IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# ENSEMBLE LEARNING TECHNIQUES FOR OBJECT DETECTION IN HIGHER RESOLUTION SATELITE IMAGES

**[1]ATHIRA A G, [2]PRAMEEJA PRASIDHAN**

**[1]Msc Scholar, [2]Assistant Professor**

[1,2]Department of Computer Science

[1,2]St.Joseph's College (Autonomous), Irinjalakuda, Thrissur, India

**ABSTRACT**: Ensembling is a system that aims to maximize the detection performance by fusing individual sensors. While infrequently mentioned in deep- literacy papers applied to remote seeing, ensembling styles have been extensively used to achieve high scores in recent data wisdom com desires, similar as Kaggle. The many remote seeing papers mentioning ensembling substantially concentrate on medial resolution images and earth observation Operations similar as land use bracket, but noway on veritably High Resolution( VHR) images for defense- related operations or object discovery. This study aims at reviewing the most applicable ensembling ways to be used for object discovery on veritably high resolution imagery and shows an illustration of the value of similar fashion on a applicable functional use case( vehicle discovery in desert areas).

**KEYWORDS**: Deep learning, Convolutional neural networks, Ensembling, Data augmentation ,Object detection ,Earth observation.

## 1 INTRODUCTION

### 1.1 Object detection

Recent increases in the quality and volume of veritably High Resolution( VHR) Earth observation images from commercially-available satellites has allowed break throughs in tools that can be used for mapping and processing objects and areas of interest to the defence and security communities.Image judges can only assay a small part of all this data and may take hours to find a vehicle in a desert area.Computer vision algorithms and deep literacy results applied to satellite images allow the development of automatic tools to cover large areas in strategic regions. Large amounts of marketable
satellite imagery of veritably high quality can be attained to marker and train algo rithms to descry, classify and identify architectures similar as aircrafts, vehicles,vessels, roads or structures. At a time when it has come easy to classify pussycats and tykes in colorful situations, earth observation raises new challenges by the diversity of the terrain types and conditions, and by the prospects of end- druggies. Ensembling is an approach using a combination of multiple algorithms to achieve better performances than could be

attained from any of the element machine learning algorithm alone. Several academic articles describe the the oretical basis of ensemble learning with first implementations starting within the
1990's. While rarely mentioned in deep-learning papers applied to remote sensing, ensembling methods are widely wont to achieve high scores in recent Kaggle competitions. Remote sensing papers mentioning ensembling mainly focus on mid resolution images and earth observation applications like land use classification, but never on VHR images for prime resolution segmentation.

### 1.2 Ensemble of neural networks

Convolutional neural networks( CNN) is the most promising machine learning algorithm for object discovery and bracket in images. It's a class of supervised algorithms that takes advantage of a large dataset of labeled objects to learn a generalized model of this data. At vaticination time, the association of convolutional kernels with effective GPU make CNN an effective tool with the eventuality for mortal-position delicacy.

Ensemble ways are used in a wide variety of problems of machine literacy. By combining several algorithms, it's known to reduce the bias of a single sensor or its friction. Those problems of bias and friction of vaticination are well known and can be caused by using a model that's too small or too large compared to the dataset or by the fact that earth observation data is too miscellaneous for the model to capture the full picture. In other words, there isn't a single model that can duly model the full set of data to reuse. For case, mounding is known to reduce the bias of the algorithm whereas bagging or Test- Time addition drop its friction.utmost ensembling styles increase the computing time( training time and/ or vaticination time) but give an effective way to push forward the state- of- theart. This study aims at reviewing the most applicable ensembling styles to be used for object discovery grounded on VHR satellite imagery. The named styles have been enforced in a Python- Keras frame and the performance gain of a combination of 3 styles is detailed in the last section.

