Shapley Consensus Deep Learning for Ensemble Pruning

Youcef Djenouri
IDEAS NCBR, Warsaw, Poland
NORCE Norwegian Research Center, Oslo, Norway
University of South-Eastern Norway, Kongsberg, Norway
youcef.djenouri@ideas-ncbr.pl

Ahmed Nabil Belbachir
NORCE Norwegian Research Center, Grimstad, Norway
nabe@norceresearch.no

Asma Belhadi
OsloMet University, Oslo, Norway
asma.belhadi@oslomet.no

Nassim Belmecheri
Simula Laboratory Research, Oslo, Norway
nassim@simula.no

Tomasz Michalak
IDEAS NCBR, Warsaw, Poland
tomasz.michalak@ideas-ncbr.pl

Abstract

This paper targets a new foundation for designing general-purpose learning systems, by establishing a consensus method that facilitates self-adaptation and flexibility to deal with different learning tasks and different data distribution. We present the Shapely Consensus Deep Learning (SCDL) as a consensus method for general-purpose solutions that do not require the help of domain experts. SCDL is two-level based learning process. In the first level, several deep learning models are trained and the Shapley Value is used to determine the contribution of each subset of models in the training. The models are pruned according to their contribution in the learning process. In the second level, the loss information of each data distribution is saved in the knowledge base. Both levels are explored to prune the models for each new observation. We present the evaluation of the generality of SCDL using different datasets with different shapes, and complexities. The results reveal the effectiveness of SCDL for weakly classification. Concretely, SCDL achieved 90% of AUC with less than 86% for the baseline solutions.

1. Introduction

Deep learning has garnered significant attention in computer vision applications, including classification [22], segmentation [23], and object detection [37]. To facilitate practical deployments, numerous deep learning architectures have been developed. However, the performance of these architectures can vary significantly depending on the data. Consensus learning has been extensively investigated to address this uncertainty in model behavior, where multiple models are combined to achieve improved performance [17, 33]. Two main challenges have not been addressed yet for ensemble learning. The first challenge is that one or more models in the ensemble can contribute negatively to the learning process for a particular data region. This is explained by the fact that the existing consensus learners do not consider the mapping between the data distribution in the training, and the learning steps. The second challenge is that these models are time and memory consuming, where all models in the consensus need to be loaded and executed during the inference phase. To tackle the previously mentioned challenges, we aim to address the following two research questions: 1. Given a collection of models that solve a specific learning task, how can we identify the most effective models? In essence, can we differentiate between models that positively contribute to the learning process and those that hinder it? 2. Once we have identified the best models, how can we leverage them to enhance performance for a particular data distribution during the inference stage? This study endeavors to delve into these research queries by introducing a novel consensus approach called Shapely Consensus Deep Learning (SCDL).

Motivations.

In our quest for ensemble pruning, we approach two pivotal levels: Firstly, we draw from cooperative game theory, leveraging the Shapley Value to equitably distribute benefits among models within a consensus. Each model acts as a player in this cooperative game, and through assessing their contributions, we gain insights into their relative importance. This evaluation extends beyond individual impacts to consider the interactions among models, providing a more comprehensive understanding of their significance. Simultaneously, we harness the power of prior knowledge to enhance deep learning performance. By integrating constraints that enforce consistency in model outputs, we can effectively guide the learning process. This incorporation of prior knowledge has proven fruitful in various machine learning challenges, inspiring us to explore its combined use with insights from data distribution and training data to select the optimal model for each new data instance.

Contributions.

To the best of our knowledge, this is the first piece of work that first thoroughly examines and compute the importance of models using Shapley Value, and then consider the mapping between the data distribution between the training, and the inference steps to effectively address challenges related to adapting automated learning systems. In short, this paper proposes the novel approach SCDL (See Figure 1 for more details) as a foundation for designing general-purpose automated learning systems to tackle the limitations of existing single-purpose models and multi-purpose models on being tailored to specific tasks, and particular data distribution. The main contributions of this research work are given as follows:

1. We propose a SCDL as two level based consensus method for model pruning, which explores the Shapley Value to first compute the contribution of each model in the set of model players that will be used in the learning process. The information derived by Shapley Value will be then used altogether with the knowledge base represented by the training information, and data distribution to efficiently select the best model for the inference purpose.

2. We introduce a novel coalition function to determine the contributions of the models. Instead of taking directly the loss of each model in the group to assess the coalition value of the group of the models, it takes into account the model output for determining the coalition value of the group of the models.

3. We conduct extensive experiment to analyze the components of SCDL using different datasets with different shapes, and complexities on weakly classification task. The results show that SCDL outperforms the baseline solutions for accuracy performance and it is very competitive in terms of the inference runtime.

4. We introduce a polynomial-time computable algorithm (proving its' correctness and time-performance guarantee) for computing another measure of algorithm's contribution for ensemble pruning, that is, the Banzhaf Value. We specifically solve the problem of computing Banzhaf Indices of players in an airport game (defined below). By the fact that computing marginal contributions of algorithms in a given portfolio (as we define it below) is equivalent to solving several airport problems and averaging between them (as was implicitly shown in [9] with respect to the Shapley Value with the use of marginal contribution networks), this also directly gives us an efficient algorithm for computing yet another measure of all models' contribution in a given ensemble.

