Proceedings of the 41st Chinese Control Conference
July 25-27, 2022, Hefei, China

BioPM: Mixer for Point Cloud Based Biomass Prediction

Yong Lei, Hongbin Ma
Beijing Institute of Technology, Beijing 100081, P. R. China
E-mail: 3120200918@bit.edu.cn

Abstract: AGB(Above-Ground Biomass) is crucial trait relevant to agricultural production and study. Benefiting from the availability of field point cloud scanned by LiDAR, it's possible to use a non-destructive and high-throughput method for predicting AGB instead of laborious and destructive methods. Inspired by deep learning methods in 3D object detection by grouping point cloud and Mixer structure achieves great performance on 2D computer vision tasks, we propose an end-to-end prediction network BioPM, which combines both advantages based on the upward growth characteristics of wheat. Our BioPM consists of two modules: 1) a feature encoding module to group point cloud as pillars and extract point-wise features of pillars; and 2) a mixer module to extract pillar-wise features and output predictions by using only MLP. Experiments on the public dataset show that our BioPM prediction outperforms non-deep learning SOTA methods and other deep learning methods.

Key Words: Biomass prediction, Mixer, Point cloud

1 Introduction

With the enduring effects of the COVID-19 pandemic, food supply security is under enormous pressure. According to the Food and Agriculture Organization of the United Nations, the prevalence of undernourishment rose from 8.4 percent to 9.9 percent in just one year, after remaining unchanged for five consecutive years[1].

One of the directions to enhance the production of crop with new technologies is to correctly predict biomass, which is important for farmers and researchers to know about the amount and optimize crop performance by timing of actions needed[2]. Moreover, biomass have a strong correlation with studying biodiversity[3] and increasing crop yied[4].

Estimating biomass in most traditional methods need to cut the culms of an experimental plot for a specific portion and weighing the plant material to a constant weight after drying in an oven[5]. This method has three main disadvantages: variable precision caused by sampling deviation, few samples caused by sampling destructiveness, and high time and labor cost caused by complex sampling processing. Besides, modern field trials or breeding trials require high-throughput screening[6].

In recent years, LiDAR-based methods have developed rapidly in agriculture[7–9]. Fig. 1 shows the schematic of how to get the point cloud of wheat. The advantages of non-destructive sampling and automation make 3D computer vision widely used in various tasks, including biomass estimation. At present, LiDAR-based biomass estimation methods are mainly divided into two categories: canopy height-based[10–12] and point cloud density-based[13–15]. In view of the limited variation of crop height in wheat breeding experiments, it is more appropriate to use the voxel-based method based on point cloud density.

However, when dealing with voxel point cloud of small grain cereals such as wheat, the following limitations currently exist:

- The accuracy of non-deep learning methods is low.
- To process voxel-grouped point cloud, deep learning methods use 3D convolution which makes network structure complex and costs high computation[16].