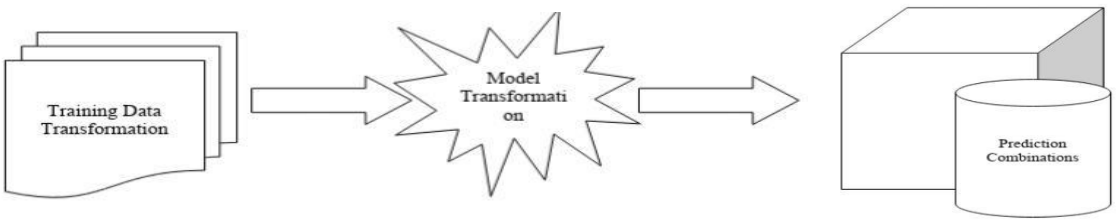
## 2 STATE OF THE ART OF ENSEMBLING TECHNIQUES

## 2.1 Introduction

All ensemble literacy styles relies on combinations of variations in the following spaces

1. Training data space metamorphoses of the data used to train the models
2. Model space metamorphoses of the models trained on the dataset
3. vaticination space metamorphoses and combinations of the vaticination

While exploration papers generally concentrate on single approaches in one of those orders, challenge-winning results use a combination of styles to achieve the loftiest performances. In the following section, the ensemble literacy styles applicable to object discovery grounded on VHR satellite imagery will be listed according to the 3 orders preliminarily introduced.




Figu.1.Architecture of Ensembling Pipeline

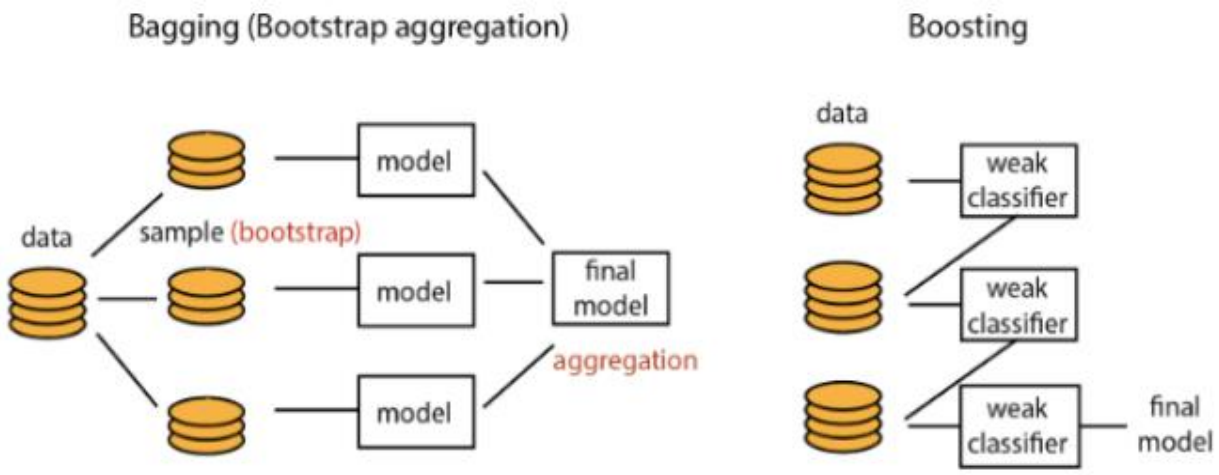



Fig. 2. Comparison between Bagging and Boosting.

**Details**:The main dataset is divided into 3 subsets that are used to train 3 clones of a model. This is done in parallel for Bagging and iteratively for Boosting.

## 2.2 Training data variation- grounded styles

**Bagging:** Bagging, or bootstrap aggregating, generates new data- sets by rooting arbitrarysub-samples from the original bone.Training the model on those arbitrarysub-samples reduces the friction of the training data and improves the delicacy of final ensembling model. Compared to boosting, which samples the original dataset in order to correct the crimes of the former models iteratively, bagging can be used to train models in parallel. Another way of performing bagging without unyoking the dataset, known as " synthetic bagging ", is to use data addition models are also trained on the same dataset with different kinds of accruals.

## 2.3 Model variation- grounded styles

In remote seeing, different CNN infrastructures are known to have different strong- points some models might be better at separating conterminous objects, de tecting lower rudiments or furnishing more precise segmentation. Varying hyperactive parameters, similar as loss and optimizer, also provides intriguing models to combine.