2. Related Work

Ensemble Pruning A lot of efforts have been invested in exploring the ensemble pruning [2, 3, 11, 13, 21, 24]. All these solutions are even search-based or multi-objective based methods, where they consider heuristics in order to prune the models that contribute negatively to the learning process. Since it is not straightforward to define general heuristics, and objectives for ensemble pruning. These methods gives unacceptable results in handling multiple scenarios. To date, only two study have explored the utilization of Shapley values for the purpose of ensemble pruning [7, 26]. The principal objective of these studies were to ascertain the relative importance of constituent models within an ensemble classifier. However, it is imperative to acknowledge that the aforementioned work is encumbered by several noteworthy limitations: 1. The studies predominantly relies upon intuitive and not robust coalitions, wherein the evaluation of model importance hinges solely on the number of correctly classified instances by each individual classifier. 2. The studies make use of the Monte Carlo approximation and traditional Shapley value techniques [27], which, regrettably, does not yield comprehensive computations of model contributions across all feasible subsets of models for the work of Rozemberczki et al. [26], and require intensive computation for the work of Djenouri et al. [7]. 3. The inquiry conspicuously abstains from an end-to-end framework, focusing exclusively on the elucidation of model contributions without articulating how this newfound knowledge might be advantageous in both the training and inference phases of the ensemble. 4. Importantly, the proposed solution is confined in its applicability, serving only the classification task.

Consensus Learning Consensus learning stands as a potent learning paradigm, harnessing a range of learning models to bolster accuracy and robustness [8, 15, 30, 34]. It encompasses three primary approaches: 1) **Aggregating**: In this method, multiple models are trained on distinct subsets of the training data. The final output is determined by

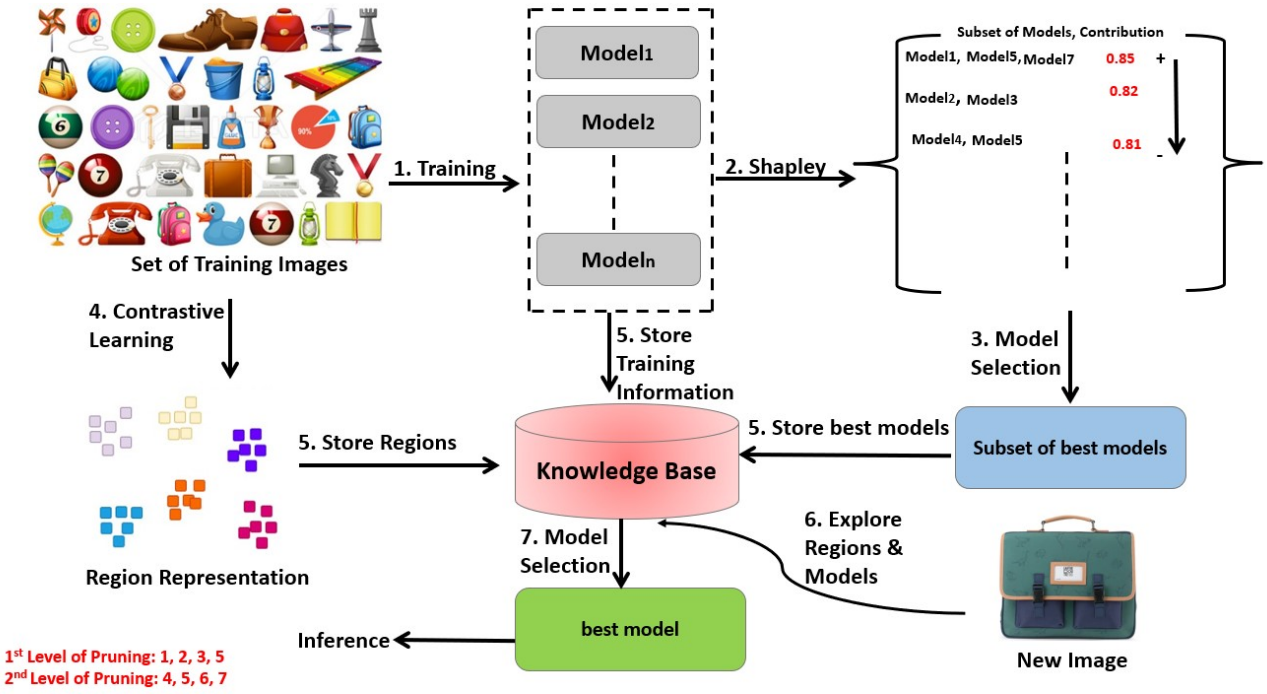

Figure 1. SCDL Overview with two pruning levels: In the first level, the Shapley Value is computed for each subset of models and ranked based on their contributions. In the second level, the training information for each data region is saved and used to select the best model for each new observation.

aggregating the outputs of all models. This aggregation process helps to mitigate overfitting and improve generalization [25, 29, 36]. 2) **Sequential Learning**: Consensus learning sequentially combines multiple weak learners, with each subsequent model focusing on rectifying errors made by its predecessors. This iterative approach contributes to enhancing model accuracy and alleviating bias [10, 28, 32]. 3) **Boosting with Optimization**: This technique amalgamates boosting with optimization strategies, where each weak learner aims to minimize the residual error left by the previous learner. Particularly potent in computer vision tasks, this approach often achieves state-of-the-art performance [6, 10, 14].