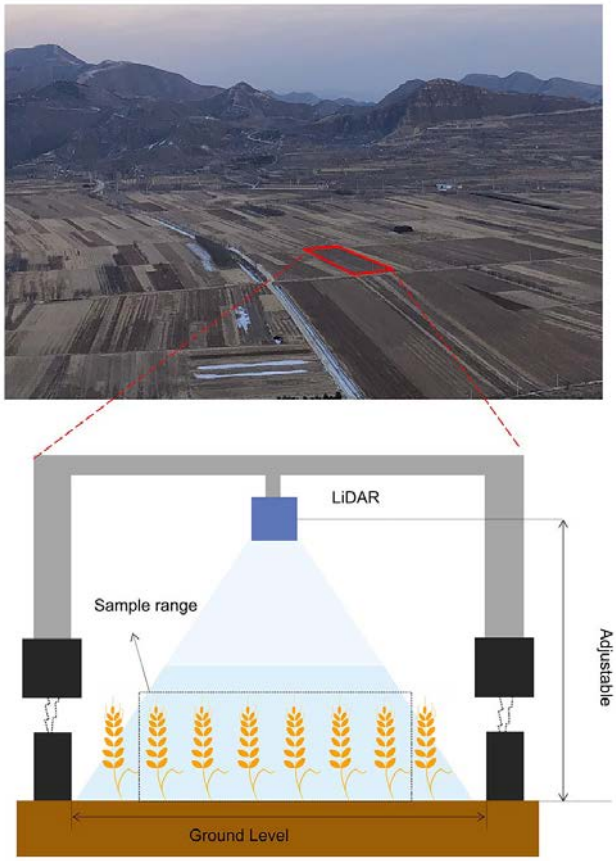

Fig. 1: Schematic of how to get the point cloud of wheat

Recently, the emergence of the MLP-Mixer[17] network draws attention to the fact that after dividing a 2D image into patches, a simple MLP structure can achieve high accuracy. However, the form of patches is not the case for nature point cloud. In the 3D object detection tasks, serval studies group point cloud as voxels[18–20], which makes the point cloud possible to be input of mixer structure.

In this paper, we propose an end-to-end network based on voxel-grouped point cloud and mixer structure, named BioPM, to predict above-ground biomass. BioPM consists of two modules: 1) Feature Encoding Module (Sec. 3.2); 2)

Mixer Module (Sec. 3.3). We use a FCN at last to output the prediction of biomass.

The key contributions of our work are as follows:

- We propose a simple and effective MLP network, named Biomass prediction mixer (BioPM), to accurately predict above-ground biomass from point cloud genarated by LiDAR.
- We provide the idea of Mixer architecture application in 3D point cloud
- Experimental results outperform the baselines on the public dataset, and surpass the SOTA method without the point cloud completion module, proving the role of the completion.

2 Related Work

2.1 Biomass Prediction with LiDAR

There are lots of studies based on adopting canopy height as a surogate for biomass[21–23]. Most of them use one or two LiDARs mounted on the moving platform to get the data [24] and use various models to predict biomass, such as bivariate regression[25], Pearson's correlation analysis and structural equation modelling (SEM)[26]. However, considering the varitation for height is limited in breeding programs, canopy height is not suitable as surogate to select improved biomass[6].

Utilizing the 3D nature point cloud, point density-based methods have a strong correlation with measured biomass, which have been widely adopted in cotton[27],wheat[14] and so on. Jimenez-Berni et al.[13] proposed a voxel-based method (3DVI) by dividing the point cloud into voxels and extracting voxel-wise and point-wise features, which is currently considered the baseline for measurin AGB and used in for its accuracy and robustness for real-world application[28, 29]. Recently, Pan et al.[6] proposed a deep leanrning based method (BioNet) and shared a new small grain cereals biomass prediction dateset. By fusing completion, regularization and projection modules, BioNet achieves great improvement on the dataset they shared compared to other methods.

2.2 Voxel-based Network

Since PointNet[30] successfully uses mutil-layer perceprton (MLP) networks to extract features from nature point cloud, lots of methods have been proposed to apply in downstream tasks and improve its performance. In 3D object detection field, voxel-grouped methods are popular. Zhou et al.[18] proposed VoxelNet by diving the point cloudinto equally spaced 3D voxels and encoding each voxel via simplified PointNet-like feature endoing layer. SECOND[20] improved the inference speed of VoxelNet but is still limited by 3D convolutions. Different from VoxelNet dividing point cloud from three-dimension, PointPillars[19] learns features on pillars by dividing point cloud in the form of verical columns, which is suitable to be as input of 2D backbone network.

2.3 Mixer architecture

MLP-Mixer[17] is a pioneering work, which proposed a MLP-like network consists of token-mixing MLPs and channel-mixing MLPs for 2D image tasks without convolutions or self-attention. Based on MLP-Mixer, many MLP-like architectures[31–33] are proposed to improve the performance. CycleMLP [34] is the first to introduce a hierarchical MLP-like architecture for dense prediction tasks.

Inspired by the success in 2D image tasks of MLP-like architecture, some researchers try to apply it into point cloud. PointMixer[35] introdcues mixer to point cloud by simply replacing token-mixing MLPs with a softmax function and achieves competitive improvement.

3 BioPM

3.1 Overview

The pipeline of BioPM is shown in Fig. 2. There are two key modules in BioPM: Feature Encoding Module and Mixer Module. Feature Encoding Module is to group point cloud as pillars and extract pillar-wise features, which acts like tokenlization of 2D image. Mixer Module is to go deep in extracting pillar-wise features and output the prediction $\widetilde{b}$ of the AGB. Here, we use b present the ground truth biomass.

