligned} \quad (4)$$

In order to ensure our optimization algorithm is based on a model where the manageable variables (i.e. nitrogen and seed rates) are really relevant (and not vanished by more relevant environmental inputs), a sensitivity index based on partial derivatives [22] is obtained by (5):

$$\zeta_i = \frac{1}{L} \sqrt{ \sum_{k=1}^{L} \left(\frac{\partial \tilde{f}}{\partial i_k}(N_k, S_k) \right)^2 }, \quad (5)$$

where L is the number of available training samples, and i is the input label. In our experiment, i indexes the set $\{\text{NR}, \text{SR}, \text{Elev.}, \text{EC.}, \text{Soil}\}$. As the value of the partial derivative depends on the point of the input space it is being evaluated, our sensitivity index takes in account the partials for every sample in our training dataset. Notice that each element i_k in the input map i has its gradient dependent on the cropped inputs (Figure 4), with N_k and S_k being the only manageable inputs. Table 2 shows the index ζ for each input for the nine studied fields, revealing that the model is in fact sensitive to the selected manageable variables. Field 1 is an exception, showing an index of zero for nitrogen rate, which is expected since a constant rate was used for this field during the OFPE.

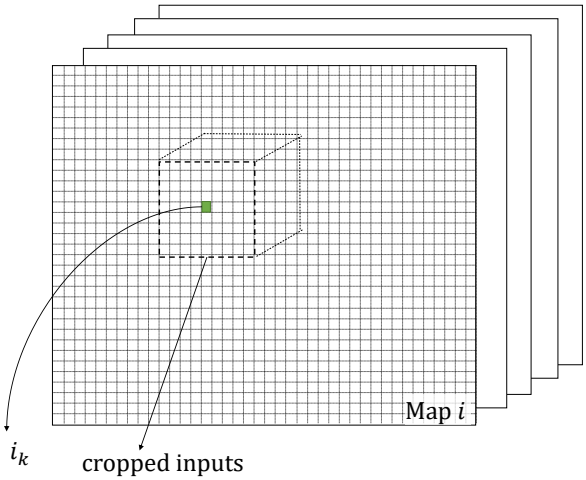


Figure 4. Region for gradient evaluation of i_k.

Table 2. Sensitivity index ζ for each input.

Field	NR	SR	Elev.	EC.	Soil
1	0.00	2.86	0.84	0.14	0.45
2	1.98	1.54	1.22	0.65	0.71
3	1.22	5.23	0.11	0.25	0.56
4	0.86	1.11	1.48	0.74	0.31
5	0.76	0.22	0.83	0.47	0.38
6	2.81	3.51	0.17	0.29	0.29
7	1.45	0.92	0.38	0.22	0.46
8	0.63	0.55	2.32	0.02	0.17
9	1.16	0.50	0.04	0.32	0.06

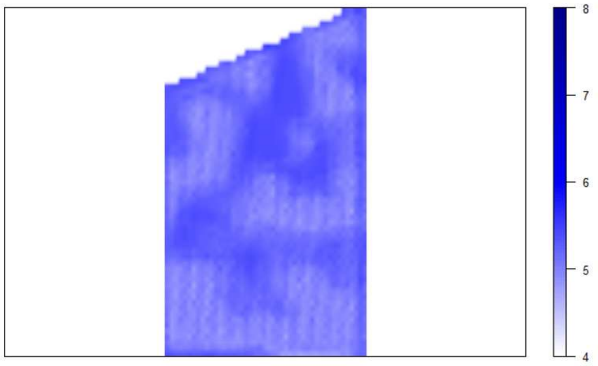


Figure 5. Optimized nitrogen map (UAN28/acre) showing low response areas on field 7.

With the model's gradient derived in (4), a gradient ascent with momentum algorithm is used in an attempt to find the global maxima. Maps A and B are initialized with the constant values usually applied by farmers and then updated according the algorithm based on the gradient in (4).

5. Experiments and Results

Experiments were conducted with the same nine fields used in Section 3 to demonstrate the monetary potential of the proposed optimization algorithm. The total profit (i.e. yield discounting costs of nitrogen and seed) is estimated for the initial condition (usual rates applied by the farmer) and for the resulting maps. Table 3 shows the total percent change on expected profit, nitrogen, seed, and yield values after optimization.

Table 3. Change in profit, nitrogen, seed and yield after optimization.