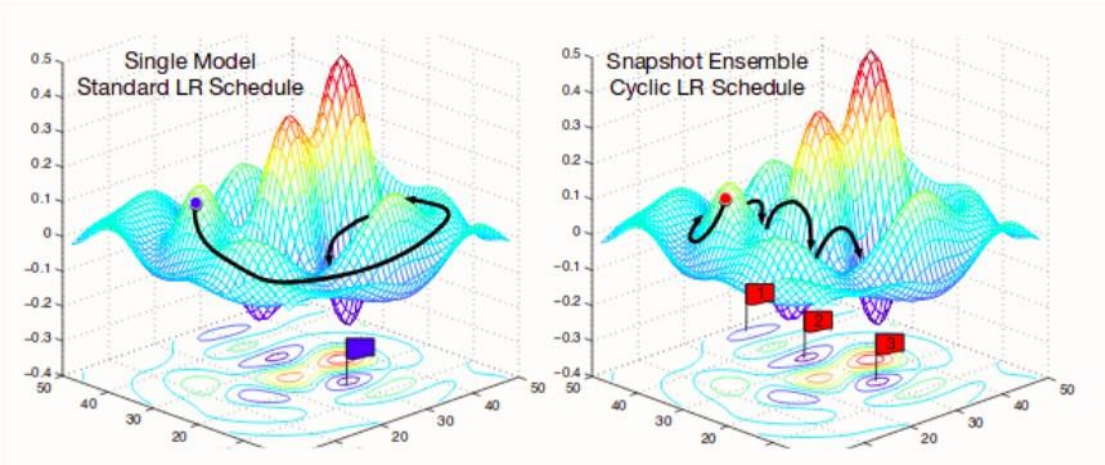



Fig.3. classical SGD optimization vs Snapshot Ensembling

## Shot Ensembling styles

The eponymsub-method shot Ensembling( SSE) is a system furnishing ' cheaply ' a large number of reciprocal CNN that could be combined to ameliorate performances. While those CNN partake the same armature, the use of a clever " cyclical scheduling " of the literacy rate allows the networks to reach different wide or narrow original optima along training. The literacy rate is gradationally dropped to allow the network to meet to some original minima. The algorithm also takes a shot of the weights, before adding the literacy rate again in order to ' jump ' to another optimum. This cycle can be repeated numerous times to produce a set of shots of the asked size. Fast Geometric Ensembling( FGE) is a variation of shot ensembling using a different scheduling scheme fleetly cycling direct piecewise variations of the literacy rate rather of the original smooth cosine one. This approach is justified by the actuality, according to the authors of( 6), of low- loss paths between intriguing minimas. FGE is reportedly briskly to train than SSE while showing performances advancements.

## Stochastic Weight Averaging( SWA)

Stochastic Weight Averaging is an optimization system approaching FGE while reducing the computational cost at vaticination time. During training, a combination of two models the first keeps a running normal of the weights at each replication, while the alternate model explores the weight space following a cycling literacy rate scheduler. At conclusion, Stochastic Weight Averaging provides a single model with averaged weights.