Discussion To our knowledge, only two works that explore ensemble pruning [7, 26] using Shapley value, however, it only provides intuitive coalition with Monte Carlo approximation. It did not explain how the determination of the model importance can be used for both the training and the inference stages. It also designed only for solving classification problem. In addition, consensus learning methods offer robustness and accuracy enhancements by harnessing the collective intelligence of multiple models. While these methods often surpass single learning models, they are accompanied by several drawbacks: 1. The memory and time complexity escalate linearly with the size of the consensus (the number of models trained). 2. Low-quality models significantly impact the performance of the best models within the consensus. Our contribution with SCDL is aligned with ensemble pruning and consensus learning. It develop a robust consensus method for ensemble pruning. Moreover, SCDL is generic and might be applied to other data representation including time series, texts, and graphs.

3. SCDL: Shapley Consensus Deep Learning

3.1. Principle

First, we will discuss the major elements of the SCDL approach for general-purpose learning systems. The developed SCDL-based consensus method makes use of deep learning, Shapley Value, and knowledge base as illustrated in Figure 1. Several deep learning models are trained in the learning phase. Once all models finish training and weights are optimized. The training set is again injected to compute the error loss derived by each model for every sample in the training set. Afterwards, the set of distributions is calculated from the training data, where the Shapley Value is used to perform the first-level offline model pruning. The average loss value is then estimated for each data distribution and saved in the knowledge base with the appropriate data distribution. During the inference, the knowledge base is explored altogether with the new observation to perform the second-level online model pruning. This section contains a detailed description of the SCDL components.

3.2. Training

We consider a set of l images used in training, denoted by $I = \{I_1, I_2, ..., I_l\}$. The training process involves a set of n models, represented as $\mathcal{M} = \{\mathcal{M}_1, \mathcal{M}_2, ..., \mathcal{M}_n\}$. Once training is completed, each image I_i is fed into each model $\mathcal{M}_j$ for training. The loss value v_{ij} is then computed by evaluating the error between the output of model $\mathcal{M}_j$ and the ground truth associated with image I_i. This loss value is determined using appropriate loss functions based on the nature of the problem being addressed. For example:

- For classification problems, the Binary Cross-Entropy Loss can be used, defined as:

$$BCE = -\frac{\sum_{i=1}^{N}(y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i))}{N} \tag{1}$$

  where N is the number of samples, y_i is the true label, and $\hat{y}_i$ is the predicted probability.

- For segmentation problems, the Dice Loss can be employed, given by:

$$Dice = 1 - \frac{2\sum_i p_i g_i}{\sum_i p_i^2 + \sum_i g_i^2} \tag{2}$$

  where p_i and g_i are the predicted and ground truth pixel values, respectively.

- For regression tasks, the Mean Squared Error (MSE) Loss is commonly used, defined as:

$$MSE = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N} \tag{3}$$

These loss functions help in quantifying the discrepancy between the model predictions and the actual ground truth, guiding the training process towards convergence.

Definition 1 (Model Projections)**.** *We define the set of projections generated by the model $\mathcal{M}_j$ as the union of all outputs of this model when trained on the set of images I, denoted by:*

$$\mathcal{Y}_j^* = \bigcup_{I_i \in I} y_{ij}^* \tag{4}$$

y_{ij}^* represents the projected value of I_i by the model $\mathcal{M}_j$.

In EQ. (4), $\mathcal{Y}_j^*$ denotes the set of projections generated by model $\mathcal{M}_j$, encompassing the projections made for all images I_i in the training set I. Each y_{ij}^* corresponds to the projected value produced by model $\mathcal{M}_j$ for image I_i.
This definition formalizes the concept of model projections, providing clarity on the projected values generated by each model during the training process.

3.3. Shapley Calculation

Let $\mathcal{P}$ denote the set of models represented as players in the Shapley paradigm. This step aims to determine the contribution of each model in $\mathcal{P}$ to the learning process. Drawing inspiration from the solution concepts or values from cooperative game theory, we measure the importance of each player or model in a coalitional game. While there are numerous ways to evaluate each player's significance, some solution concepts are considered more basic due to the axiom systems that define them specifically. The Shapley Value is an intriguing game-theory concept that has sparked a great deal of interest in the area of deep learning [1]. In the following, we will show how we can adapt the Shapley value to calculate the importance of each model from $\mathcal{MP}$ in the learning process.

Definition 2 (Shapley for Models)**.** *Let $\langle c, \mathcal{P}\rangle$ denote a coalition game, where $c : 2^{\mathcal{P}} \rightarrow \mathbb{R}$ is a set function that assigns utility to each coalition of player subsets in $\mathcal{P}$. We define the coalition value of a subset of players p in $\mathcal{P}$ by the minimum loss values of all models in p compared to the set $\mathcal{Y}$, composed by the ground-truth of all data in I. Let F be all the possible subsets of $\mathcal{P}$ after removing a model player p.*
The associated Shapley Value for each subset of model players $p \in \mathcal{P}$ is defined as:

$$\phi_p = \frac{1}{|\mathcal{P}|}\sum_{F \subseteq \mathcal{P}\setminus\{p\}} \frac{(|\mathcal{P}|-1)!}{(|F|)!(|\mathcal{P}|-|F|-1)!} \tag{5}$$

In Equation (5), ϕ_p represents the Shapley Value for model p in $\mathcal{P}$. It quantifies the marginal contribution of model p to different coalitions of models. The term $\frac{(|\mathcal{P}|-1)!}{(|F|)!(|\mathcal{P}|-|F|-1)!}$ represents the number of possible permutations of coalitions containing p, ensuring that each model's contribution is fairly weighted. The function c evaluates the utility or significance of each coalition, capturing how much better the coalition performs with the addition of model p compared to without it.

