TELKOMNIKA Telecommunication, Computing, Electronics and Control
Vol. 19, No. 2, April 2021, pp. 479~490
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v19i2.18321

PhosopNet: An improved grain localization and classification by image augmentation

Pakpoom Mookdarsanit[1], Lawankorn Mookdarsanit[2]
[1]Faculty of Science, Chandrakasem Rajabhat University, Bangkok, Thailand
[2]Faculty of Management Science, Chandrakasem Rajabhat University, Bangkok, Thailand

Article Info

Article history:

Received Aug 3, 2020
Revised Sep 30, 2020
Accepted Oct 19, 2020

Keywords:

Feature transformation
Grain classification
Grain localization
Image augmentation
Transfer adaptation learning

ABSTRACT

Rice is a staple food for around 3.5 billion people in eastern, southern and south-east Asia. Prior to being rice, the rice-grain (grain) is previously husked and/or milled by the milling machine. Relevantly, the grain quality depends on its pureness of particular grain specie (without the mixing between different grain species). For the demand of grain purity inspection by an image, many researchers have proposed the grain classification (sometimes with localization) methods based on convolutional neural network (CNN). However, those papers are necessary to have a large number of labeling that was too expensive to be manually collected. In this paper, the image augmentation (rotation, brightness adjustment and horizontal flipping) is appiled to generate more number of grain images from the less data. From the results, image augmentation improves the performance in CNN and bag-of-words model. For the future moving forward, the grain recognition can be easily done by less number of images.

Corresponding Author:

Pakpoom Mookdarsanit
Computer Science, Faculty of Science
Chandrakasem Rajabhat University
39/1 Rachadapisek Road, Chan Kasem District, Chatuchak, Bangkok, 10900, Thailand
Email: pakpoom.m@chandra.ac.th

1. INTRODUCTION

A bowl/dish of cooked rice is easily seen as the cultural gastronomy in many Asian countries, e.g., Japan, China, India, Bangladesh, Pakistan and other ASEAN countries. From the oldest historical evidence, rice-grain (or grain) was grown in [1] Yangtze river, China; longer than 10,000 years ago. A folk wisdom on grain agriculture was originally farmed on volcanic soil in Kyushu, Japan [1] during Yayoi period. And the flow of Mekong river [1] (shared by Vietnam, Laos, Thailand, Myanmar and Cambodia) was also one of the most important grain-cultivated-lands in a long time ago. It is not surprise that most winners of world's best rice conferences within the last 4 years were from the Mekong river shared region: Jasmine [2] (Thailand, 2016-2017), Malys Angkor [2] (Cambodia, 2018) and ST25 [2] (Vietnam, 2019). Traditionally, rice was linked to the goddess belief in Japan [1] who sowed grain in the fields of heaven. In Indian culture as Pongal [1], rice was an offering to the god as a thanksgiving. As well as Thai, Cambodian and Balinese had the similar cultural worship of rice's mother [3, 4] as Mae Phosop, Po Ino Nogar and Dewi Sri, respectively. Economically, rice is not only a staple food but also an important agricultural production for 3.5 billion people in Asia (half of the world population). Prior to being rice, the grain is previously husked and/or milled by the milling machine. There are so many grain taxonomies; one of the world's widest diversity is absolutely Asian grain varieties (both paddy and glutinous grain). In the real market, the diversity of grain

species looks different physically genetic features (like size, texture and shape) that make them have the different prices. The most well-known trick in grain (and rice) trading is mixing the pure grain specie product with other species [5] for making the higher price by the heavier ton of product. The faulty impurity by mixing absolutely violates the product quality. As to TAS 4004-2017– one of Thai agriculture standard [6] that is defined for the grain purity inspection by randomnizing some samples the 5% of ton. Generally, the validation of many physical grains is still based on human vision as a manual labor.

Many researchers leveraged computer vision as an inspection solution since the beginning of "Japanese rice grading problem" in 2002 [7]. From the literature, all papers could be categorized by methods [8] into 2 groups: bag of words [9-23] and convolutional neural network (CNN) [24-29]. The former was used by early researches [30] (which require less number of labeled data [31]) that were still useful in some open-world industry [20-22] such as iRSVPred [23]. The latter was exponentially increased by current researches (which required large volume of labeled data [32, 33]) that had already been proven to be higher performance than bag-of-words methods [34]. The significant limitation in previous works in both groups is that they need a large number of labeled data that is really expensive for the high-quality manual labeling too many small grains by human labor.

To expand those previous works (in both bag of words [9-23] and CNN [24-29]), this paper proposes PhosopNet to do more with less labeled data by image augmentation, as shown in Figure 1. The augmentation is proposed to increase the size and variety of training rice-grain (or grain) data by grain rotation in different angels, brightness adjustment in power law distribution and horizontal flipping in x-axis, respectively. For testing, all grains are localized/detected by mask region convolutional neural network (Mask R-CNN). Each grain is classified by densely connected convolutional neural network (DenseNet). Note that the name "Phosop" is dedicated to the rice's mother [3, 4] in antique Thai culture who produced the rain over the land; in order to grow those grain seeds.