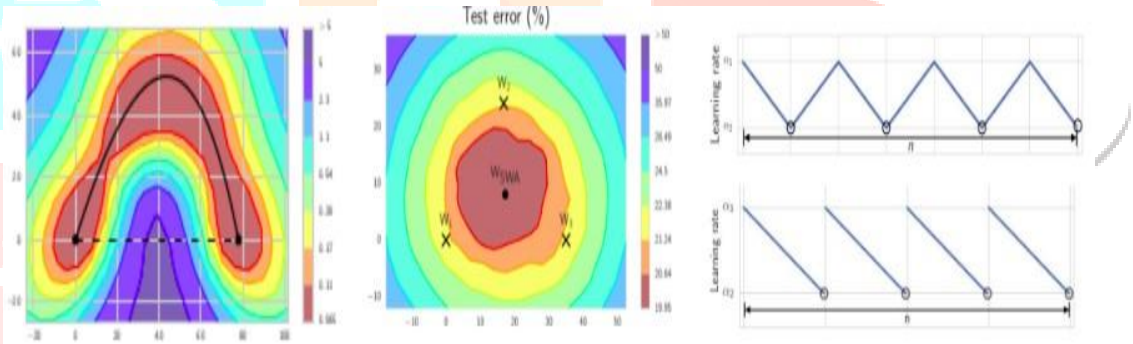



Fig. 4. Illustration of FGE and SWA styles. Left illustration of a low error rate path between 2 original minimas( FGE system). Middle illustration of 3 weights networks W1, W2 and W3, and their weights average WSWA that's on a wider original minima with lower error rate( SWA system). Right the corresponding scheduling schemes of the literacy rate( FGE on the top, SWA on the bottom).

## Hydra

Hydra is a CNN ensembling system that consists of training an original " coarse " model, called the " body " a shot of the training with average performances. From this body, a set of " heads " are fined- tuned by applying colorful data accruals. Just like shot ensembling, this approach provides a set of original minima to combine from a single neural network model.

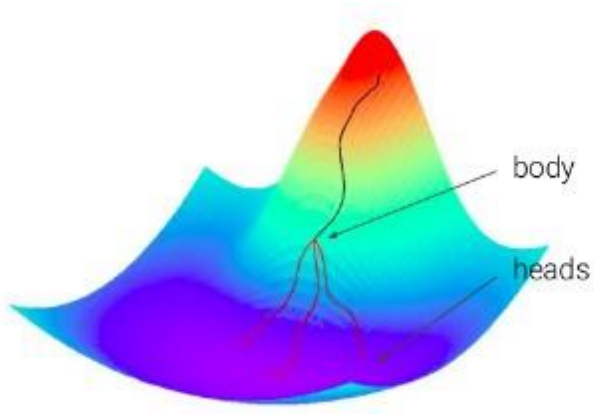



Fig.5 : Illustration of the optimization process in the Hydra frame.

### 2.4 vaticination variation- grounded styles

### Voting- grounded styles

This veritably simple emulsion system combines prognostications by counting the agreeing sensors. This vote could be counted moreover on the raw pixel labors of the sensor orobject-wised from the vectorised shapefile.

### Test Time addition( TTA)

CNN aren't naturally steady to restatement, gyration, scale and other metamorphoses. Applying similar metamorphoses to the input image before feeding it to the network can correct this variability and increase the performances of the model. While TTA doesn't need any new training, it increases linearly the vaticination time.

### Piled conception

mounding relies on a " meta- learner ", which take the prognostications of the existent sensors as its input and is trained to affair a refined affair. While mounding has the eventuality of learning the optimal emulsion of sensors, It's veritably specific to the input sensors and is quite a rigid system as the " meta- learner " needs new to bere-trained if an input sensor is modified.

### Powerhouse as a Bayesian Approximation

Powerhouse is the process of switching off arbitrary neurons within the network during training. During training, this system helps limiting the overfitting and can be seen as an ensemble of networksub-sampled in the model space.introduces the idea of using powerhouse not only during training but also in vaticination stage several prognostications are run by removing different arbitrary neurons from the network. Those prognostications can also be equaled to achieve better generalisation.

## 2 APPLICATION

In this composition, we present the results of the tests of three ensembling ways on the vehicle discovery tasks in VHR images.

## Use case

In object discovery for military purposes, a use case of interest is detecting vehicles in the desert. On the functional side, it's veritably delicate and time consuming to manually find those small objects in the middle of veritably wide areas, as it can be seen onFig. 10. On the machine learning side, the task of detecting vehicles in similar terrain is veritably grueling , indeed for high performance sensors, since a lot of objects look like vehicles and are veritably delicate to distinguish( small trees and jewels, veritably small structures,etc.), which leads to an increase of false cons. An illustration of similar objects is shown onfig. 7.

