Towards Generalized Artificial Intelligence by Assessment Aggregation with Applications to Standard and Extreme Classifications

Anonymous authors
Paper under double-blind review

Abstract

The paper provides generalized assessment integration frameworks for multimodel, multimodal and distributed sets of specialized convolutional neural networks. This framework involves two processing stages. In the first stage called specialization, any convolutional neural network learns to predict independently from others. In the second stage called generalization, an integration network learns to predict from assessment measures forwarded by specialized networks: in contrast with the standard learning paradigm based on prediction errors, the integration network learns from assessment measures being considered as deep and densely connected network activations. The multimodel framework is built from different representation functionals associated with different network architectures. The multimodal framework applies on data with different types considered respectively as inputs of a series of cloned or distinct networks. The distributed framework corresponds to local collaborations with assessment exchanges between several specialized networks: it is such that the aggregation aims at determining relevant joint assessments for mapping a given input to a single or a multiple output category. The paper also addresses the difficult and open extreme classification issue in image analysis by proposing a describable texture classification framework. The *describable textures dataset* is used as a case study thanks to its multiple output category annotations. Performance obtained in both standard and extreme classification allows to state that assessment aggregation is a consistent direction for building generalized artificial intelligence.

1 Introduction

Artificial Intelligence (AI) is a science in effervescence due to the considerable advances obtained the last 10 years in connectionist Machine Learning (ML) through deep Convolutional Neural Networks (CNN). After a decade of fine-tuning CNN hyperparameters, we have now empirically proven that CNN variants such as Chollet (2017), Szegedy et al. (2015), Huang et al. (2017), Simonyan & Zisserman (2014), Krizhevsky et al. (2012) are stable and performant in visual object classification. These CNNs, associated with near optimal weights at large, have some interesting generalization properties through transfer learning, but their performance strongly depends on the transfer intricacy and none of them is unquestionably the best. These CNNs will thus be called of *Specialized* AI (SAI) type hereafter.

There exist several plausible scenarios for the construction of a Generalized AI (GAI) that will be unconditionally the best in terms of generalization proof by transfer learning. Indeed, one direction to seek GAI is expanding the numbers of classes, parameters and learning examples. However, this raises commensurability issues and such a direction is, furthermore, biologically questionable since human brain does not only rely on learning by examples, but also on concurrent advice to make decisions in complex situations.

Another direction in establishing GAI is the consideration of collaborative AI systems, where one can essentially distinguish presently from the literature, two widely used approaches. The first approach is associated with *Collaborative ML frameworks*: they consist in building AIs from (i) data that are inaccessible entirely in a limited time Huang et al. (2019), (ii) generative adversarial networks

Chae et al. (2019), (iii) a unified latent space Li et al. (2019), (iv) multivariate data such as those associated with spatial and spectral dimensions Shen et al. (2019). Foundations of these approaches can be found in Wolpert (1992).

The second approach relates to *Collaborative AI agents*: presented as a rule-based request management on multi-agent systems that can be supervised by experts as in Vallam et al. (2019), constrained to data sharing limitations depending on agent reliability Zhao et al. (2020), subject to reinforcement learning as in Chen et al. (2019) (for multi-agent collaborative exploration) and Pan et al. (2019) (for finding homogeneous agents) or managed by semantic collaborative inference Gehrmann et al. (2020).

For collaborative ML frameworks, the goal is to adapt to the distributed availability of data/information in the construction of a ML system and not the deduction of GAIs by aggregation of features from complementary SAI. Therefore, the authors do not develop a proof of GAI concept in the sense of balancing *generalization-versus-specialization* properties. In addition, the same network[1] is shared at central and terminal nodes and this may limit performance since there exists no *universally the most performant network architecture* whatever the data/classification issue at present time.

For collaborative AI agents, agent decisions are counterbalanced by intelligent rules or completed every now and then when complementary: there is no concern on listening agent 'narration' (feature level, not decision level) and learning deeply, the aggregation parameters of a GAI.

We consider in this paper, aggregative learning of very deep interactions between SAI high level features. In this learning framework, the GAI is a set of interconnected independent SAIs whose upstream neurons are Integrated by an AI (IAI). The IAI learns appropriate weights (penalized integration) to compensate the defaults of any SAI under consideration: the collaboration will be performed with respect to the minimization of the same classification objective as that of individual SAI considered so that the GAI will inherit the most important classification capabilities of the integrated SAIs. Thus, the GAI is expected to be more performant than the SAI at large, which suffices to prove the generalization property. This framework will be proven efficient both for establishing the basics of GAIs by aggregation learning and to increase performance on standard and extreme classification tasks.

The organization of the paper is the following. Section 2 presents the GAI framework proposed and its related integration principle. The remaining sections provide different GAI designs from this principle: Section 3 provides multimodel and modal GAI designs by aggregation of SAIs and with respect to the standard classification issue. Section 4 presents some adaptations of the GAI for the extreme classification issue in a possibly distributed framework. Section 5 concludes the work.

Table 1: State-of-the-art performance on DTD standard classification issue.