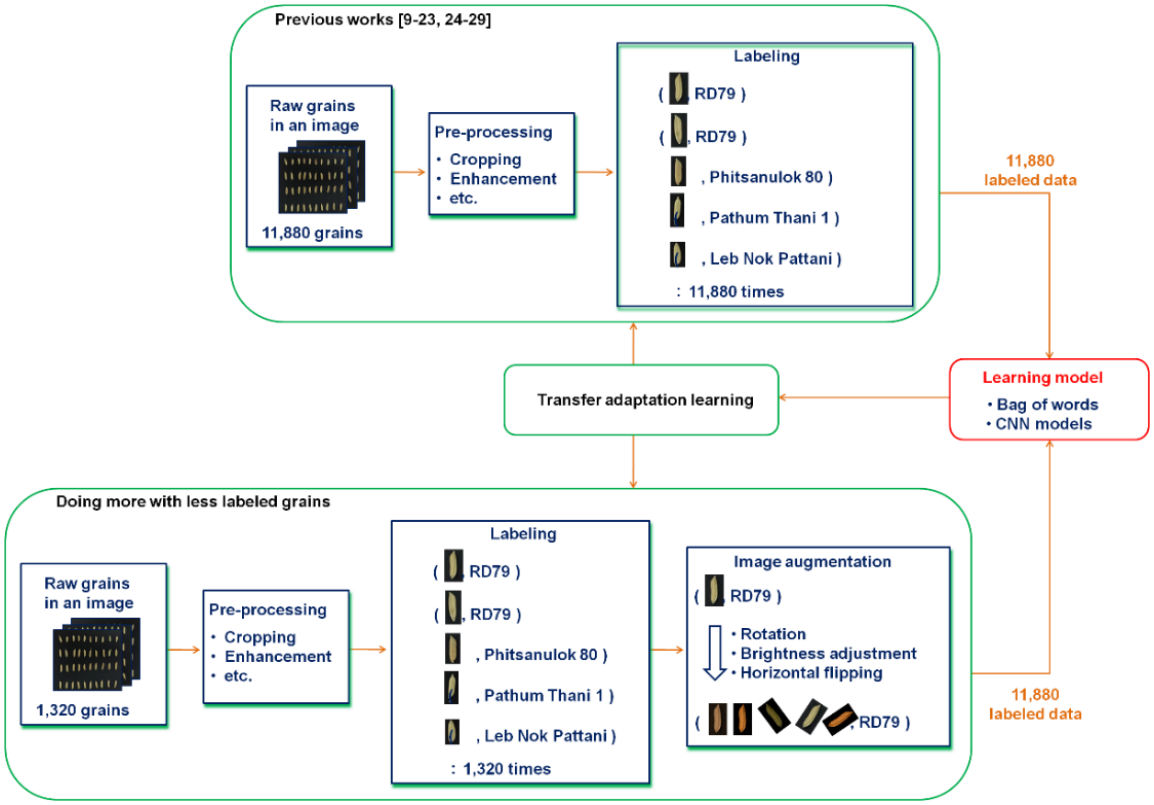


Figure 1. Overview of the improved grain localization and classification by image augmentation

The contribution of PhosopNet can be summarized as follows:

– The proposed augmentation operations can generate ten-thoundsand or thoundsand training grains from a little thoundsand or hundred raw labeled grain data, respectively.

- To do more with less data, the PhosopNet achieved the high localization and classification performance using only the less number of labeled grain data.
- For the grain recognition performance enhancement, not only convolutional neural network but also bag of words can be improved by image augmentation.

This paper is organized as follows. Related works are in section 2. Image augmentation and learning model are described in section 3 and 4. Section 5 talks about experimental settings and results. And the conclusion is in section 6.

2. RELATED WORKS

The history began from "Japanese rice grading problem" [7] that the authors firstly introduced the way to use computer vision as the main solution in 2002. Later, there were many papers concerning rice recognition. Those papers can be categorized by methods [8] into 2 groups: bag of words and convolutional neural network.

2.1. Bag of words

For traditional bag of words, Japanese grain grading was originally introduced by handcrafted feature with neural network [7] as a supervised model. Neural network [19] was the main classfier for shape feature [17] and the principal component analysis (PCA) was used for dimensional reduction on the features [9] in both morphology and multi-color channel. The result showed that neural network with PCA provided a better recognition rate. Not only was the grain recognition, neural network also found to be high correctness in germination prediction [18]. In contrast, the complexity of neural network was found to be a main problem in speed and resource consumption. Instead of the long time and resource processing in neural learning, many statistical with image processing techniques were also proposed [12-13, 15] as the alternative ways. Zernike polynomials were also orthogonally computed to quickly extract features [10] and the threshold-based segmentation [11] from the physical grain images. By the way, neural network was still the highest accuracy. Until 2004, a novel support vector machine (SVM) was proven to be higher performance (in speed and time) than MLP [35], especially in a larger number of target classes. Moreover, SVM also had [36] transfer adaptation learning (TAL) mechanism as well as convolutional neural network (CNN), called adaptive-SVM (Ada-SVM) [37]. SVM for grain recognition was used to learn features from colors, morphology and texture with sparse coding [16]. One highlight in bag of words was based on SVM [14]: the saturation channel from hue-saturation-value (HSV) model as a threshold for segmentation, the color histogram of the green (−) to red (+) and blue (−) to yellow (+) from international commission on illumination lab (CIELAB) model, the shape description by histogram of curvature and the texture was described by scale invariant feature (SIFT) [38], speed up robust features (SURF) [39] and root-SIFT [40], respectively. In 2012, a convolutional neural network (CNN) in AlexNet architecture [41] was the winner of ImageNet large scale visual recognition challenge (ILSVRC) that outperformed those bag-of-words models, especially in larger data volume [32, 33]. Many computer vision papers have been gradually shifted from traditional bag of words to CNN paradigm [8] to solve object localization and classification problems in big data. Argubly, the industrial requirements concern user experience, environmental implementation and software maintenance friendliness; it was sometimes better to be implemented by histogram of gradients (HoG) [42] with traditional machine learning as a bag of words model [20-22], for the open-world grain inspection [23].

2.2. Convolutional neural network

For convolutional neural network (CNN), grains were calibaratedly acquisted by hyperspectral camera and sent to CNN [24-25] that totally needed the cost for data acquistion. CNN was proven to be higher performance than traditional machine learning like k-NN and SVM [25] based on those hyperspectral images. For a digital image, GoogLeNet [43] (as Inception v.4 [44]) was used as the CNN architecture for germ integrity [26]. Later, the comparison between CNN architectures under the same environment [34] were done for grain image classification and densely connected convolutional networks (DenseNet) [45] showed the highest accuracy; higher than ResNet [46], GoogLeNet [43], Neural architecture search network (NasNet) [47] and visual geometry group (VGG) [48]. Moreover, the deeper model did not guarantee the more correctness of grain image classification (such VGG-16 higher correctness than VGG-19 [49]). Not only classification but also localization was necessary for grain quality inspection. As the highlight, Mask R-CNN [50] with ResNet [46] was used for grain localization and classification (called MIMR [29]). But a large number of manual labeling on too many small grains [51] was still necassary. To do more with less data, this paper named PhosopNet proposes the image augmentation that generates the thoundsand grain data from hundred one, instead of manually labeling those ten-thoundsand small grain images. For the expansion

of previous works, the computer vision applied to rice or grain problems (both bag of words [9-23] and CNN [24-29]) can achieve high performance by training the less labeled grain data.