The loss function of our model is given by

$$L(\widetilde{b}, b) = \frac{1}{M}\sum_{m=1}^{M} S(b_m - \widetilde{b}_m) \tag{1}$$

where M is the number of plots and the subscript $m \in \{1, ..., M\}$ is the plot index. $S(.)$ is the smooth l_1 loss function which is robust to outiers in regression tasks[36].

3.2 Feature Learning Module

To create a set of pillars without binning in the z direction, we have to discretize the point cloud into an evenly spaced grid in the x-y plane by resolution v_h and v_w.

We define the point cloud encompasses x-y plane with range H, W along the y, x axes respectively. Assuming H, W are a multiple of v_h, v_w for simplicity, the resulting pillar grid is of size $H' = H/v_h$, $W' = W/v_w$. The amount of pillars is defined by $N = H' \times W'$.

The number of point cloud collected by LiDAR of wheat plot is about 50k. Considering sparsity of point cloud, we set limitations on the number of pillars per sample (N) and on the number of points per pillar (T). To keep a dense tensor of size $(N, T, 3)$, if a pillar is allocated too much points, the points are randomly sampled to T. Conversely, we pad zero points into the pillar until it's enough to populate the tensor.

After grouping, we use a stacked Pillar Feature Encoding (PFE) Layer to extract point-wise features. Let's denote $V = \{ p_i = [x_i, y_i, z_i]^T \}_{i=1...t}$ as a non-empty pillar containing $t \leq T$ points, where p_i contains xyz coordinates for the i-th point. Each p_i is transformed through a simplified version of PoinNet, which contains a linear layer followed by BatchNorm and ReLU, to generate point features f_i. Then, we use MaxPooling across all f_i to get the locally aggregated features f'. By conbing the f_i and f', we get the point-wise features f_i^{out} and output feature set $V_{out} = \{f_i^{out}\}_{i...t}$. We use PFE-i($c_{in}, c_{out}$) to represent the i-th PFE layer.By using FCN and element-wise Maxpool on the output of PFE-n, we obtain pillar-wise features with C dimenstions.