Example 1. *Consider the following scenario with three models ($\mathcal{P} = \{\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3\}$) where c evaluates the accuracy improvement of each coalition compared to the individual models:*

$$\begin{aligned} c(\{\mathcal{M}_1\}) &= 0.2 \\ c(\{\mathcal{M}_2\}) &= 0.3 \\ c(\{\mathcal{M}_3\}) &= 0.4 \\ c(\{\mathcal{M}_1, \mathcal{M}_2\}) &= 0.6 \\ c(\{\mathcal{M}_1, \mathcal{M}_3\}) &= 0.7 \\ c(\{\mathcal{M}_2, \mathcal{M}_3\}) &= 0.8 \\ c(\{\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3\}) &= 1.0 \end{aligned}$$

Using Equation (5), we can calculate the Shapley Value for each model:

$$\begin{aligned} \phi_{\mathcal{M}_1} &= \frac{1}{3}\left(\frac{2!(0.6-0.2)}{0!2!} + \frac{2!(0.7-0.6)}{1!1!} + \frac{2!(1.0-0.7)}{2!0!}\right) \\ &= \frac{1}{3}(0.4+0.2+0.3) \\ &= \frac{9}{30} = 0.3 \\ \phi_{\mathcal{M}_2} &= \frac{1}{3}\left(\frac{2!(0.6-0.3)}{0!2!} + \frac{2!(0.6-0.3)}{1!1!} + \frac{2!(1.0-0.8)}{2!0!}\right) \\ &= \frac{1}{3}(0.3+0.2+0.2) \\ &= \frac{7}{30} \approx 0.233 \\ \phi_{\mathcal{M}_3} &= \frac{1}{3}\left(\frac{2!(0.7-0.4)}{0!2!} + \frac{2!(0.8-0.7)}{1!1!} + \frac{2!(1.0-0.8)}{2!0!}\right) \\ &= \frac{1}{3}(0.3+0.1+0.2) \\ &= \frac{6}{30} = 0.2 \end{aligned}$$

In this example, $\phi_{\mathcal{M}_1}$, $\phi_{\mathcal{M}_2}$, and $\phi_{\mathcal{M}_3}$ represent the Shapley Values for models $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3$, respectively. Similarity, we will use the same process for computing the Shapley value for all subsets of models in $\mathcal{M}$. We will then rank the models according to their shapley values. We select the best subsets of models with a reasonable memory and training time costs.

Computing Shapley Values is generally computationally intensive in particular for high number of models. To determine the complexity of calculating the Shapley Value with varying the number of models, we need to analyze the number of coalitions and the computation required for each coalition. Let n be the number of models (players) in the coalition game. To compute the Shapley Value for each model, we need to consider all possible coalitions excluding that model. There are $2^n - 1$ possible coalitions excluding the empty coalition. For each coalition F, we need to compute the marginal contribution $c(F \cup \{p\}) - c(F)$ for adding the model p. The computation of $c(F \cup \{p\})$ and $c(F)$ depends on the specific problem and the way the coalition's value is defined. However, the key point is that for each coalition, we need to compute the value of the coalition with and without the model p, which may involve processing the entire dataset or performing other computations. Therefore, the overall complexity of computing the Shapley Value for each model scales with the number of models and the computation required for each coalition. The complexity can vary depending on the specific problem and the computational resources available. The overall complexity can be expressed as $O((2^n - 1) \times \text{computation\_per\_coalition})$, where **computation_per_coalition** represents the computation required for each coalition.

To reduce the computational time of Shapley, we use the Banzhaf Value which offers an alternative perspective. This value calculates individual contributions differently, taking into account the total contributions such as:

$$\beta(p) = \frac{1}{2^{|\mathcal{P}|-1}} \sum_{F \subseteq \mathcal{P} \setminus \{p\}} (c(F \cup \{p\}) - c(F)) \tag{6}$$

The complexity of calculating the Banzhaf and Shapley Values is similar in terms of the number of coalitions considered. Both values require evaluating the contribution of each player/model by considering all possible coalitions excluding that player/model. However, there is a difference in the computational aspect. While the Shapley Value involves computing the marginal contribution for each player/model in every coalition, the Banzhaf Value considers the total contribution of each player/model across all coalitions. Therefore, the overall computational complexity of calculating the Banzhaf Value can be lower compared to the Shapley Value due to the difference in the computational approach.