## Original model

We performed the tests of the ensembling ways applied on a veritably general and high performance model possessed by Earthcube( 92 of F1- score on a different terrain testing set of size31.2 km2 detailed infig. 6). It's a custom ResNet- Unet armature model with8.5 M parameters. It has been trained for the segmentation task on a veritably large training dataset( 330K civil vehicles labeled at a pixel scale) comprising VHR satellite images of different types of surroundings( civic, vegetal, desertic, littoral) and conditions( aerosol, summer, snow,etc.) from spots located each around the world.

**Testing set:** In order to assess and compare the tested ensembling techniques on the use case, we used a very challenging testing set composed by images of desert environment that is particularly difficult for vehicle detection (detailed in fig. 6). We used a very large testing set (composed of 36 different sites for a total area of 608.3 km2) in order to make the assessment of the ensembling techniques relevant.

| | Diverse environment | Desert environment |
|---|---|---|
| Type | diverse *(coastal, desertic, urban, vegetal, etc. )* | desert |
| # of images | 30 | 36 |
| Satellite | WV3 | WV3 |
| Pixel size | 0.3 m | 0.3 m |
| Size | 31.2 km² | 608.3 km² |
| # vehicle per km² | 583.6 per km² | 4.0 per km² |

Fig. 6. Information on the testing sets. Left Information on the different terrain testing set( original model has 92 of F1-score on this testing set). Right Information on the desert terrain testing set( veritably grueling for vehicle discovery since objects look analogous to vehicles).

### Tested ensembling ways

Grounded on the state of the art check, the ensembling ways that are suitable to the considered use case have been named. Since the training dataset is veritably large, the fresh number of ages to train the model on is a crucial constraint. Three ensembling ways have been tested TTA combined with Bayesian Powerhouse, FGE and SWA. TTA and Bayesian Powerhouse don't need any fresh training since are applied at vaticination time. Grounded on former trials, the performance is increased when these two ways are combined. On the other hand, FGE and SWA are promising and only need fine- tuning of the original model so they aren't time- consuming to apply. Other ensembling ways that need a large number of ages to train, similar as Bagging, SSE, and Hydra, have been discarded( and may need a full composition by themselves). Bagging is both time consuming and complex to train since the training dataset size is modified so new tests have to be performed to optimize the model size for eachsub-datasets. SSE needs to train the model from scrape several times in a row and provides lower

performance increase than FGE and SWA according to primary study. Eventually, Hydra needs to insure confluence of veritably different models to diversify accruals and vary abecedarian parameters which is both complex and time- consuming to tune and examiner.

**Result of the tests**

The results of the tests are gathered on the table infig. 8. The named ensembling ways have been tested on two high recall operating points. TTA combined with Bayesian Powerhouse achieves a veritably large performance increase on similar grueling desert terrain0.86 of recall while dwindling the number of false cons by 25 for an operating point with around 92 of recall. This result means that this fashion has functional value since it can boost veritably significantly on grueling terrain a sensor that has formerly functional performance.

Fig. 7. Hard exemplifications in the desert terrain testing set where numerous objects can be confused with vehicles. Top illustration of small trees that look analogous to vehicles. nethermost illustration of veritably small structures that look analogous to vehicles( unheroic points are the vehicles in the ground verity).

FGE achieves a very similar performance boost to TTA / Bayesian Dropout: -31% of false positives while achieving the same recall for the 92% recall operating point. So it provides similar operational value than TTA / Bayesian.

| | Ensembling techniques | Computation time | | Detection performance | | | |
|---|---|---|---|---|---|---|---|
| | | # predictions | # Add. epochs to train | FP/km$^2$ | Recall | FP/km$^2$ relative change | Recall absolute change |
| Operating point 1 | original model | 1 | 0 | 15.04 | 91.85% | - | - |
| | TTA / Bayesian Dropout | 7 | 0 | 11.16 | 92.71% | -25.80% | +0.86% |
| | FGE | 5 | 20 | 10.32 | 91.85% | -31.38% | +0.00% |
| | SWA | 1 | 40 | 14.76 | 92.05% | -1.86% | +0.20% |
| Operating point 2 | original model | 1 | 0 | 6.43 | 86.15% | - | - |
| | TTA / Bayesian Dropout | 7 | 0 | 3.40 | 85.46% | -47.12% | -0.69% |
| | FGE | 5 | 20 | 4.77 | 85.87% | -25.81% | -0.26% |
| | SWA | 1 | 40 | 5.72 | 86.23% | -11.04% | +0.08% |

Fig. 8. Results of the tests on ensembling ways on the desert terrain testing set.