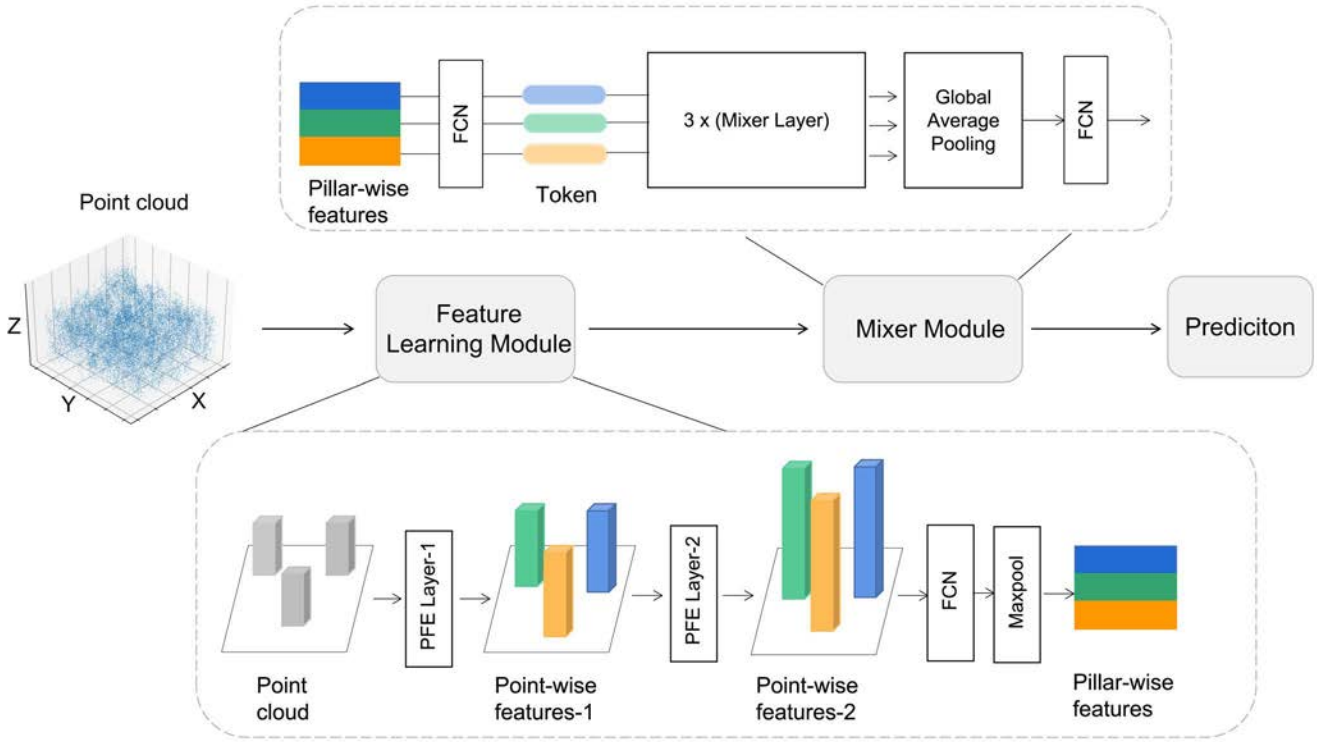

Fig. 2: Architecture of BioPM

3.3 Mixer Module

Mixer module accepts tensor (N, C) from Feature Learning Module as input. We use MLP-Mixer as backbone of mixer module. The backbone consists of token-mixing MLPs and channel-mixing MLPs. N is the number of token and C is the number of token features dimensions, which are required by backbone.

After a linear layer mapping, the tokens passed will be tranposed the spatial axis and channel axis to mix spatial information by token-mixing block which contains linear layer, LayerNorm[37] and GeLU[38]. Followed by channel-mixing block, input tokens are mixed in both spatial and channel dimensions. The tokens are passed to global average polling after repeating being mixed for several times. At last, we use a linear layer to process the output of backbone to get the prediction.

4 Experiments

In this section, we introduce the public dataset, which we evaluate on and augment with. Moreover, the detail of model settings and performance of our model is shown too.

4.1 Dataset

We evaluate our model on the small grain cereals biomass prediction (SGCBP) dataset[6], which consists of 306 point cloud and corresponding ground-truth AGB, split into 204 training samples and 102 testing samples.

The dataset was based on a field experiment,which sowed 26 varieties of small grain cerals at 78 plots for either seeding densities of 250 ('high') and 50 ('low') seeds/m^2, 156 plots in total. Point cloud samples of the dataset were scanned from plots at two different growing stage, vegetative stage and flowering stage by LiDAR.

4.2 Data Augmentation

As is shown Fig. 3, the width of point cloud at flowering stage is almost half of that at vegetative stage because of the same LiDAR height, which leads to using fixed pillar number is unworkable. To solve this problem and keep our model simple, we duplicate the point cloud of samples at flowering stage. Then we normalize all samples after employing statistical filtering to remove outliers from the point cloud.

Data augmentation is important for improving the performance especially on small dataset. There are less than 300 training point cloud, which results in that training our model from scratch will inevitably suffer from overfitting. Thus, we employ two forms of data augmentation. On the one hand, we shift the normalized point cloud by a uniformally distributed random variable $\Delta s \in$ [0.9,1.1]. On the other hand, we scale the point cloud with a random variable drawn from uniform distribution [0.8,1.25]

4.3 Settings

Evaluation metris. Here we use mean absolute error 'MAE' to evaluate the quality of the AGB prediction using BioPM. 'MAE' is defined as

$$MAE = \frac{1}{M}\sum_{m=1}^{M}|\widetilde{b}_m - b_m| \tag{2}$$

where $\widetilde{b}_m$ is the m-th predicted biomass and b_m is the m-th manullay measured ground-truth biomass. Same as formula (1), M is the number of plots and the subscript $m \in \{1, ..., M\}$ is the plot index.

Implementation details. We choose a pillar size of v_w = 0.125, v_h = 0.125 for normalized samples, which leads to W' = 8, H' = 8 and N(the number of pillars) = 64. We set T

= 32 as the maximum number of pillars by default. We employ two PFE layers, PFE-1(3,32) and PFE-2(32,128), and a FCN layer to genarate a sparse tensor of shape $128 \times 8 \times 8$. Then we use three MLP-Mixer blocks, of which input dimenstion is 128, to aggregate pillar-wise features. At last, we employ a FCN(128,1) layer to get the prediction.