3.4. Region Representation

Contrastive learning is a popular technique in the field of computer vision for various tasks, including image region representation. It aims to learn meaningful representations of data by contrasting positive pairs (similar samples) and negative pairs (dissimilar samples). In this part, we will describe contrastive learning for image region representation, providing detailed formulas and explanations. The core of contrastive learning is the contrastive loss function, which encourages similar samples to have similar representations and dissimilar samples to have dissimilar representations. Let $f(x)$ be the feature representation of an image x, and τ be a temperature parameter that controls the sharpness of the similarity scores. The contrastive loss for a pair of images x_i and x_j with labels y_i and y_j can be defined as follows:

$$\mathcal{L}(x_i, x_j, y_i, y_j) = -\log\left(\frac{e^{f(x_i)\cdot f(x_j)/\tau}}{\sum_{b=1}^{l} e^{f(x_i)\cdot f(x_b)/\tau}}\right) \tag{7}$$

In Equation 7, l represents the total number of images in the dataset. When x_i and x_j belong to the same cluster ($y_i = y_j$), the goal is to maximize the similarity score $f(x_i) \cdot f(x_j)$. Conversely, when x_i and x_j belong to different clusters ($y_i \neq y_j$), the goal is to minimize this similarity score.

To use contrastive learning for image region representation, we need to define an objective function that leverages the contrastive loss. Let R be the number of regions, and $\mathbf{z}_i$ be the learned representation for image x_i. We can define the region representation objective function as follows:

$$\mathcal{J} = \sum_{i=1}^{N}\sum_{j=1}^{R} w_{ij} \cdot \mathcal{L}(\mathbf{z}_i, \mathbf{r}_j) \tag{8}$$

$\mathbf{r}_j$ represents the representative image of the region j, and w_{ij} is a weight that measures the similarity between image x_i and $\mathbf{r}_j$. One common choice for w_{ij} is the softmax function:

$$w_{ij} = \frac{e^{s\cdot\mathcal{L}(\mathbf{z}_i,\mathbf{r}_j)}}{\sum_{k=1}^{R} e^{s\cdot\mathcal{L}(\mathbf{z}_i,\mathbf{r}_k)}} \tag{9}$$

In this equation, s is a scaling parameter that controls the temperature of the softmax function. Finally, the proposed contrastive learning algorithm returns the partitioned regions $\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_k$.

3.5. Knowledge Base Creation.

We define the knowledge base $\mathcal{KB}$, which contains $|D|$ rows. Each row corresponds to a specific distribution $\mathcal{D}_i$ of a region $\mathcal{R}_i$. Let $\mathcal{D}_i = \{I_{i1}, I_{i2}, \ldots, I_{in_i}\}$ denote the set of images in region $\mathcal{R}_i$, where n_i is the number of images in that region. The i^{th} row of $\mathcal{KB}$ contains the following information about $\mathcal{D}_i$:

1. Mean (μ_i): The average value of images in $\mathcal{D}_i$, calculated as:

$$\mu_i = \frac{1}{n_i}\sum_{j=1}^{n_i} PixelValues(I_{ij}) \tag{10}$$

2. Standard Deviation (σ_i): A measure of the dispersion of images around the mean μ_i, calculated as:

$$\sigma_i = \sqrt{\frac{1}{n_i}\sum_{j=1}^{n_i}(PixelValues(I_{ij}) - \mu_i)^2} \tag{11}$$

3. Set of Average Loss Values: The set of average loss values $(\mathcal{L}_i(\mathcal{M}_1^{p^*}, \mathcal{D}_i), \mathcal{L}_i(\mathcal{M}_2^{p^*}, \mathcal{D}_i), \ldots, \mathcal{L}_i(\mathcal{M}_{|p^*|}^{p^*}, \mathcal{D}_i))$ associated with each model $\mathcal{M}_j^{p^*}$ in the set p^* when trained on the distribution $\mathcal{D}_i$.

This definition provides detailed information about the knowledge base $\mathcal{KB}$, including the statistical properties of each region's distribution and the associated average loss values for training models on those distributions.

3.6. Inference.

For each new image I_{new}, we first determine which distribution D_{best} in $\mathcal{D}$ fits to I_{new}. We used the moment matching strategy. If the moments (e.g., mean, variance) of the pixel values of the new image match those of the given distribution, then it follows such distribution. Let us consider p_{best} the best models of the selected distribution D_{best}. $y'_{best}(i)$ is the inference output of the model $\mathcal{M}^i_{best}$ on I_{new}. We will use weighted voting for inference the final output y'_{best}. In the weight voting, the output of the best models are averaged by considering the importance of each model in the set p_{best} as follows:

$$y'_{best} = \frac{\sum_{\mathcal{M}^i_{best}\in p_{best}} w^i_{best} \cdot y'_{best}(i)}{\sum_{\mathcal{M}^i_{best}\in p_{best}} w^i_{best}} \tag{12}$$

w^i_{best} represents the weight of the model $\mathcal{M}^i_{best}$. It is calculated by the ratio of the importance of the model $\mathcal{M}^i_{best}$ in the set p_{best}, and it is given as,

$$w^i_{best} = \frac{\mathcal{L}_{best}(\mathcal{M}_i^{p_{best}}, \mathcal{D}_{best})}{\phi_{best}}. \tag{13}$$

In case we consider the Banzhaf function, ϕ_{best} is replaced by β_{best}. The best models p_{best} that will be used in the inference.

4. Performance Evaluation

Datasets. We systematically undertake a series of comprehensive experiments, wherein we employ comparative analyses involving multiple specialized models. These experiments are performed across four prominent computational vision tasks, specifically encompassing instance segmentation, panoptic segmentation, and semantic segmentation, with the utilization of benchmark datasets such as COCO [19], ADE20K [35], and Cityscapes [5].

