










# SSTop3: Sole-Top-Three and Sum-Top-Three Class Prediction Ensemble Method Using Deep Learning Classification Models

Abdulaziz Anorboev<sup>1</sup> , Javokhir Musaev<sup>1</sup> , Jeongkyu Hong<sup>1</sup>  ,  
Ngoc Thanh Nguyen<sup>2</sup> , and Dosam Hwang<sup>1</sup>  

<sup>1</sup> Yeungnam University, Daegu, Republic of Korea

jhong@yu.ac.kr, dosamhwang@gmail.com

<sup>2</sup> Wroclaw University of Science and Technology, Wroclaw, Poland

Ngoc-Thanh.Nguyen@pwr.edu.pl

**Abstract.** Computer Vision (CV) has been employed in several different industries, with remarkable success in image classification applications, such as medicine, production quality control, transportation systems, etc. CV models rely on excessive images to train prospective models. Usually, the process of acquiring images is expensive and time-consuming. In this study, we propose a method that consists of multiple steps to increase image classification accuracy with a small amount of data. In the initial step, we set up multiple datasets from an existing dataset. Because an image carries pixel values between 0 and 255, we divided the images into pixel intervals depending on dataset type. If the dataset is grayscale, the pixel interval is divided into two parts, whereas it is divided into five intervals when the dataset consists of RGB images. In the next step, we trained the model using the original dataset and each created datasets separately. In the training process, each image illustrates a non-identical prediction space where we propose a top-three prediction probability ensemble method. Top-three predictions of newly generated images are ensemble to the corresponding probabilities of the original image. Results demonstrate that learning patterns from each pixel interval and ensemble the top three prediction vastly improves the performance and accuracy and the method can be applied to any model.

**Keywords:** Deep learning ensemble method · Classification task · Image pixel interval

## 1 Introduction

In this work, we focus mainly on two key issues for knowledge transferring and ensemble: what to transfer and when to ensemble. We propose Image Pixel Interval Power (IPIP) that is divided into subsection according to the data type: Image Pixels' Double Representation (IPDR) and Image Pixels' Multiple Representation (IPMR). Detailed explanation of IPDR and IPMR was given in Sect. 3. The second method is Top Three

Prediction that is ensemble prediction probabilities from multiple CNN model to target high accuracy classification.

In our current work, we applied IPDR and IPMR sub methods combined with configuration changes in the Top Three Prediction Probability method. Because our current method consists of multitasks, it begins with dividing the dataset into sub methods, depending on the type of dataset. If the dataset consists of single color images, the IPDR sub method is applied. If the dataset is colorful that is the RGB type, the IPMR sub method is applied. After dividing the image pixels interval, each dataset is trained using the custom-made model. At the next stage of the method, we ensemble top-three or three maximum prediction probabilities of each trained image into the corresponding position within the prediction probabilities of the main model.

The key point to mention in the paper is the increase in accuracy. In deep learning ensemble models, it is challenging to achieve better results with few data samples. The prediction scope of an ensemble is usually located in the prediction scope of the main model and does not allow increasing the accuracy. However, with our method, we partially solved this problem and obtained better results for our models. Also the method does not affect the training of the main model because it is trained separately and includes nearly all knowledge of the main model.

The rest of this work is organized as follows. Section 2 presents previous research works that is related to our method. Section 3 explains the methodology of our work. The experiments and results are reported in Sect. 4. Section 5 presents a summary of this work and discussion for the future work.

## 2 Related Works

Several authors addressed in the past the issues of what, how and when to ensemble and several image preprocessing methods to increase the amount of the data. We review below the most prominent approaches.

Ensemble learning of image pixel values and prediction probabilities-based methods have shown their advantages in numerous research areas, such as image classification, natural language processing, speech recognition, and remote sensing. In recent years, many machine learning [1, 2] and deep learning models [3], including the convolutional neural network [4] and recurrent neural network have been evaluated for image classification tasks. To tackle the shortcomings of conventional classification approaches, [5] proposed a novel ensemble learning paradigm for medical diagnosis with imbalanced data, which consists of three phases: data pre-processing, training base classifier, and final ensemble. CNNs or deep residual networks are used as individual classifiers and random subspaces; to diversify the ensemble system in a simple yet effective manner to further improve the classification accuracy, ensemble and transfer learning is evaluated in [6] to transfer the learnt weights from one individual classifier to another (i.e., CNNs). The generalization of the existing semi-supervised ensembles can be strongly affected by incorrect label estimates produced by ensemble algorithms to train the supervised base learners. [7] proposed cluster-based boosting (Cboost), a multiclass classification algorithm with cluster regularization. In contrast to existing algorithms, Cboost and its base learners jointly perform a cluster-based semi-supervised optimization.