BioPM is trained from scratch employing the Adam optimizer[39] and implemented in PyTorch. We set 0.001 as initial learning rate and halve the learning rate per 40 epochs.

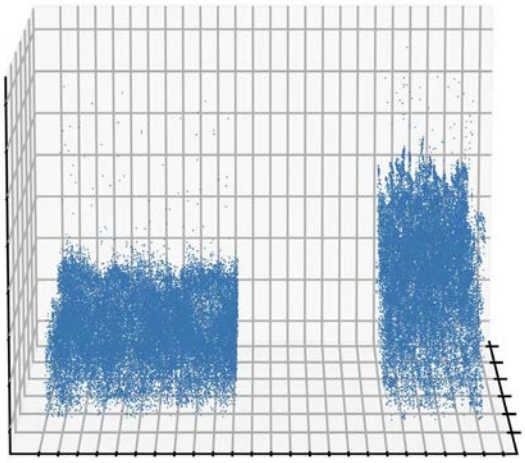

Fig. 3: Point cloud of wheat at vegetative (left) stage and flowering stage (right) from SGCBP dataset

4.4 Results

Baselines. Considering there is no open-source biomass prediction baseline available for now, we follow the baselines build from [6], which includes PointNet[30], PointNet++[40], DGCNN[41] and GS-Net[42]. Besides, non-deep SOTA method 3DVI is compared with too.

We evaluate our model on the public dataset SGCBP. As is shown in Tables 1, BioPM performs competitively compared with the state-of-the-art methods at biomass prediction. Especially, our method improve 33% compared with SOTA non-deep method (3DVI).

Table 1: Results of biomass prediction methods on SGCBP

Method	MAE↓
PointNet[30]	139.61
PointNet++[40]	142.80
DGCNN[41]	129.66
GS-Net[42]	145.63
3DVI[13]	115.15
BioNet[6]	71.23
BioNet$_{pr}$[6]	**98.27**
BioPM	**79.80**

Considering architecture we proposed is possible to expand a point cloud completion module, we specially compare our method with 'BioNet$_{pr}$' which refers to the BioNet without completion module. As bold MAE shown in Tables 1, the relative improvement is 19%. Even compared with the MAE of BioNet, ours is very close to, which proves the role of point cloud completion module. What's more, our model runs very fast due to simple network architecture.

5 Conclusion

In this paper, we propose BioPM, a simple but effective deep learning network to estimate above-groud biomass(AGB) based on point cloud. By utilizing voxel-grouped point cloud, we explore introducing up-to-date network architecture (Mixer) in our method. Our model archives competitive AGB prediction on the public dataset SGCBP compared with SOTA methods. The simple architecture also makes our model expandable. In the future, we aim to expand the model with more modules to improve prediction accuracy.

References

[1] I. FAO, *The State of Food Security and Nutrition in the World 2021: Transforming food systems for food security, improved nutrition and affordable healthy diets for all*, ser. The State of Food Security and Nutrition in the World (SOFI). Rome, Italy: FAO, 2021, no. 2021.

[2] K. Johansen, M. J. L. Morton, Y. Malbeteau, B. Aragon, S. Al-Mashharawi, M. G. Ziliani, Y. Angel, G. Fiene, S. Negrão, M. A. A. Mousa, M. A. Tester, and M. F. McCabe, "Predicting Biomass and Yield in a Tomato Phenotyping Experiment Using UAV Imagery and Random Forest," *Frontiers in Artificial Intelligence*, vol. 3, 2020.

[3] J. Dauber, M. B. Jones, and J. C. Stout, "The impact of biomass crop cultivation on temperate biodiversity," *GCB Bioenergy*, vol. 2, no. 6, pp. 289–309, 2010.

[4] M. A. Parry, M. Reynolds, M. E. Salvucci, C. Raines, P. J. Andralojc, X.-G. Zhu, G. D. Price, A. G. Condon, and R. T. Furbank, "Raising yield potential of wheat. II. Increasing photosynthetic capacity and efficiency," *Journal of experimental botany*, vol. 62, no. 2, pp. 453–467, 2011.