3. IMAGE AUGMENTATION

Since the previous papers on grain (or rice-grain) recognition (in both bag of words [9-23] and convolutional neural network [24-29]) require a large number of human-annotated labels (labeled data), the proposed PhosopNet leverages the image augmentation by feature transformation to artificially generate a variety of grain instances as a larger dataset and train them to the model. Practically, a grain object that is performed by image processing operations: rotation, brightness adjustment and horizontal flipping, in order to increase a number of data, as shown in Figure 2.

Figure 2. To increase training labeled data, the augmented examples by a RD79 grain image; (a) an original RD79 grain, (b) rotation, (c) brightness adjustment, and (d) horizontal flipping

3.1. Rotation

For rotation of a grain object, the pivot point is at the top-left pixel of an image as $(x, y) = (0,0)$. The rotation $Rotate_{Grain}(x, y, \theta)$ steps are measured in degree that moves in counter clockwise direction (θ) from 0 to 360. Each object pixel is moved to the new position (u, v) by (1).

$$Rotate_{(Grain)}(x, y, \theta) = (u, v); \quad where \begin{cases} u = (x \cdot cos\,\theta) - (y \cdot sin\,\theta) \\ v = (x \cdot sin\,\theta) + (y \cdot cos\,\theta) \end{cases} \tag{1}$$

3.2. Brightness adjustment

For brightness adjustment ($Adjust_{Brightness\ \ (Grain)}(x, y)$), power law distribution is used to tune a pixel ($Pixel(x, y)$) into image brightness values by gamma threshold (γ) and a constant (c). The lower γ value provides more darkness and the higher one provides more lightness, vice and versa. And the c is normally set to 1.

$$Adjust_{Brightness\ \ (Grain)}(x, y) = c \cdot (Pixel(x, y))^{\gamma} \tag{2}$$

3.3. Horizontal flipping

For horizontal flipping ($Flip_{Horizon\ \ (Grain)}([x_0,\ x_1,\ x_2, \ldots, x_{n-1},\ x_n])$), the grain object is flipped horizontally by (3). The flipped grain seems to be a new grain object. Practically, the horizontal flipping is done by the descending order of pixels in x-positions of entire image.

$$Flip_{Horizon\ \ (Grain)}([x_0,\ x_1,\ x_2, \ldots, x_{n-1},\ x_n]) = [x_n,\ x_{n-1}, \ldots,\ x_2,\ x_1,\ x_0] \tag{3}$$

4. LEARNING MODEL

Convolutional neural network (CNN) achieves performance over conventional bag of words [32, 33], especially in large volume of data. For bag of words, the positions of all grains are localized and transformed into the numerical values by handcrafted feature extraction [52] (e.g., SIFT [38], SURF [39] or HoG [42]) those values are used to classify using traditional supervised machine learning (e.g., MLP [7, 35] or SVM [14, 35]). For the CNN, all grains within an image are localized by CNN detection (e.g., Faster R-CNN or Mask R-CNN); each grain object is directly represented in term of features and classified by CNN classification (e.g., ResNet [46] or DenseNet [45]). Moreover, CNN conveys the role

of transfer adaptation learning [36] with pre-trained weights of COCO dataset that model representation can be retrained in many times.

4.1. Localization

To localize the grains (or rice-grains) within an image, mask region convolutional network (Mask R-CNN [50]) based on DenseNet [45] was used to detect all grains with their positions in the proposed PhosopNet. Mask R-CNN is one of region-proposal based (two-stage) [53] detection pipeline that was designed to preserve the lowest instance (or pixel) level spatial correspondence. Although two-stage pipeline was shown to be higher average precision (AP) than one-stage pipeline [33], (e.g., you only look once (YOLO), and single shot multibox detector (SSD), one-stage detection was better in speed; and mostly used in real-time applications. For the grain recognition, texture within a grain object was small and very similar between species; the localization accuracy was necessary to use two-stage detection. Originally, the two-stage detection pipeline inherited from R-CNN [54] that firstly introduced to use the regions as CNN features. However, R-CNN had the expensive and slow problem on training support vector machine (SVM) for localization of all grains. For the improvement, Fast R-CNN [55] used region of interest (RoI) pooling, instead of unorganized RoI; and also used soft-max loss, instead of the full SVM classifier. Later, region proposal network (RPN) and multi-reference detection were the main contribution in Faster R-CNN [56] that completely solved the redundancy and bottleneck of Fast R-CNN. Since the flat Faster R-CNN cannot tackle pixel-wise instance in grain localization, Mask R-CNN [50] was extended from both Faster R-CNN and Fast R-CNN that achieved results by including feature pyramid network (FPN) [57] for feature fusion, RoI alignment and bi-linear upsampling. To identify the boundary of grain, Mask R-CNN uses RPN to generate the bounding box of each object as the first stage and the class parallel prediction in the second stage, respectively.

4.2. Classification

For the grain (or rice-grain) classification, densely connected convolutional network (DenseNet) [45] is a main architecture in the proposed PhosopNet which enables transfer domain learning. Originally, visual geometry group network (VGGNet) [48] used only 3x3 convolutional kernels. Unlike AlexNet [41], the larger kernel size (such as 5x5 or 7x7) caused the larger model and too many parameters. Moreover, too larger stride made the network lost the useful features from the lower layers. Although VGGNet was proven that the deeper networks obtained better performance, it was later found to spawn the problem as gradient vanishing and explosion that were finally solved by skip connection in residual network (ResNet) [46]. Unfortunately, most architectures are neither hierarchical (e.g., AlexNet [41], VGGNet [48], ResNet [46]) nor parallel (e.g., GoogLeNet [43]) architectures that make the low-level grain features to be disable for reusing in the high-level layers. For the solution by DenseNet, the feature maps from previous layers were also sent to the next convolutional blocks. Moreover, the transition layers after dense layer were proposed to reduce the number of feature maps in grain features that completely made the shallow layers focus on low-level features and the deeper layers focus on high-level features. The DenseNet architectures were shown in Table 1.