In computer vision, visual information is captured as pixel arrays. These pixel arrays are then processed by convolutions, the de-facto deep learning operator for computer vision. Convolutions treat all image pixels equally regardless of importance, explicitly model all concepts across all images, regardless of content, and struggle to relate spatially distant concepts. Majority of the above reviewed papers skipped the image pixel interval variance knowledge and ensemble of their top three prediction outcomes, which could be an effective tool to increase the classification accuracy of CNN models.

[11] proposed the Image Pixel Interval Power (IPIP) Ensemble method for DL classification tasks. Two sub methods (IPDR and IPMR), which describes IPIP to make other datasets out of original dataset that is used for a DL classification task. In this research, we studied the effect of above-mentioned method and applied ensemble of prediction outcomes in both separate and assemble methods to fulfill the gap in DL.

### 3 Proposed Methodology

**Data Pre-processing:** In the data-processing, we applied the IPIP method, that studies image pixel variance and includes two sub methods: IPDR and IPMR. IPIP is described using IPDR and IPMR. The main contribution of IPIP is the use of datasets copied from the original dataset, leaving certain interval pixel values. The difference in the number of intervals encouraged us to make an initially double representation of the main dataset and multiple representation of the main dataset.

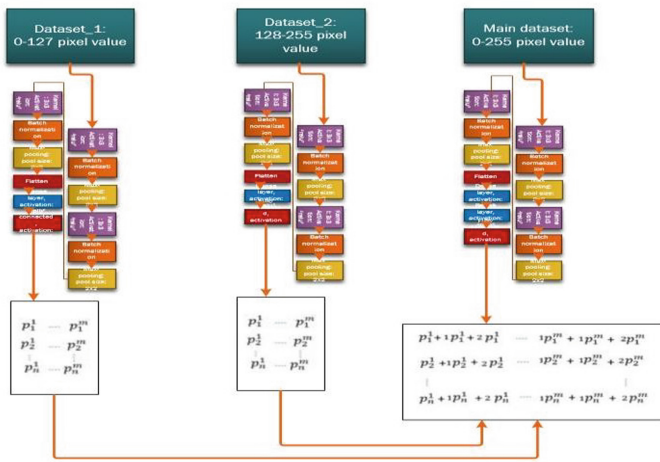


Fig. 1. Sole-Top3 ensemble method for MNIST dataset.

IPDR is a simple double representation of the main dataset. With IPDR, we create two zero arrays (dataset\_1 and dataset\_2) with the same size as the main dataset. In our experiments, we used the MNIST dataset. For this dataset we created two arrays with a size of  $60000 \times 32 \times 32 \times 1$  all filled with zeros. For

dataset\_1, we took only pixel values from the main dataset that belongs to the [0:127] interval and copied and pasted them at the same position in dataset\_1 and all other image pixels in the dataset changed to zeros. Dataset\_2 was also built using the same

method as dataset\_1, except the pixel value interval for dataset\_2 was [128:255]. All values higher than 127 were copied and pasted at the appropriate position in dataset\_2 and all other image pixels in the dataset changed to zeros.

With the IPMR method, we applied the same method as previously described, although instead of two intervals, we used multiple intervals of 50 (i.e., [0:50], [51:100], [101:150], [151:200], and [201:255]) for the Cifar10 dataset. The number of intervals depends on the type of dataset. During our experiments, we found out that to achieve high accuracy in training the RGB channel image, they should be divided into five parts.

**Training the Pre-processed Data with the Model:** In the pre-processed data training, we used two different model architectures with different numbers of parameters. The main model includes 226,122 trainable parameters and is designated for the main dataset. A model of the other created datasets has 160,330 trainable parameters. We chose a smaller model for the created datasets to avoid an overuse of time and power during the IPIP implementation. Generally, the larger the model is, the higher the results achieved. The architecture for the main model consists of three convolutional layers with 32, 64, and 128 filters, and four dense layers with 256, 256, 128, and the number of class nodes for each dataset, respectively. Additionally, we used max pooling with a  $2 \times 2$  filter and batch normalization layers in both model architectures. The filter used for the convolutional layers in both models had a size of  $3 \times 3$ . The architecture of the model includes three convolutional layers with 32, 64, and 128 filters, and three dense layers, which have following numbers of nodes: 256, 128, and the number of classes in each dataset.