[5] A. Pask, J. Pietragalla, D. Mullan, and M. P. Reynolds, "Physiological breeding II: A field guide to wheat phenotyping," 2012.

[6] L. Pan, L. Liu, A. G. Condon, G. M. Estavillo, R. A. Coe, G. Bull, E. A. Stone, L. Petersson, and V. Rolland, "Biomass Prediction With 3D Point Clouds From LiDAR," p. 11.

[7] T. Canata, M. Martello, L. Maldaner, J. Moreira, and J. Molin, "3D Data Processing to Characterize the Spatial Variability of Sugarcane Fields," *Sugar Tech*, Sep. 2021.

[8] M. Alonzo, R. J. Dial, B. K. Schulz, H.-E. Andersen, E. Lewis-Clark, B. D. Cook, and D. C. Morton, "Mapping tall shrub biomass in Alaska at landscape scale using structure-from-motion photogrammetry and lidar," *Remote Sensing of Environment*, vol. 245, p. 111841, 2020.

[9] Y. Zhao, X. Liu, Y. Wang, Z. Zheng, S. Zheng, D. Zhao, and Y. Bai, "UAV-based individual shrub aboveground biomass estimation calibrated against terrestrial LiDAR in a shrub-encroached grassland," *International Journal of Applied Earth Observation and Geoinformation*, vol. 101, p. 102358, 2021.

[10] C. Wang, S. Nie, X. Xi, S. Luo, and X. Sun, "Estimating the biomass of maize with hyperspectral and LiDAR data," *Remote Sensing*, vol. 9, no. 1, p. 11, 2017.

[11] J. U. Eitel, T. S. Magney, L. A. Vierling, H. E. Greaves, and G. Zheng, "An automated method to quantify crop height and calibrate satellite-derived biomass using hypertemporal lidar," *Remote Sensing of Environment*, vol. 187, pp. 414–422, 2016.

[12] N. Tilly, D. Hoffmeister, Q. Cao, S. Huang, V. Lenz-Wiedemann, Y. Miao, and G. Bareth, "Multitemporal crop surface models: Accurate plant height measurement and biomass estimation with terrestrial laser scanning in paddy rice," *Journal of Applied Remote Sensing*, vol. 8, no. 1, p. 083671, 2014.

[13] J. A. J. Berni, D. Deery, P. Rozas-Larraondo, A. Condon, G. Rebetzke, R. James, W. Bovill, R. Furbank, and X. Sirault, "High Throughput Determination of Plant Height, Ground Cover, and Above-Ground Biomass in Wheat with LiDAR," *Frontiers in Plant Science*, vol. 9, Feb. 2018.

[14] J. D. Walter, J. Edwards, G. McDonald, and H. Kuchel, "Estimating biomass and canopy height with LiDAR for field crop breeding," *Frontiers in plant science*, vol. 10, p. 1145, 2019.

[15] J. U. H. Eitel, T. S. Magney, L. A. Vierling, T. T. Brown, and D. R. Huggins, "LiDAR based biomass and crop nitrogen estimates for rapid, non-destructive assessment of wheat nitrogen status," *Field Crops Research*, vol. 159, pp. 21–32, Mar. 2014.

[16] Z. Liu, H. Tang, Y. Lin, and S. Han, "Point-Voxel CNN for Efficient 3D Deep Learning," in *Advances in Neural Information Processing Systems*, vol. 32. Curran Associates, Inc., 2019.

[17] I. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, A. Steiner, D. Keysers, J. Uszkoreit, M. Lucic, and A. Dosovitskiy, "MLP-Mixer: An all-MLP Architecture for Vision," *arXiv:2105.01601 [cs]*, May 2021.

[18] Y. Zhou and O. Tuzel, "VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4490–4499.

[19] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "PointPillars: Fast Encoders for Object Detection From Point Clouds," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA: IEEE, Jun. 2019, pp. 12 689–12 697.

[20] Y. Yan, Y. Mao, and B. Li, "Second: Sparsely embedded convolutional detection," *Sensors*, vol. 18, no. 10, p. 3337, 2018.

[21] E. L. Loudermilk, J. K. Hiers, J. J. O'Brien, R. J. Mitchell, A. Singhania, J. C. Fernandez, W. P. Cropper, and K. C. Slatton, "Ground-based LIDAR: A novel approach to quantify fine-scale fuelbed characteristics," *International Journal of Wildland Fire*, vol. 18, no. 6, pp. 676–685, 2009.