Results Intensive experiments have been carried out on the three use cases instance, panoptic, and semantic segmentation. We compare SCDL with the SOTA solutions: Mask2Former [4], SoftTeacher [31], SwinV2 [20],

Algorithms	Mask AP	Box AP	AP_{50}^{mask}	AP_{75}^{mask}	params(M)	FPS
Mask2Former	42.5	43.8	61.7	48.9	48	**15**
SoftTeacher	43.8	44.4	62.3	45.3	51	12
SwinV2	41.5	45.8	64.7	47.6	47	8
MaskDino	44.6	46.9	0.89	60.5	**44**	7
SCDL(Shapley)	**45.1**	**48.3**	**68.9**	**51.3**	72	6
SCDL (Banzhaf)	45.0	48.2	68.7	50.5	76	8

Table 1. Comparison of the SOTA instance segmentation models.

Algorithms	Mask AP	Box AP	AP_{50}^{mask}	AP_{75}^{mask}	params(M)	FPS
PanopticSegFormer	44.1	45.6	60.3	47.5	**118**	10
Mask2Former	**48.2**	47.6	61.5	49.8	129	8
SwinV2	42.1	49.2	60.0	51.5	135	6
MaskDino	44.0	45.8	61.3	**52.9**	137	**5**
SCDL(Shapley)	48.1	**50.3**	**63.8**	51.5	144	5
SCDL (Banzhaf)	48.1	50.2	63.7	50.9	152	9

Table 2. Comparison of the SOTA panoptic segmentation models.

Models	Mask AP	Box AP	AP_{50}^{mask}	AP_{75}^{mask}	params	FPS
SeMask	44.2	45.7	60.6	47.9	**131**	9
Mask2Form.	**48.3**	47.7	61.8	49.9	139	10
SwinV2	42.5	49.3	60.8	51.7	141	8
MaskDino	43.9	44.6	60.5	**53.2**	141	**6**
$\text{SCDL}_{Shap.}$	48.5	**50.7**	**64.5**	50.4	152	8
$\text{SCDL}_{Banz.}$	48.4	50.3	63.9	51.2	158	7

Table 3. Comparison of the SOTA semantic segmentation models.

MaskDino [16], PanopticSegFormer [18], SeMask [12]. Tables 1, 2, 3 show the detailed results of this experiment. SCDL outperforms the other solutions in terms of accuracy performance (Mask AP, Box AP, AP_{50}^{mask}, and AP_{75}^{mask}). In addition, SCDL is very competitive to the other solutions compared to the number of trainable parameters, and the inference runtime. These results are obtained thanks to the efficient combination between Shapley Value and knowledge guided mechanisms to select the most relevant models that will be used in the inference stage. Indeed, collaborative intelligence among the models is carried out using the Shapley Value in the first pruning stage. Based on the coalition function, only the models that contribute positively more in the learning process are selected. The returned models are then injected to the second pruning stage, where information related to training are used to select the best models that will be used in the inference.

Ablation Study In this section, we ablate the effectiveness of components in SCDL using COCO, ADE20K and Cityscapes benchmarks. Table 4 shows the effect of each component of SCDL in improving the AP_{75}^{mask}. We examine eight model configurations to test the success of each component in the SCDL by either choosing the Shapley or Banzhaf for the first pruning strategy, and either choosing the contrastive learning or naive partitioning for region representation in the second pruning strategy. Note that the naive partition aims to randomly split the images into k different regions. The comprehensive model consistently attains the highest performance scores across all experimental

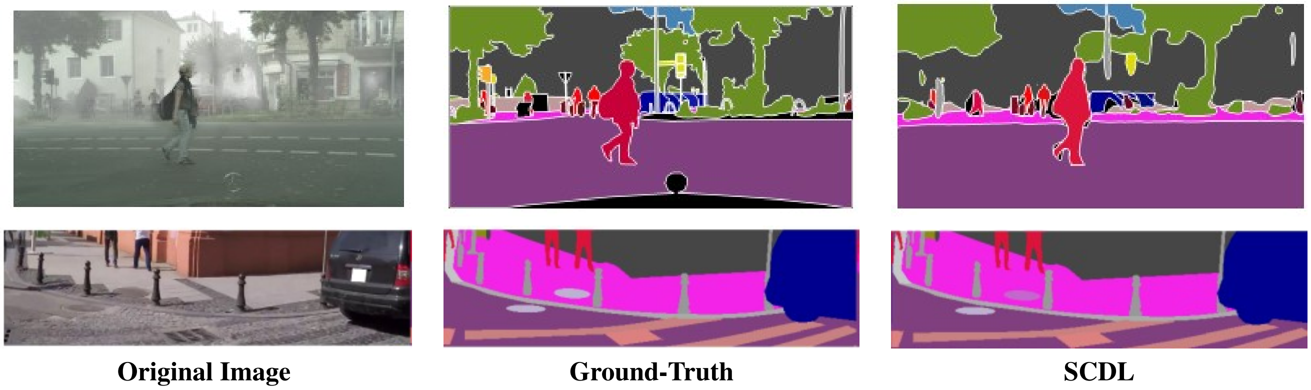

Figure 2. Qualitative results of the SCDL where the other baseline solutions fail.