Field Id	Profit	Nitrogen	Seed	Yield
1	3.6%	0%	10.5%	4.6%
2	5.2%	-17.3%	-18.2%	-0.9%
3	3.8%	18.0%	13.4%	5.6%
4	4.2%	-20.8%	-11.4%	-0.2%
5	2.7%	-12.8%	-11.8%	0.0%
6	1.8%	10.1%	4.6%	2.4%
7	2.2%	-30.4%	-17.6%	-0.7%
8	3.7%	-42.3%	-13.0%	0.1%
9	4.1%	-12.7%	-11.7%	-0.2%

Results show a 3.5% average increase in the expected profit, with field 2 going up to 5.2%. It can be observed that the profit increase is obtained through different strategies, depending on the field. For some fields, the rate of nitrogen and seeds was increased to generate more yield, while others showed to be unresponsive to high rates of nutrients. In such cases, the algorithm was able to select areas where nutrient could be reduced without reducing the yield. Field 7 is a good example of how the algorithm reduced the nitrogen rate at unresponsive areas. Figure 5 shows the final nitrogen map for this field.

With tested values for the step size and momentum terms, the algorithm converged within an average of 30 iterations. Additional experiments were made initializing the input maps with random values. In such experiments, the optimization algorithm also converged within 30 iterations to very similar maps to the ones obtained with different initial conditions. These results indicate that the algorithm is able to find either the global maxima or a flat local maxima. The second case could be considered even better than a sharp global maxima since it is more robust to undesired variations during the nitrogen application process.

Notice that the results depend on the initial conditions and are based on the model's predictions, rather than on real experiments. Nevertheless, this experiment aims to evaluate the optimization framework regarding its ability to drive the output of the yield prediction model to maximize an objective function. The performance of this algorithm in real experiments will depend on the model's accuracy, which was evaluated in Section 3 and showed to be better than other machine learning methods.

6. Conclusions

A novel spatial dependency model for yield prediction based on pre-season treatments and environmental variables was proposed. Such model leverages a multi-stream architecture of CNN in order to model nonlinear dependencies among input variables, while accounting for variable-wise feature extraction. We provided experimental evidence to show the superior performance of the LF realization of the MSCNN, achieving a reduction up to 26% on the MSE value when compared to a conventional 2D CNN with stacked input channels. The MSCNN-LF architecture is also appropriate for using transfer learning since the multi-stream approach works as a feature extractor for each individual input. So, a good performance is expected when using a pre-trained MSCNN-LF to model a field even when the dataset is limited.

We presented an optimization framework to find the manageable inputs of the MSCNN-LF that maximize the expected yield return. When compared to traditional farming practices, this framework showed an increase in crop yield return discounting the costs for all fields. The benefits of decreasing nitrogen while maintaining yield are of great importance to subdue environmental impact caused by water pollution. Future work includes investigating the proposed optimization framework as a CNN analysis tool.

Acknowledgment

This work was supported by the 2020 University of Illinois Center of Digital Agriculture Seed Funding.

References

[1] Lynette K. Abbott and Daniel V. Murphy. *What is Soil Biological Fertility?*, pages 1–15. Springer Netherlands, Dordrecht, 2007. 1

[2] P.K. Aggarwal. Uncertainties in crop, soil and weather inputs used in growth models: Implications for simulated outputs and their applications. *Agricultural Systems*, 48:361–384, 1995. 1

[3] John M. Antle, James W. Jones, and Cynthia E. Rosenzweig. Next generation agricultural system data, models and knowledge products: Introduction. *Agricultural Systems*, 155:186–190, 2017. 1

[4] Badi H. Baltagi. *A companion to theoretical econometrics*. Blackwell, 2003. 1

[5] E. Betiku and A. E. Taiwo. Modeling and optimization of bioethanol production from breadfruit starch hydrolyzate vis-à-vis response surface methodology and artificial neural network. *Renewable Energy*, 74:87–94, 2015. 3

[6] David S Bullock, Maria Boerngen, Haiying Tao, Bruce Maxwell, Joe D Luck, Luciano Shiratsuchi, Laila Puntel, and Nicolas F Martin. The data-intensive farm management project: Changing agronomic research through on-farm precision experimentation. *Agronomy Journal*, 111(6):2736–2746, 2019. 2

[7] S. R. Carpenter, N. F. Caraco, D. L. Correll, R. W. Howarth, A. N. Sharpley, and V. H. Smith. Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications*, 8(3):559–568, 1998. 1

[8] Lucas Abreu da Silva, Patrik Olã Bressan, Diogo Nunes Gonçalves, Daniel Matte Freitas, Bruno Brandoli Machado, and Wesley Nunes Gonçalves. Estimating soybean leaf defoliation using convolutional neural networks and synthetic images. *Computers and Electronics in Agriculture*, 156:360–368, 2019. 2

[9] S.S. Dahikar and V.S. Rode. Agricultural crop yield prediction using artificial neural network approach. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, 2:683–686, 2014. 3

[10] Christoph Feichtenhofer, Axel Pinz, and Andrew Zisserman. Convolutional two-stream network fusion for video action recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1933–1941, 2016. 3

[11] Michael D Frank, Bruce R Beattie, and Mary E Embleton. A comparison of alternative crop response models. *American Journal of Agricultural Economics*, 72(3):597–603, 1990. 2