[22] R. Gebbers, D. Ehlert, and R. Adamek, "Rapid mapping of the leaf area index in agricultural crops," *Agronomy Journal*, vol. 103, no. 5, pp. 1532–1541, 2011.

[23] N. Tilly, D. Hoffmeister, Q. Cao, S. Huang, V. Lenz-Wiedemann, Y. Miao, and G. Bareth, "Multitemporal crop surface models: Accurate plant height measurement and biomass estimation with terrestrial laser scanning in paddy rice," *Journal of Applied Remote Sensing*, vol. 8, no. 1, p. 083671, 2014.

[24] W. Saeys, B. Lenaerts, G. Craessaerts, and J. De Baerdemaeker, "Estimation of the crop density of small grains using LiDAR sensors," *Biosystems Engineering*, vol. 102, no. 1, pp. 22–30, Jan. 2009.

[25] N. Tilly, H. Aasen, and G. Bareth, "Fusion of plant height and vegetation indices for the estimation of barley biomass," *Remote Sensing*, vol. 7, no. 9, pp. 11 449–11 480, 2015.

[26] W. Li, Z. Niu, N. Huang, C. Wang, S. Gao, and C. Wu, "Airborne LiDAR technique for estimating biomass components of maize: A case study in Zhangye City, Northwest China," *Ecological indicators*, vol. 57, pp. 486–496, 2015.

[27] S. Sun, C. Li, A. H. Paterson, Y. Jiang, R. Xu, J. S. Robertson, J. L. Snider, and P. W. Chee, "In-field high throughput phenotyping and cotton plant growth analysis using LiDAR," *Frontiers in Plant Science*, vol. 9, p. 16, 2018.

[28] J. ten Harkel, H. Bartholomeus, and L. Kooistra, "Biomass and crop height estimation of different crops using UAV-based LiDAR," *Remote Sensing*, vol. 12, no. 1, p. 17, 2020.

[29] J. D. Walter, J. Edwards, G. McDonald, and H. Kuchel, "Estimating biomass and canopy height with LiDAR for field crop breeding," *Frontiers in plant science*, vol. 10, p. 1145, 2019.

[30] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," *arXiv:1612.00593 [cs]*, Apr. 2017.

[31] J. Guo, Y. Tang, K. Han, X. Chen, H. Wu, C. Xu, C. Xu, and Y. Wang, "Hire-mlp: Vision mlp via hierarchical rearrangement," *arXiv preprint arXiv:2108.13341*, 2021.

[32] Q. Hou, Z. Jiang, L. Yuan, M.-M. Cheng, S. Yan, and J. Feng, "Vision permutator: A permutable mlp-like architecture for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[33] H. Touvron, P. Bojanowski, M. Caron, M. Cord, A. El-Nouby, E. Grave, G. Izacard, A. Joulin, G. Synnaeve, and J. Verbeek, "Resmlp: Feedforward networks for image classification with data-efficient training," *arXiv preprint arXiv:2105.03404*, 2021.

[34] S. Chen, E. Xie, C. Ge, D. Liang, and P. Luo, "Cyclemlp: A mlp-like architecture for dense prediction," *arXiv preprint arXiv:2107.10224*, 2021.

[35] J. Choe, C. Park, F. Rameau, J. Park, and I. S. Kweon, "PointMixer: MLP-Mixer for Point Cloud Understanding," *arXiv:2111.11187 [cs]*, Nov. 2021.

[36] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1440–1448.

[37] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," *arXiv preprint arXiv:1607.06450*, 2016.

[38] D. Hendrycks and K. Gimpel, "Gaussian Error Linear Units (GELUs)," *arXiv:1606.08415 [cs]*, Jul. 2020.

[39] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[40] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," *arXiv preprint arXiv:1706.02413*, 2017.

[41] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph cnn for learning on point clouds," *Acm Transactions On Graphics (tog)*, vol. 38, no. 5, pp. 1–12, 2019.

[42] M. Xu, Z. Zhou, and Y. Qiao, "Geometry sharing network for 3D point cloud classification and segmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, 2020, pp. 12 500–12 507.