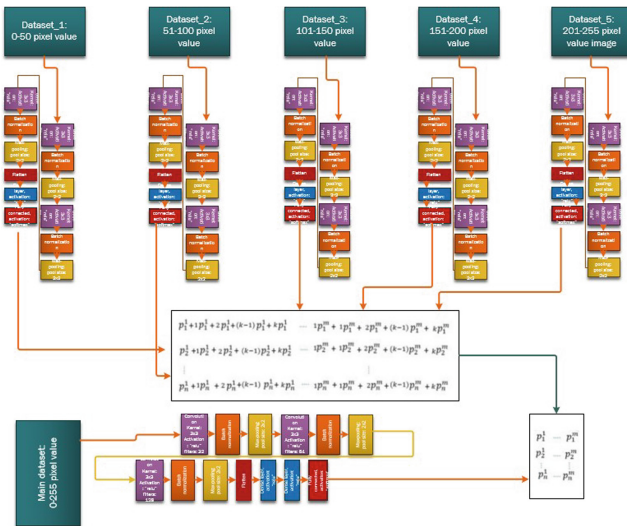


Fig. 2. Sum-Top3 ensemble method for Cifar-10 dataset.

**Prediction Probability Ensemble:** In this part of our method, we propose the Top-Three Prediction Probability Ensemble. Each trained dataset, in our case three datasets for MNIST dataset and six datasets for Cifar10 dataset, represents different prediction spaces. We used the top three prediction probabilities of each image that is trained in the sub model and ensemble them into main models corresponding to the prediction probabilities. The ensemble process was experimented in two distinct ways; Sole-Ensemble and Sum-Ensemble. The workflow of the sole-ensemble is shown in Fig. 1; each sub model’s top three prediction probabilities merged to the main model’s corresponding prediction probability. In sum-ensemble as shown in Fig. 2, the aggregate top three prediction probabilities of all sub models were ensemble into the main model’s corresponding prediction probabilities.

## 4 Experiments and Results

### 4.1 Evaluation Metrics

In this work, we proposed a method that mainly focuses on the accuracy of the model. In many other studies including different metrics like the F1 score, Recall, IoU, ROC, etc. For our research, we chose two metrics that meaningfully explains the method’s achievements in different datasets. The accuracy is the ratio of the true predictions to the total number of cases that are used to evaluate the model. Equation 1 shows the calculation of the accuracy.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

*TP-true predicted positive results*  
*TN-true predicted negative results*  
*FP-false predicted positive results*  
*FN-false predicted negative results*

$$UTP(X, Y) = X - X \cap Y \quad (2)$$

*UTP-Unique True Prediction*  
*X-Prediction Scope of a model X*  
*Y-Prediction Scope of a model Y.*

The next evaluation metric is UTP, which identifies the percentage of unique predictions for each model with respect to another one. In Eq. 2,  $UTP(X, Y)$  finds unique true predictions of model X with respect to model Y. This metrics explain why our proposed model achieved better results than the main model where we trained only the main dataset. The indexes of the true predicted images are different in each model and even have the same accuracy. This leads the ensemble to achieve better results. The main motivation of this work was to utilize the knowledge from the pixel level variance by changing the data representation and to achieve better accuracy metrics for the classification task of the work. After changing the data representation, we proposed the ensemble of each sub dataset’s outcomes top three prediction to the main model’s corresponding three predictions. We used classic training as the baseline method for the experimental

evaluations. Classic training was chosen because of the difficulty of finding alternative similar methods that can be used to compare the results. Most of the DL ensemble models focused on model architectures and data representations by image level preprocessing not researching the pixel level variances. The main objective of this method is to apply it in various combinations and with many other DL ensembles simultaneously.