[12] H. Charles J. Godfray, John R. Beddington, Ian R. Crute, Lawrence Haddad, David Lawrence, James F. Muir, Jules Pretty, Sherman Robinson, Sandy M. Thomas, and Camilla Toulmin. Food security: The challenge of feeding 9 billion people. *Science*, 327(5967):812–818, 2010. 1

[13] Kamlesh Golhani, Siva K. Balasundram, Ganesan Vadamalai, and Biswajeet Pradhan. A review of neural networks in plant disease detection using hyperspectral data. *Information Processing in Agriculture*, 5(3):354–371, 2018. 2

[14] James W. Jones, John M. Antle, Bruno Basso, Kenneth J. Boote, Richard T. Conant, Ian Foster, H. Charles J. Godfray, Mario Herrero, Richard E. Howitt, Sander Janssen, Brian A. Keating, Rafael Munoz-Carpena, Cheryl H. Porter, Cynthia Rosenzweig, and Tim R. Wheeler. Brief history of agricultural systems modeling. *Agricultural Systems*, 155:240–254, 2017. 1

[15] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1725–1732, 2014. 3

[16] Monisha Kaul, Robert L. Hill, and Charlie Walthall. Artificial neural network for corn and soybean yield prediction. *Agricultural Systems*, 85:1–18, 2005. 1

[17] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 4

[18] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012. 1

[19] Sue Han Lee, Chee Seng Chan, Paul Wilkin, and Paolo Remagnino. Deep-plant: Plant identification with convolutional neural networks. In *IEEE international conference on image processing (ICIP)*, pages 452–456, 2015. 2

[20] Yang Lu, Shujuan Yi, Nianyin Zeng, Yurong Liu, and Yongdong Zhang. Identification of rice diseases using deep convolutional neural networks. *Neurocomputing*, 267:378–384, 2017. 2

[21] A Meyer-Aurich, N Muhammad, and R Herbst. On-farm experimentation for identification of site-specific fertilization potentials. In *Proceedings of the 9th International Conference on Precision Agriculture*, pages 20–23, 2008. 1

[22] Grégoire Montavon, Wojciech Samek, and Klaus-Robert Müller. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73:1–15, 2018. 6

[23] S.L. Osborne, J.s Schepers, D Francis, and Michael R. Schlemmer. Use of spectral radiance to estimate in-season biomass and grain yield in nitrogen- and water-stressed corn. *Crop science*, 42:165–171, 2002. 1

[24] Diego Inacio Patricio and Rafael Rieder. Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Computers and Electronics in Agriculture*, 153:69–81, 2018. 1

[25] Amy Peerlinck, John Sheppard, and Bruce Maxwell. Using deep learning in yield and protein prediction of winter wheat based on fertilization prescriptions in precision agriculture. In *Proceedings of the 14th International Conference on Precision Agriculture*, pages 1–13, 2018. 3

[26] CA. Planet Team, San Francisco. Planet application program interface: In space for life on earth. https://api.planet.com, 2017. 2

[27] R. Plant. *Spatial Data Analysis in Ecology and Agriculture Using R.* Boca Raton: CRC Press, 2012. 1, 2

[28] Divina Gracia P Rodriguez, David S Bullock, and Maria A Boerngen. The origins, implications, and consequences of yield-based nitrogen fertilizer management. *Agronomy Journal*, 111(2):725–735, 2019. 1

[29] Karen Simonyan and Andrew Zisserman. Two-stream convolutional networks for action recognition in videos. In *Advances in neural information processing systems*, pages 568–576, 2014. 3

[30] Yu Sun, Lin Zhu, Guan Wang, and Fang Zhao. Multi-input convolutional neural network for flower grading. *Journal of Electrical and Computer Engineering*, 2017:1–8, 2017. 3

[31] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. Learning spatiotemporal features with 3d convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 4489–4497, 2015. 3

[32] Sankar Vani, Rajeev Kumar Sukumaran, and Sivaraman Savithri. Prediction of sugar yields during hydrolysis of lignocellulosic biomass using artificial neural network modeling. *Bioresource technology*, 188:128–135, 2015. 3

[33] Slobodan Vucetic, Tim Fiez, and Zoran Obradovic. A data partitioning scheme for spatial regression. In *IJCNN'99 International Joint Conference on Neural Networks*, volume 4, pages 2474–2479. IEEE, 1999. 4

[34] Yujian Yang, Jianhua Zhu, Xueqin Tong, and Dianchang Wang. The spatial pattern characteristics of soil nutrients at the field scale. In Daoliang Li and Chunjiang Zhao, editors, *Computer and Computing Technologies in Agriculture II, Volume 1*, pages 125–134, Boston, MA, 2009. Springer US. 1