Table 1. DenseNet configuration

Layer	Detail	Output size
Convolution	7×7 *CONV*, *stride* 2	112×112
Pooling	3×3 *Max Pool*, *stride* 3	56×56
DenseBlock (1)	$\begin{bmatrix} 1\times1\ CONV \\ 3\times3\ CONV \end{bmatrix}\times6$	56×56
Transition layer (1)	1×1 *CONV*	56×56
	2×2 *Average Pool*, *stride* 2	28×28
DenseBlock (2)	$\begin{bmatrix} 1\times1\ CONV \\ 3\times3\ CONV \end{bmatrix}\times12$	28×28
Transition layer (2)	1×1 *CONV*	28×28
	2×2 *Average Pool*, *stride* 2	14×14
DenseBlock (3)	$\begin{bmatrix} 1\times1\ CONV \\ 3\times3\ CONV \end{bmatrix}\times24$	14×14
Transition layer (3)	1×1 *CONV*	14×14
	2×2 *Average Pool*, *stride* 2	7×7
DenseBlock (4)	$\begin{bmatrix} 1\times1\ CONV \\ 3\times3\ CONV \end{bmatrix}\times6$	7×7

4.3. Transfer adaptation learning

Transfer adaptation learning [58] enables to transfer knowledge from one training task into another one. For the first training, the pre-trained weights are set as the initial network. The source domain contains some useful grain features that are used for retraining the second time. Technically, all weights from the source domain can be reused and retrained with the new labeled grain data (and sometimes with their target classes). The usefulness of transfer adaptation learning in grain recognition is that the retraining task can be performed in many times. This makes a less number of small labeled grain data to be iteratively trained to the model, instead of one time (or big-bang) training from the large-scale data. Furthermore, transfer adaptation learning [36] can be divided into transfer learning (TL) and domain adaptation (DA), as shown in Figure 3.

For transfer learning $(Train_{TL}(\bullet))$, the pair of grain feature and class in source $(I_{source\ \ (Grain)}, C_{source\ \ (Grain)})$ and target domain $(I_{t\,arg\,et\ \ (Grain)}, C_{t\,arg\,et\ \ (Grain)})$ are different, by (4).

$$Train_{TL}(I_{source\ \ (Grain)}, C_{source\ \ (Grain)}) \neq Train_{TL}(I_{t\,arg\,et\ \ (Grain)}, C_{t\,arg\,et\ \ (Grain)}); \quad (4)$$

where $I_{source\ \ (Grain)} \neq I_{t\,arg\,et\ \ (Grain)}$, $C_{source\ \ (Grain)} \neq C_{t\,arg\,et\ \ (Grain)}$. For domain adaptation $(Train_{DA}(\bullet))$, the grain feature from the source domain $\left(I_{source\ \ (Grain)}\right)$ is different from the target domain $\left(I_{t\,arg\,et\ \ (Grain)}\right)$ but they are members in the same class, by (5).

$$Train_{DA}(I_{source\ \ (Grain)}, C_{source\ \ (Grain)}) = Train_{DA}(I_{t\,arg\,et\ \ (Grain)}, C_{t\,arg\,et\ \ (Grain)}); \quad (5)$$

where $I_{source\ \ (Grain)} \neq I_{t\,arg\,et\ \ (Grain)}$, $C_{source\ \ (Grain)} = C_{t\,arg\,et\ \ (Grain)}$

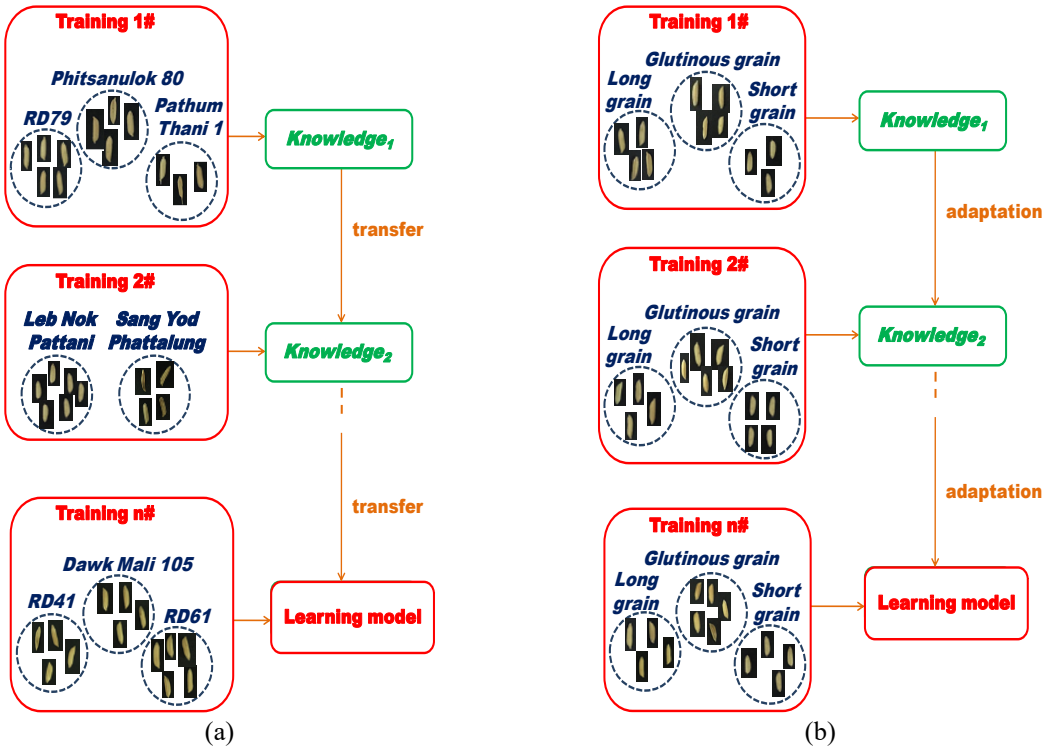


Figure 3. The difference between transfer learning and domain adaptation; (a) transfer learning (more samples with the new target classes) and (b) domain adaptation (more samples with the same target classes)

5. EXPERIMENTAL SETTINGS AND RESULTS

According to the real-world inspection problems in the industry [5, 6], this section talked about the experimental results and discussion in PhosopNet. Image augmentation with transfer adaptation learning was used to increase the data volume and variety. The detail could be categorized into 6 the main issues.