**Interpretation:** Those ways are compared on two operating points. The first bone has a veritably high recall close to 92, the alternate bone has a lower recall close to 86 with lower false findings. For each fashion and each operating point, we assessed both the discovery performance and the calculation time.

**Discovery performance metric:** For the discovery performance, we assessed the performance gain by using both the absolute increase in recall, and the relative drop of false cons per km2. We haven't used the F1- score metric since it isn't suitable for desert areas there are veritably many vehicles in veritably large areas, which make the F1- score veritably unstable( see fig. 6).

**Calculation time metric :**To assess the vaticination time, we use the number of prognostications to perform so that it doesn't depend on the available tackle or the image size. For training time, we used the number of fresh ages to train the model.

**Parameters optimization:** The parameters of each fashion( number of shots to fuse for FGE and SWA, and number and nature of the metamorphoses for TTA/ Bayesian Powerhouse) have been preliminarily optimized. For illustration, a rate of gain of performance versus fresh calculation time has been maximized.

Eventually, SWA achieves a significant performance boost indeed though it's lower than TTA/ Bayesian Powerhouse and FGE-11 of false cons on the 85 recall operating point, and small performance increase on the 92 recall operating point. still, this fashion only needs one vaticination( rather than independently 7 and 5 for TTA/ Bayesian Powerhouse and FGE). So this fashion is largely recommended for operations that can not use fresh time during vaticination. One could conclude that for a veritably high position of recall( 90), substantial performance boosts can only be achieved through voting- grounded ways( similar as TTA/ Bayesian and FGE) and that using one unique model might not be sufficient.