Models	COCO	ADE20K	Cityscapes
Shapley	48.3	49.3	48.1
Banzhaf	47.9	48.5	48.3
Naive	47.2	48.0	48.6
Contrastive Learning	49.2	49.1	48.8
Shapley-Naive	51.0	51.2	50.1
Banzhaf-Naive	50.2	50.5	50.9
Shapley-Contrastive	51.3	51.5	50.4
Banzhaf-Contrastive	50.5	50.9	51.2

Table 4. AP_{75}^{mask} performance of each component of SCDL.

datasets, exhibiting a notable enhancement in the AP_{75}^{mask} metric when juxtaposed with initial and intermediate levels of analysis employing a simplistic partitioning strategy. Moreover, our findings underscore the significance of leveraging contrastive learning methodologies to delineate regions within the image domain.

Qualitative Results Further experiments were carried out to specifically tackle segmentation tasks, with the results depicted in Figure 2. The figure showcases scenarios where SCDL surpassed other baseline solutions. These results highlight SCDL's ability to accurately delineate various shapes of differing sizes, a task that poses challenges for conventional baseline methods. This notable success can be attributed to the synergistic interaction between Shapley-based techniques and the knowledge base. These techniques effectively trim and guide the SCDL models towards convergence with the global optimum. Consequently, the selection of the inference model depends on the collaborative dynamics among models and the coherence of their training data.

5. Conclusion

This study addresses the challenges associated with the establishment of versatile and adaptable learning vision systems using existing models and introduces SCDL framework for achieving consensus modeling that is agnostic to both task-specificity and data characteristics. Within the SCDL framework, multiple deep learning models are employed for training on each set of images. The Shapley Value is computed to quantify the contribution of each subset of models to the training process. The training information pertaining to the regions of images is retained and utilized during the inference phase for the final model pruning. To enhance the efficiency of Shapley computation, an investigation into the Banzhaf function is conducted. The performance of the SCDL approach is assessed using diverse datasets with varying shapes and complexities. The experimental results validate the superior accuracy of SCDL when compared to baseline methods. Given the critical runtime considerations associated with SCDL, particularly as the number of models increases, and for real-time processing applications, we intend to enhance model exploration by exploring alternative heuristics beyond Shapley and Banzhaf Values. Additionally, we plan to explore more efficient methods for image space and knowledge base analysis to address scalability concerns in SCDL. We will also investigate SCDL's potential for other learning tasks and conduct additional case studies as part of our future research agenda.

Acknowledgments

This work is funded by the Research Council of Norway under the project entitled ”Next Generation 3D Machine Vision with Embedded Visual Computing” with grant number 325748. It is also partially funded by the European Commission through the AI4CCAM project (Trustworthy AI for Connected, Cooperative Automated Mobility) under grant agreement No 101076911.

References

[1] Lucas Agussurja, Xinyi Xu, and Bryan Kian Hsiang Low. On the convergence of the shapley value in parametric bayesian learning games. In *International Conference on Machine Learning*, pages 180–196. PMLR, 2022. 4

[2] Yijun Bian, Yijun Wang, Yaqiang Yao, and Huanhuan Chen. Ensemble pruning based on objection maximization with a general distributed framework. *IEEE transactions on neural networks and learning systems*, 31(9):3766–3774, 2019. 2

[3] Sebastian Buschjäger and Katharina Morik. Joint leaf-refinement and ensemble pruning through l 1 regularization. *Data Mining and Knowledge Discovery*, 37(3):1230–1261, 2023. 2

[4] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1290–1299, 2022. 6

[5] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016. 6

[6] Tusar Kanti Dash, Chinmay Chakraborty, Satyajit Mahapatra, and Ganapati Panda. Gradient boosting machine and efficient combination of features for speech-based detection of covid-19. *IEEE Journal of Biomedical and Health Informatics*, 26(11):5364–5371, 2022. 3

[7] Youcef Djenouri, Ahmed Nabil Belbachir, Tomasz Michalak, and Anis Yazidi. Shapley deep learning: A consensus for general-purpose vision systems. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1224–1233, 2023. 2, 3

[8] Dongyang Fan, Celestine Mendler-Dünner, and Martin Jaggi. Collaborative learning via prediction consensus. *Advances in Neural Information Processing Systems*, 36, 2024. 2

[9] Alexandre Fréchette, Lars Kotthoff, Tomasz Michalak, Talal Rahwan, Holger Hoos, and Kevin Leyton-Brown. Using the shapley value to analyze algorithm portfolios. *Proceedings of the AAAI Conference on Artificial Intelligence*, 30(1), Mar. 2016. 2

[10] Magzhan Gabidolla and Miguel Á Carreira-Perpiñán. Pushing the envelope of gradient boosting forests via globally-optimized oblique trees. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 285–294, 2022. 3

[11] Huaping Guo, Hongbing Liu, Ran Li, Changan Wu, Yibo Guo, and Mingliang Xu. Margin & diversity based ordering ensemble pruning. *Neurocomputing*, 275:237–246, 2018. 2

[12] Jitesh Jain, Anukriti Singh, Nikita Orlov, Zilong Huang, Jiachen Li, Steven Walton, and Humphrey Shi. Semask: Semantically masked transformers for semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 752–761, 2023. 7

[13] Ajay Kumar Jaiswal, Shiwei Liu, Tianlong Chen, Ying Ding, and Zhangyang Wang. Instant soup: Cheap pruning ensembles in a single pass can draw lottery tickets from large models. In *International Conference on Machine Learning*, pages 14691–14701. PMLR, 2023. 2