5.1. Datasets

The raw grains with their target classes in this experiment could be divided into 3 different paradigm settings according to the rice-grain standard inspection [5, 6], named "Phosop i-th" (in Table 2). These raw samples were classified and sent from a grain inspection laboratory. Those physical grains were trained to the supervised model in a format of digital image as the primary dataset. The grains were put on the black scene. Within an image, each row contained 10 grains which were the same target class. The distance between image and camera positions was 25 cm.

5.2. Experimental settings

For the experimental settings, PhosopNet was such a supervised learning (or supervision). Mask R-CNN [50] and DenseNet [45] were applied for localization and classification, respectively. The supervised model generally consisted of training and testing.

5.2.1. Training

All grains in each row were laid on the same orientation. The image and camera positions were vertical; and the distance between them was 25 cm. Each row refered to one target class that had 10 grain samples, as shown in Figure 4. For the labeling, all cropped grains in each row were labeled one by one in text file and trained by CNN-based supervised model, where the grain same row was the same target class. To do more with less data, each grain was further augmented to increase the dataset size by rotation, brightness adjustment and horizontal flipping.

Table 2. Grain datasets with the experimental settings

Problem	Target classes	# training grains by manual labeled grain	Augmentation operations	Transfer adaptation learning	# training grains, the labeled grains increased by augmentation	# testing grains
Phosop-1	2 classes: glutinous grain and paddy grain	300	(1) Rotation, (2) Brightness Adjustment and (3) Horizontal Flipping	Domain Adaptation	2,400	500
Phosop-2	3 classes: glutinous grain, long-paddy grain and short-paddy grain	450			3,600	700
Phosop-3	11 classes: Dawk Mali 105, Pathum Thani 1, Chiang Phatthalung, Leb Nok Pattani, Sang Yod Phattalung, Phitsanulok 80, RD41, RD61, RD 79, RD 6 and San Pah Tawng 1	1,320		Domain Adaptation and Transfer Learning	11,880	1,500

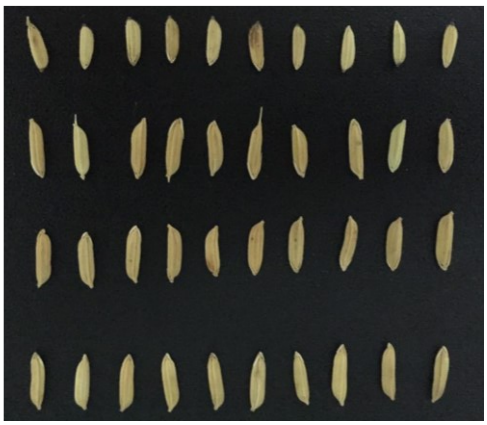


Figure 4. Such an example of raw grain image in 4 rows (from the first to the last row*:* Leb Nok Pattani, Pathum Thani 1, Phitsanulok 80 and RD79), all grains in each row has the same target class

5.2.2. Testing

All grains could be laid on any orientations but they should not have been overlapped one another. The image and camera positions were either vertical or non-vertical in any background colors. The distance between image and camera could be varying according to the real-world inspection. All grains in any orientations were detected and generated in the same orientations with new positions, as shown in Figure 5.

For testing, Mask R-CNN [50] firstly localized all grains within the image. Each grain was classified by DenseNet [45], as shown in Figure 5. With the help of image augmentation, the number of training by manual labeled grain could be less than that of testing. For the localization evaluation, the intersection over union (IoU) between proposal locations and the associated ground-truth labeling was set to 50%; the performance of grain objectiveness localization was evaluated by mean average precision (mAP) metric [53]. In the same way, the accuracy metric was used to measure the classification correctness [33].

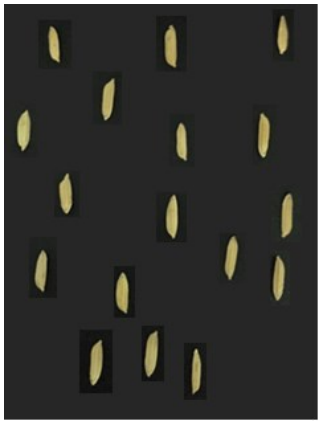
5.3. Phosop-1: Purity between glutinous and paddy grain

Purity between glutinous and paddy grain was one of the main industrial problem in grain inspection. As to the physical appearance, the glutinous grains were both fatter and longer than the paddy grains. As related to the Phosop-1 problem, the augmentation could improve both localization for 32% and classification for 31%, as shown in Table 3. The augmentation operations (rotation, brightness adjustment and horizontal flipping) increased the data size from 300 grains to 2,400 grains that totally boosted the Mask R-CNN [50] to localize the grain objects better from an image by larger size and variety of image training data. To do more with less data by augmentation, the purity between glutinous and paddy rain in 500 testing grains could be correctly classified as 100%, using only 300 manually-labelled grains.

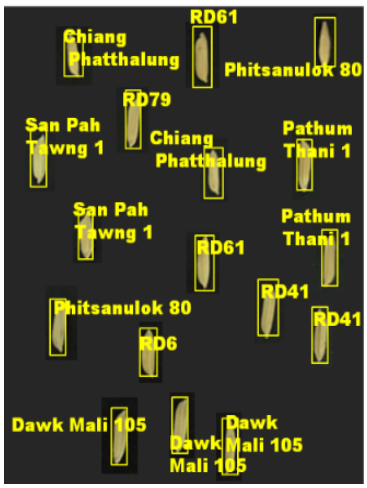
Table 3. Phosop-1 – improved by augmentation using 500 testing grains

Training set (with 2 target classes)	# training grains	mAP (IoU=0.5)	Accuracy
Manually labeled training set	300	0.714	0.76
Labeled training set with augmentation operations	2,400	0.943	1

(a)



(b)



(c)

Figure 5. Grain testing; (a) raw grains tested in many orienataion, (b) grains generated in the same orientation, and (c) localization and classification

5.4. Phosop-2: Paddy grain grading

According to the paddy grain standard inspection, the grain grading was also visionally checked by the grain size that could be divided into high quality (or long-paddy grain) and low quality (short-paddy grain). Furthermore, the glutinous grains were fatter than long-paddy grains and short-paddy grains; but the glutinous grains were often mixed in paddy grain products. For the physical difference between long-paddy and glutinous grain, most length of long-paddy grains were equal or longer than that of glutinonous grains; but the glutinous grains were clearly fatter than the long-paddy grains. Both long-paddy and glutinous grains were longer than short-paddy grains. For an important limitation, the length of paddy grain species (like Phitsanulok 80, Chiang Phattalung and Sang Yod Phattalung) occasionally looked half of short-paddy and long-paddy grain grades that made the Phosop-2 model have the overfitting error as 2%. As well as the Phosop-1, augmentation also improved both localization and classification (as shown in Table 4), the data size was increased from 450 to 3,600 training grains.