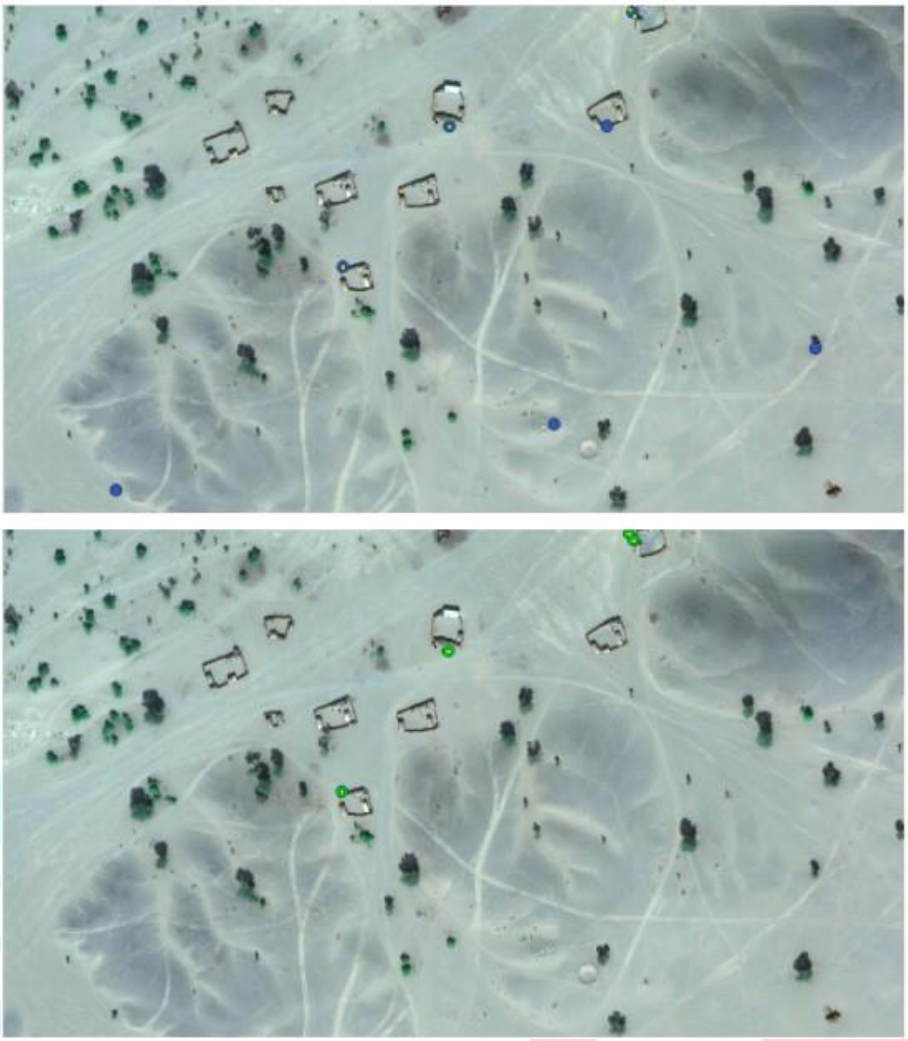


Fig. 9. illustration of findings on the desert terrain testing set.

**Top:** Discovery of the original model in blue( operating point 1) and ground verity in unheroic.

**Bottom:** Discovery of the model with TTA and Bayesian Powerhouse combined in green( operating point 1) and ground verity in unheroic.

**Interpretation:** We observe that the original model has a high recall and provides some false findings on small trees and structures that look analogous to vehicles. The combination of TTA with Bayesian Powerhouse removes these false dictions and increase the recall by detecting an fresh vehicle on the top.

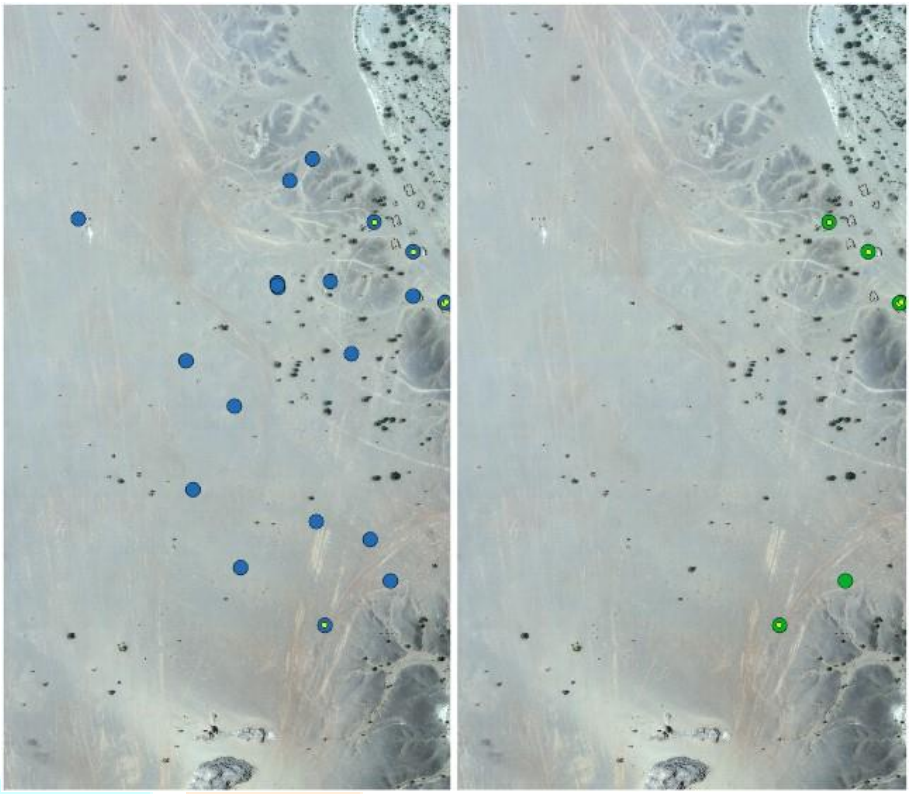


Fig. 10. illustration of findings on the desert terrain testing set.