Our work began from preparing the dataset for training by dividing into different parts, depending on the image. Cifar-10 was divided into six types. By applying the best epoch to the test prediction, we chose the top three predictions from each sub model and ensemble into the main model's corresponding class. In the Sole-Top3 ensemble method, we achieved 73,45% accuracy, presented in Table 1. With the help of the Sole-Top3 method we successfully increased the prediction knowledge of the model compared to the IPIP method with 73,38% accuracy. Additionally, we applied the Sum-Top3 method for the trained model. The main difference between the Sum-Top3 method from the Sole-Top3 method is in the ensemble of the prediction probabilities. In the Sole-Top3 method, each sub model's top three predictions were merged to the main model's corresponding prediction class. In the Sum-Top3, the sum of the sub models' top three predictions were ensemble to the main model. Consequently, we achieved 73,58% accuracy much better than in previous methods. The same experiments were conducted with the MNIST dataset, although the data preprocessing was different. The main model was trained using the original MNIST dataset with all pixel values. In the Sole-Top3 method, we achieved 99.01% where the accuracy of the IPIP methods was 98.90%. By applying Sum-Top3 method, prediction metrics figure the accuracy was increased up to 99.07%. The comparison between our previous work's method and the Sole-Top3 and Sum-Top3 is presented in Table 1.

**Table 1.** Test set accuracy for MNIST and CIFAR-10 datasets and proposed methods.

Dataset	Methods	PA
MNIST	IPIP	0.9890
	Sole-Top3	0.9901
	Sum-Top3	0.9907
CIFAR-10	IPIP	0.7338
	Sole-Top3	0.7345
	Sum-Top3	0.7358

## 5 Conclusion

This short work proposed a method for improving the performance of classification tasks based on image pixel interval power and top-three-prediction ensembles in CNN models. The feature ensemble model was built by splitting a database into image intervals and gathering the top three predictions individually and in-group as well. We achieved better

results using both the Sole-Top3 and Sum-Top3 ensembles, by merging the probabilities of sub models to the main model. It is worth noting that the use of any knowledge to improve the metrics of the CNN model will help with the development of Computer Vision in all studies. Selecting the top three prediction rather than whole class prediction probabilities adds more knowledge to the model. We attempted to solve the generalization problem of deep learning using an ensemble of prediction probabilities. There is still a huge gap in this work that needs to be addressed in this field. In the future work we plan to use ontology structures for building a knowledge base for the metadata about the images, which should be an additional sources for the classification tasks [8–10].

**Acknowledgements.** This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No.2022R1F1A1074641).

## References

1. Gaikwad, D.P., Thool, R.C.: Intrusion detection system using bagging ensemble method of machine learning. In: Proceedings – 1st International Conference on Computing, Communication, Control and Automation, ICCUBEA 2015, Jul 2015, pp. 291–295. <https://doi.org/10.1109/ICCUBEA.2015.61>
2. Tolstikhin, I., et al.: MLP-Mixer: an all-MLP architecture for vision, May 2021 [Online]. Available: <http://arxiv.org/abs/2105.01601>
3. Chollet, F.: Xception: deep learning with depthwise separable convolutions, Oct 2016 [Online]. Available: <http://arxiv.org/abs/1610.02357>
4. Krogh, A.: Neural network ensembles, cross validation, and active learning
5. Liu, N., Li, X., Qi, E., Xu, M., Li, L., Gao, B.: A novel ensemble learning paradigm for medical diagnosis with imbalanced data. *IEEE Access* **8**, 171263–171280 (2020). <https://doi.org/10.1109/ACCESS.2020.3014362>
6. Chen, Y., Wang, Y., Gu, Y., He, X., Ghamisi, P., Jia, X.: Deep learning ensemble for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **12**(6), 1882–1897 (2019). <https://doi.org/10.1109/JSTARS.2019.2915259>
7. Soares, R.G.F., Chen, H., Yao, X.: A cluster-based semisupervised ensemble for multiclass classification. *IEEE Trans. Emerg. Top. Comput. Intell.* **1**(6), 408–420 (2017). <https://doi.org/10.1109/TETCI.2017.2743219>
8. Nguyen, N.T., Sobecki, J.: Using consensus methods to construct adaptive interfaces in multimodal web-based systems. *J. Univ. Access Inf. Soc.* **2**(4), 342–358 (2003)
9. Nguyen, N.T.: Conflicts of ontologies – classification and consensus-based methods for resolving. In: Gabrys, B., Howlett, R.J., Jain, L.C. (eds.) *KES 2006. LNCS (LNAI)*, vol. 4252, pp. 267–274. Springer, Heidelberg (2006). [https://doi.org/10.1007/11893004\\_34](https://doi.org/10.1007/11893004_34)
10. Pietranik, M., Nguyen, N.T.: A multi-attribute based framework for ontology aligning. *Neurocomputing* **146**, 276–290 (2014)
11. Anorboev, A., Musaev, J.: An image pixel interval power (IPIP) method using deep learning classification models. Accepted to the proceeding of ACIIDS 2022 Conference