Table 4. Phosop-2 – improved by augmentation using 700 testing grains

Training set (with 3 target classes)	# training grains	mAP (IoU=0.5)	Accuracy
Manually labeled training set	450	0.733	0.80
Labeled training set with augmentation operations	**3,600**	**0.972**	**0.98**

5.5. Phosop-3: Grain specie classification

For the seed growing, grain specie purity was really important for farmers because the different grain species affect the different prizes and volumes of productivities. From Tables 3-5, not only the augmentation operations but also the higher number of data could improve the Mask R-CNN localization performance, almost 100%. Using only 1,320 manually-labeled grains covering 11 species, Phosop-3 achieved the accuracy at 94% with the help of augmentation. However, some very similar grain appearance like Dawk Mali 105, Pathum Thani 1 and RD 79 also could not be classified by experts' inspection that were difficult to be classified by the supervised model. Furthermore, PhosopNet was a transfer learning architecture as a source domain that could be transferred to learn more species/samples in the next target domain.

Table 5. Phosop-3 – improved by augmentation using 1,500 testing grains

Training set (with 11 target classes)	# training grains	mAP (IoU=0.5)	Accuracy
Manually labeled training set	1,320	0.874	0.73
Labeled training set with augmentation operations	**11,880**	**0.996**	**0.94**

5.6. Experimental comparisons

The proposed PhosopNet was compared to previous highlight paper MIMR [29] that was based on ResNet-50 [46] for classification. Since the PhosopNet classification was DenseNet that also had image augmentation to increase the number of grains in training set, instead of the full labeling by human. Both MIMR and PhosopNet were localized by Mask R-CNN that already had been proven to be the highest mAP for object localization (compared to other two-stage detections, e.g., R-CNN [50], Fast R-CNN [55], Faster R-CNN [56]), especially for the small objects (like grains) within an image. From Table 6, MIMR [29] did not have the augmentation to increase the size of dataset that made it provide lower accuracy trained by the less data. Another reason was ResNet has only skip connection, while DenseNet [45] had with dense block between the layers that could easily use the feature maps from the low-level layers.

Table 6. Comparisson between MIMR [29] and PhosopNet

Model	Grain localization	Classification	Purity between glutinous and paddy grain	Paddy grain grading	Grain specie classification
MIMR [29]	Mask R-CNN	ResNet-50 with transfer learning	0.71	0.74	0.68
PhosopNet		DenseNet with augmentation and transfer adaptation learning	1	0.98	0.94

According to economic condition, the grain inspection based on bag-of-words model (traditional machine learning with feature extraction) is still required [20-22] in the open-world industry such as iRSVPred [23]. The previous papers [29] has already showed the high correctness (higher than 0.7) in the problems: purity between glutinous and paddy grain (Phosop-1); and paddy grain grading (Phosop-2). The bag of words models still had a problem on too many target classes, like the 11 classes in grain classification

(Phosop-3). Furthermore, some grain species (like Dawk Mali 105 and Pathum Thani 1) were looked very similar. For the solution, image augmentation operations could improve the accuracy classification in a large number of images with target classes. Most traditional machine learning algorithms were support vector machine (SVM) and multi-layer perceptron (MLP) that were frequently used in computer vision. SVM was already proven to be stronger than MLP, especially in a larger number of target classes. And SVM also had adaptive-SVM (Ada-SVM) [35] as transfer adaptation learning mechnism like CNN. For localization, most feature extraction algorithms were originated from scale invariant feature transform (SIFT) [38]. There were many versions of SIFTs, e.g., PCA-SIFT (SIFT with dimension reduction by principal component analysis) [8], speed-up robust feature (SURF) [39], root-SIFT (SIFT with ℓ_1-normalization and square-root) and histogram of gradient (HoG) [42]. From Table 7, not only convolutional neural network but also bag-of-words model could be improved by image augmentation, where HoG with SVM provided the highest accuracy as 84%.

Table 7. Improvement on traditional machine learning with feature extraction tested by 1,500 testing grains

Bag of words	Phosop-3 (11 target classes)	
	without augmentation	with augmentation
SIFT + SVM	0.58	0.76
PCA-SIFT + SVM	0.52	0.71
SURF + SVM	0.55	0.67
root-SIFT + SVM	0.64	0.81
HoG + SVM	0.61	0.84

6. CONCLUSION

As referred to the expensive labeling on too many small grains, the proposed PhosopNet has achieved the high performance in terms of grain localization and classification using the less labeled data training. The augmentation is the behind technique to generate more grain data by rotation, brightness adjustment and horizontal flipping. PhosopNet has Mask R-CNN for grain localization and DenseNet for grain classification. DenseNet is a transfer learning architecture that consists of transfer learning–learning some data with target classes in one stage and more data with new target classes in the next stage; and domain adaptaion–learning some data with target classes in one stage and more data with the same classes in the next stage. According to the grain standard inspection in the real-world, the experiments are divided into 3 groups: Phosop-1 as glutinous grain and paddy grain classification, Phosop-2 as glutinous grain, long-grain paddy and short-grain paddy classification and Phosop-3 as 11 grain specie classification. Moreover, the augmentation improves not only convolutional neural network but also bag of words. For the main finding, the less labeled data is possible to achieve high correctness in both localization and classification. The shortcoming like the similar grain appearance may be alleviated by pseudo labeling (or self-supervision) that some labeled data is trained in the learning model; another unlabeled data is later classified and pseudo-labeled by the model. For the outlook and direction, the seed recognition (both CNN and bag of words) like rice-grains, weeds or beans will absolutely not needs the iteratively manual labeling process by human labor for training those large-scale small seeds.