**Left:** Discovery of the original model in blue( operating point 1) and ground verity in unheroic.

**Right:** Discovery of the model with TTA and Bayesian Powerhouse combined in green( operating point 1) and ground verity in unheroic.

**Interpretation :**We observe on the left side that the original model achieves a veritably good recall with some false findings. The discovery using ensembling removes utmost of these false findings without affecting the recall and increase the overall performance.

# 3 Conclusion

Ensembling ways give a toolbox to increase performances of a given model with a given training dataset, at each position of the discovery channel( training data, model, vaticination). After detailing the most applicable ensembling ways for object discovery in VHR images, three ways have been named for the use case of vehicle discovery on VHR images with an original veritably high performance model trained on a veritably large dataset. These ways have been assessed and compared and give substantial performance increases on a veritably grueling desert terrain testing set. At the price of redundant training/ vaticination time, ensembling situations up state of the art algorithms and brings automatic vehicle discovery at the position of a mortal expert. In general, the work presented then allows for better performances and advanced trustability for any object discovery result. This can give the fortified forces more tools to perform their operations, enabling them to automatically prize applicable information from satellite images.

This can be used for specific tasks similar as relating isolated vehicles in desertic areas or measuring exertion at a given position but also for more astronomically defined tasks similar as strategic spots covering. It can be used for any type of object( vehicles, vessels, structures, roads,etc.) and can either be stationed alone without any mortal looking at the images or be applied as a tool to help image judges in their operations.

# References


1. Arthur Vilhelm, Matthieu Limbert, Clément Audebert, Tugdual Ceillier *arXiv preprint arXiv:2202.10554, 2022*
2. Breiman, Leo. "Bagging predictors." Machine learning 24.2 (1996): 123-140.
3. Li, Hanxi and Li, Yi and Porikli, Fatih and Mingwen, Wang. (2016). "ConvolutionalNeural Net Bagging for Online Visual Tracking:." Computer Vision and Image Understanding. 153. 10.1016/j.cviu.2016.07.002.
4. Perez, L., and Wang, J. (2017). The effectiveness of data augmentation in imageclassification using deep learning. arXiv preprint arXiv:1712.04621.
5. Luke Taylor and Geoff Nitschke. Improving deep learning using generic data augmentation, 2017.
6. Gao Huang, Yixuan Li, Geoff Pleiss, Zhuang Liu, John E. Hopcroft, and Kilian Q.Weinberger. Snapshot ensembles: Train 1, get M for free. CoRR, abs/1704.00109, 2017.
7. Timur Garipov, Pavel Izmailov, Dmitrii Podoprikhin, Dmitry Vetrov, and AndrewGordon Wilson. Loss surfaces, mode connectivity, and fast ensembling of dnns, 2018.
8. Pavel Izmailov, Dmitrii Podoprikhin, Timur Garipov, Dmitry Vetrov, and AndrewGordon Wilson. Averaging weights leads to wider optima and better generalization, 2018.
9. Ben Athiwaratkun, Marc Finzi, Pavel Izmailov, and Andrew Gordon Wilson. Improving consistency-based semi-supervised learning with weight averaging, 2018.
10. Rodrigo Minetto, Mauricio Pamplona Segundo, and Sudeep Sarkar. Hydra: an ensemble of convolutional neural networks for geospatial land classification, 2018.
11. Roberto Battiti and Anna Maria Colla. Colla, m.: Democracy in neural nets: Voting schemes for classification. neural networks 7(4), 691-707. Neural Networks, 7:691–707, 12 1994.
12. Murat Seckin Ayhan and Philipp Berens. Test-time data augmentation for estimation of heteroscedastic aleatoric uncertainty in deep neural networks. 2018.
13. David H. Wolpert. Stacked generalization. Neural Networks, 5:241–259, 1992.
14. Milad Mohammadi and Subhasis Das. SNN: stacked neural networks. CoRR,abs/1605.08512, 2016.
15. Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning, 2015.