[14] Xiaojun Jia, Yong Zhang, Baoyuan Wu, Jue Wang, and Xiaochun Cao. Boosting fast adversarial training with learnable adversarial initialization. *IEEE Transactions on Image Processing*, 31:4417–4430, 2022. 3

[15] Lingjing Kong, Tao Lin, Anastasia Koloskova, Martin Jaggi, and Sebastian Stich. Consensus control for decentralized deep learning. In *International Conference on Machine Learning*, pages 5686–5696. PMLR, 2021. 2

[16] Feng Li, Hao Zhang, Huaizhe Xu, Shilong Liu, Lei Zhang, Lionel M Ni, and Heung-Yeung Shum. Mask dino: Towards a unified transformer-based framework for object detection and segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3041–3050, 2023. 7

[17] Zongyi Li, Yuxuan Shi, Hefei Ling, Jiazhong Chen, Qian Wang, and Fengfan Zhou. Reliability exploration with self-ensemble learning for domain adaptive person re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 1527–1535, 2022. 1

[18] Zhiqi Li, Wenhai Wang, Enze Xie, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, Ping Luo, and Tong Lu. Panoptic segformer: Delving deeper into panoptic segmentation with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1280–1289, 2022. 7

[19] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014. 6

[20] Ze Liu, Han Hu, Yutong Lin, Zhuliang Yao, Zhenda Xie, Yixuan Wei, Jia Ning, Yue Cao, Zheng Zhang, Li Dong, et al. Swin transformer v2: Scaling up capacity and resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12009–12019, 2022. 6

[21] Gonzalo Martinez-Munoz, Daniel Hernández-Lobato, and Alberto Suárez. An analysis of ensemble pruning techniques based on ordered aggregation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):245–259, 2008. 2

[22] Sparsh Mittal, Srishti Srivastava, and J Phani Jayanth. A survey of deep learning techniques for underwater image classification. *IEEE Transactions on Neural Networks and Learning Systems*, 2022. 1

[23] Yujian Mo, Yan Wu, Xinneng Yang, Feilin Liu, and Yujun Liao. Review the state-of-the-art technologies of semantic segmentation based on deep learning. *Neurocomputing*, 493:626–646, 2022. 1

[24] Chao Qian, Yang Yu, and Zhi-Hua Zhou. Pareto ensemble pruning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, 2015. 2

[25] Vahid Rowghanian. Underwater image restoration with haar wavelet transform and ensemble of triple correction algorithms using bootstrap aggregation and random forests. *Scientific Reports*, 12(1):8952, 2022. 3

[26] Benedek Rozemberczki and Rik Sarkar. The shapley value of classifiers in ensemble games. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 1558–1567, 2021. 2, 3

[27] Sofiane Touati, Mohammed Said Radjef, and SAIS Lakhdar. A bayesian monte carlo method for computing the shapley value: Application to weighted voting and bin packing games. *Computers & Operations Research*, 125:105094, 2021. 2

[28] Pichao Wang, Xue Wang, Fan Wang, Ming Lin, Shuning Chang, Hao Li, and Rong Jin. Kvt: k-nn attention for boosting vision transformers. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXIV*, pages 285–302. Springer, 2022. 3

[29] Thomas Westfechtel, Hao-Wei Yeh, Qier Meng, Yusuke Mukuta, and Tatsuya Harada. Backprop induced feature weighting for adversarial domain adaptation with iterative label distribution alignment. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 392–401, 2023. 3

[30] Lei Xing, Yawen Song, Badong Chen, Changyuan Yu, and Jing Qin. Incomplete multi-view clustering via correntropy and complement consensus learning. *IEEE Transactions on Multimedia*, 2024. 2

[31] Mengde Xu, Zheng Zhang, Han Hu, Jianfeng Wang, Lijuan Wang, Fangyun Wei, Xiang Bai, and Zicheng Liu. End-to-end semi-supervised object detection with soft teacher. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3060–3069, 2021. 6

[32] Shuyuan Zhang, Lei Wang, Haihui Wang, and Bai Xue. Consensus control for heterogeneous multivehicle systems: An iterative learning approach. *IEEE Transactions on Neural Networks and Learning Systems*, 32(12):5356–5368, 2021. 3

[33] Xingjian Zhen, Zihang Meng, Rudrasis Chakraborty, and Vikas Singh. On the versatile uses of partial distance correlation in deep learning. In *Computer Vision–ECCV 2022: 17th European Conference, 2022, Proceedings, Part XXVI*, pages 327–346. Springer, 2022. 1

[34] Peng Zheng, Jie Qin, Shuo Wang, Tian-Zhu Xiang, and Huan Xiong. Memory-aided contrastive consensus learning for co-salient object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 3687–3695, 2023. 2

[35] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ade20k dataset. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 633–641, 2017. 6

[36] Lei Zhu, Qian Chen, Lujia Jin, Yunfei You, and Yanye Lu. Bagging regional classification activation maps for weakly supervised object localization. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part X*, pages 176–192. Springer, 2022. 3

[37] Zhengxia Zou, Keyan Chen, Zhenwei Shi, Yuhong Guo, and Jieping Ye. Object detection in 20 years: A survey. *Proceedings of the IEEE*, 2023. 1