7. ACKNOWLEDGEMENTS

To do more with less number of labeled grains, the “PhosopNet” was proposed by extending the previous works on grain recognition in both bag of words [9-23] and CNN models [24-29]. All images in the paper were watermarked and copyrighted. All grains in this paper were visionly classified by the expert from Grain Quailty Inspection Laboratory, Rice Department, Ministry of Agriculture and Cooperatives, Thailand. All computational resources were supported by Chandrakasem Rajabhat University.

REFERENCES

[1] J. Roche, “History of Rice,” in *the International Rice Trade*. Cambridge, United Kingdom: Woodhead, 1992, pp. 14–21.
[2] The 11th World Rice Conference. The Rice Trader. 2019. [Online]. Available: https://thericetrader.com/conferences/2019-wrc-manila/worlds-best-rice/. Accessed: Jan 14, 2020.
[3] X. Romero-Frias, “On the Role of Food Habits in the Context of the Identity and Cultural Heritage of South and South East Asia,” 2013. [Online]. Available: https://www.academia.edu/6002651/On_the_Role_of_Food_Habits_in_the_Context_of_the_Identity_and_Cultural_Heritage_of_South_and_South_East_Asia. Accessed: Jan 14, 2020.

[4] K. Davis, "Rice Goddesses of Indonesia, Cambodia and Thailand," 2011. [Online]. Available: https://www.devata.org/rice-goddesses-of-indonesia-cambodia-and-thailand/. Accessed: Jan 14, 2020.
[5] V. Chamarerk, *et al*., "Development of Molecular Markers for Purity Testing in Thai Jasmine Rice," *Advances in Ecological and Environmental Research,* pp. 35-44, 2016.
[6] P. Praphasanobol, *et al*., "Small-Scaled Analysis for Amylose Content in Brown Rice and Thai Rice Clustering based on Amylose Content," *Srinakharinwirot Science Journal,* vol. 36, no. 1, pp. 75-94, 2020.
[7] F. Takeda, *et al*., "A Proposal of Grading System for Fallen Rice using Neural Network," in *2002 International Joint Conference on Neural Networks (IJCNN'02),* 2002, pp. 709-714.
[8] L. Zheng, Y. Yang and Q. Tian, "SIFT Meets CNN: A Decade Survey of Instance Retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 40, no. 5, pp. 1224-1244, 2018.
[9] Z. Y. Liu, *et al*., "Identification of rice seed varieties using neural network," *Journal of Zhejiang University Science B,* vol. 6, no. 11, pp. 1095-1100, 2005.
[10] C-Y Wee, *et al*., "Fast Computation of Zernike Moments for Rice Sorting System," in *2007 IEEE International Conference on Image Processing (ICIP'07),* 2007, pp. 16-168.
[11] D. Xiaopeng and X. Dahong, "Research on Conjoint Grains of Rice based on Machine Vision," in *2010 2nd International Conference on Signal Processing Systems,* 2010, pp. 760-763.
[12] A-G. OuYang, *et al*., "An Automatic Method for Identifying Different Variety of Rice Seeds using Machine Vision Technology," in *2010 6th International Conference on Natural Computation,* 2010, pp. 84-88.
[13] P. T. T. Hong, *et al*., "Identification of Seeds of Different Rice Varieties using Image Processing and Computer Vision Techniques," *Science and Technology Development Journal (STDJ), Vietnam National University,* vol. 13, no. 6, pp. 1036-1042, 2015.
[14] X. Yi, *et al*., "Identification of Morphologically Similar Seeds using Multi-kernel Learning," in *Canadian Conference on Computer and Robot Vision,* 2014, pp. 143-150.
[15] P. T. T. Hong, *et al*., "Comparative Study on Vision Based Rice Seed Varieties Identification," in *2015 International Conference on Knowledge and Systems Engineering (KSE'15),* 2015, pp. 377-382.
[16] T. Y. Kuo, *et al*., "Identifying Rice Grains using Image Analysis and Sparse-representation-based Classification," *Computers and Electronics in Agriculture,* vol.127, pp. 716-725, 2016.
[17] B. Lurstwut and C. Pornpanomchai, "Application of Image Processing and Computer Vision on Rice Seed Germination Analysis," *International Journal of Applied Engineering Research,* vol. 11, no. 9, pp. 6800-6807, 2016.
[18] B. Lurstwut and C. Pornpanomchai, "Image Analysis based on Color, Shape and Texture for Rice Seed (Oryza Sativa L.) Germination Evaluation," *Agriculture and Natural Resources,* vol. 51, no. 5, pp. 383-389, 2017.
[19] A. A. Aznan*, et al*., "Rice Seed Varieties Identification based on Extracted Colour Features using Image Processing and Artificial Neural Network (ANN)," *International Journal on Advanced Science Engineering and Information Technology,* vol. 7, no. 6, pp. 2220-2225, 2017.
[20] H. Nguyen-Quoc, "Rice Seed Images Recognition based on HoG Descriptor and Trimming Imputation Approach," in *11th International Academic Conference Global Goals, Local Actions: Looking Back and Moving Forward,* 2020, pp. 126-134.
[21] H. Nguyen-Quoc and V. T. Hoang, "Rice Seed Image Classification based on HoG Descriptor with Missing Values Imputation," *TELKOMNIKA Telecommunication, Computing, Electronics and Control,* vol. 18, no. 4, pp.1897-1903, 2020.
[22] S. D. Fabiyi, *et al*., "Varietal Classification of Rice Seeds using RGB and Hyperspectral Images," *IEEE Access,* vol. 8, pp. 22493-22505, 2020.
[23] A. Sharma, *et al*., "iRSVPred: A Web Server for Artificial Intelligence based Prediction of Major Basmati Paddy Seed Varieties," *Frontiers in Plant Science,* vol. 10, pp. 1791-1800, 2019.
[24] I. Chatnuntawech, *et al*., "Rice Classification using Spatio-Spectral Deep Convolutional Neural Network," 2018. [Online]. Available: https://arxiv.org/ftp/arxiv/papers/1805/1805.11491.pdf. Accessed: Jan 14, 2020.
[25] Z. Qiu, *et al*., "Variety Identification of Single Rice Seed Using Hyperspectral Imaging Combined with Convolutional Neural Network," *Applied Science,* vol. 8, no. 2, pp. 121-224, 2018.
[26] B. Li and S. Li, "Recognition Algorithm of Rice Germ Integrity base on Improved Inception v3," in *2019 International Conference on Intelligent Computing, Automation and Systems (ICICAS'19),* 2019, pp. 497-501.
[27] Y. Chen, *et al*., "A Deep Multi-View Learning Method for Rice Grading," in *2019 IEEE International Conference on Real-time Computing and Robotics (RCAR),* 2019, pp. 726-730.
[28] S. Mathulaprangsan, *et al*., "Rice Diseases Recognition using Effective Deep Learning Models," in *2020 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON'20),* 2020, pp. 386-389.
[29] K. Aukkapinyo, *et al*., "Localization and Classification of Rice-grain Images using Region Proposals-based Convolutional Neural Network," *International Journal of Automation and Computing,* vol. 17, pp. 233-246, 2020.
[30] A. Chaugule, "Application of Image Processing in Seed Technology: A Survey," *International Journal of Emerging Technology and Advanced Engineering,* vol. 2 no. 4. pp. 153-159, 2012.
[31] A. Olaode and G. Naghdy, "Review of the Application of Machine Learning to the Automatic Semantic Annotation of Images," *IET Image Processing,* vol. 13, no. 8, pp. 1232-1245, 2020.
[32] Md. Z. Alom, *et al*., "The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches," 2017. [Online]. Available: https://arxiv.org/ftp/arxiv/papers/1803/1803.01164.pdf. Accessed: Jan 14, 2020.

[33] Z. Zou, *et al*., "Object Detection in 20 Years: A Survey," 2019. [Online]. Available: https://arxiv.org/pdf/1905.05055.pdf. Accessed: Jan 14, 2020.
[34] D. P. V. Hoai, *et al*., "A Comparative Study of Rice Variety Classification based on Deep Learning and Hand-crafted Features," *ECTI Transactions on Computer and Information Technology,* vol. 14, no. 1, pp. 1-10, 2020.
[35] R. Collobert and S. Bengio, "Links between Perceptrons, MLPs and SVMs," in *2004 International Conference on Machine Learning (ICML'04),* 2004, pp. 23.
[36] L. Zhang, "Transfer Adaptation Learning: A Deacade Survey," 2019. [Online]. Available: https://arxiv.org/pdf/1903.04687.pdf. Accessed: Jan 14, 2020.
[37] J. Yang, R. Yan and A. G. Hauptmann, "Cross-domain Video Concept Detection using Adaptive SVMs," in *2007 ACM International Conference on Multimedia,* 2007, pp.188-197.
[38] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision,* vol. 60, pp. 91-110, 2004.
[39] H. Bay, T. Tuytelaars and L. V. Gool, "SURF: Speeded Up Robust Features," in *European Conference on Computer Vision (ECCV'06),* 2006, pp. 404-417.
[40] R. Arandjelović and A. Zisserman, "Three Things Everyone Should Know to Improve Object Retrieval," in *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'12),* 2012, pp. 2911-2918.
[41] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *2012 International Conference on Neural Information Processing Systems (NIPS'12),* pp. 1097-1105, 2012.
[42] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05),* pp. 886-893, 2005.
[43] C. Szegedy, *et al*., "Going Deeper with Convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15),* 2015, pp. 1-9.
[44] C. Szegedy, *et al*., "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," 2016. [Online]. Available: https://arxiv.org/pdf/1602.07261.pdf. Accessed: Jan 14, 2020.
[45] G. Huang, *et al*., "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'17),* 2017, pp. 2261-2269.
[46] K. He, *et al.*, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'16),* 2016, pp. 770-778.
[47] B. Zoph, *et al.*, "Learning Transferable Architectures for Scalable Image Recognition," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'18),* 2018, pp. 8697-8710.
[48] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2015. [Online]. Available: https://arxiv.org/pdf/1409.1556.pdf. Accessed: Jan 14, 2020.
[49] L. Jing and Y. Tian, "Self-supervised Visual Feature Learning with Deep Neural Networks: A Survey," 2019. [Online]. Available: https://arxiv.org/pdf/1902.06162.pdf. Accessed: Jan 14, 2020.
[50] K. He, *et al.*, "Mask R-CNN," in *2017 IEEE International Conference on Computer Vision (ICCV'17),* 2017, pp. 2980-2988.
[51] Y. Toda, *et al.*, "Training Instance Segmentation Neural Network with Synthetic Datasets for Crop Seed Phenotyping," *Nature Communications Biology,* vol. 3, no. 1, pp. 173-185, 2020.
[52] A. Olaode, G. Naghdy and C. Todd, "Unsupervised Classification of Images: A Review," *International Journal of Image Processing,* vol. 8, no. 5, pp. 325-342, 2014.
[53] L. Liu, *et al.*, "Deep Learning for Generic Object Detection: A Survey," *International Journal of Computer Vision,* vol. 128, pp. 216-318, 2020.
[54] R. Girshick, *et al.*, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," 2014. [Online]. Available: https://arxiv.org/pdf/1311.2524.pdf. Accessed: Jan 14, 2020.
[55] R. Girshick, "Fast R-CNN," in *2015 IEEE International Conference on Computer Vision (ICCV'15),* 2015, pp. 1440-1448.
[56] S. Ren, *et al.,* "Faster R-CNN: towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 39, no. 6, pp. 1137-1149, 2017.
[57] T-Y. Lin, *et al.,* "Feature Pyramid Networks for Object Detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'17),* 2017, pp. 936-944.
[58] M. Raghu and E. Schmidt, "A Survey of Deep Learning for Scientific Discovery," 2020. [Online]. Available: https://arxiv.org/pdf/2003.11755.pdf. Accessed: April 21, 2020.