SIFT based methods				
SIFT-KCB	SIFT-LLC	SIFT-BoVW	SIFT-VLAD	SIFT-IFV
51.08 ± 1.88	52.89 ± 1.52	53.55 ± 1.53	57.34 ± 1.49	58.59 ± 2.04
CNN methods				
From scratch				Transfer
DeepTEN	HistRes	DEP	FV-CNN	LFV-CNN
69.60	71.98 ± 1.23	73.2	73.8 ± 1.0	75.2 ± 1.2

2 GENERALIZED ARTIFICIAL INTELLIGENCE BY ASSESSMENT AGGREGATION

Denote $\mathbb{N}_K = \{1, 2, \ldots, K\}$, where K will represent in practice the number of categories in a standard classification process. A collection $\mathbf{S} = (S_k)_{k \in \mathbb{N}_K}$ of random variables will be called a stochastic assessment process if:

---

[1]The belief is that the network made up of the average of the parameters of several of its clones is a more robust network than each of the cloned network, given that learning is performed at the clone level Konecny et al. (2015), Savazzi et al. (2020).
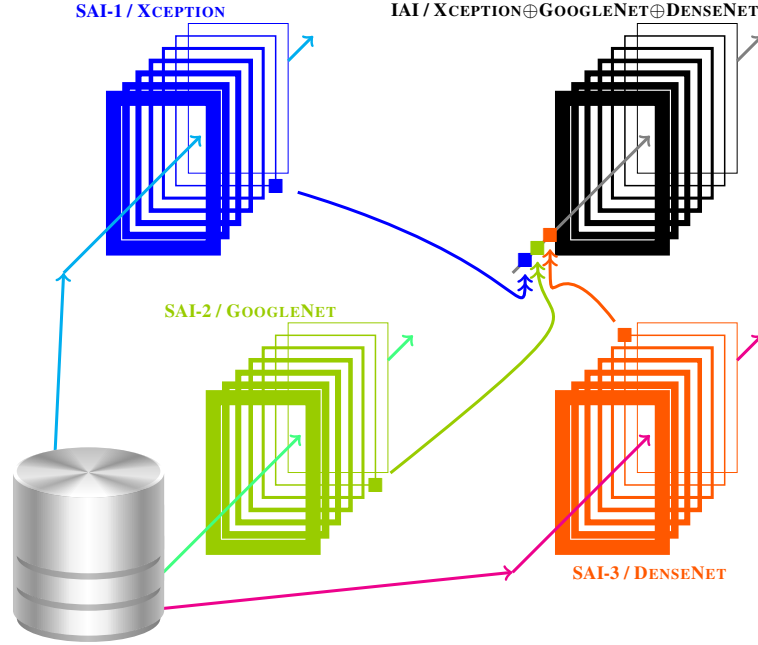
Figure 1: Multimodel assessment GAI composed by $B = 3$ SAIs and a single IAI. Training, validation and testing examples are assumed to pertain to the same observation modality.

- $S_k : \Omega \longrightarrow [0,1]$ represents an objective concept or the set of properties defining a single object or a physical phenomenon (characteristics of a being) and
- for any $\omega \in \Omega$, we have:

$$\sum_{k=1}^{K} S_k(\omega) = 1.$$

The space $[0,1]$ used in this definition represents the set of admissible probabilities for consideration as a being with respect to the essential properties defining S. The universal set Ω contains all elements of a given dataset. For instance, in a multinomial assignment framework, the assessment process is the step consisting in computing the categorical probabilities for an arbitrary query ω to belong to the category indexed by $k \in \mathbb{N}_K$.

One can note that an assessment process is placed in general just before a standard decision/classification making. This is the case for instance in ML frameworks where softmax probabilities:

$$p_k(\omega) = e^{X_k(\omega)} \Big/ \sum_{k=1}^{K} e^{X_k(\omega)}$$

associated with the feature variables $(X_k)_{k \in \mathbb{N}_K}$ are computed prior to mutual-exclusive category assignation. We will use such softmax probabilities to define several stochastic assessment processes $\Lambda = \left\{ \mathbf{S}^{[b]} : b = 1, 2, \ldots, B \right\}$ from B distinct ML frameworks. The goal will then consist in learning an intelligent aggregation system from Λ.

In what follows, any assessment process $\mathbf{S}^{[b]} : \omega \mapsto \mathbf{S}^{[b]}(\omega)$ issued from a standard ML based SAI framework will be called a witness[2]. Given that several witnesses have observed and interpreted some intricate scenes from Ω, the aim of this paper is the derivation of GAI (tribunal or wise juror) by aggregation of witness interpretations. A GAI operating on the countable set Λ of witnesses is a functional

$$\mathcal{F}(\Lambda) = \mathcal{F} \left\{ \mathbf{S}^{[b]}, b = 1, 2, \ldots, B \right\}$$

---

[2]This terminology is chosen, in contrast with 'agent', because the decisions of witnesses are not integrated by the IAI: only their interpretations of the observed scene are integrated.

where $\mathcal{F}$ will be associated with neuro-convolutional frameworks illustrated by Figures 1, 2 and 3 respectively for multimodel, multimodal and distributed aggregation issues. The main difference of the forthcoming developments, in comparison with standard multi-agent processes, consists in the construction of an integration CNN (called IAI) inside the GAI and learning how to correct the probabilities forwarded by SAIs and this, at the IAI level.

Given an observation $\mathcal{I} \in \Omega$, the assessment process sequence $\Lambda \mathcal{I}$) issued from our witnesses (specialized CNNs) is the input (Layer #1) of the IAI network that has in charge the assessment aggregation. Layer #2 of the IAI has input $\Lambda(\mathcal{I})$ and it computes, for any b and any d pertaining to $\{1, 2, \ldots, B\}$, a Geometric Convergence Matrix (GCM):

$$\Theta[b, d](\mathcal{I}) = \sqrt{\mathbf{S}^{[b]}(\mathcal{I})} \sqrt{{}^t\mathbf{S}^{[d]}(\mathcal{I})} = (\Theta_{k,\ell}[b, d](\mathcal{I}))_{k,\ell \in \mathbb{N}_K}$$

satisfying diag $(\Theta[b, b](\mathcal{I})) = \mathbf{S}^{[b]}(\mathcal{I})$, where diag is the diagonal operator of a matrix. Thus, the specialized assessment processes are located on the diagonal of $\Theta[b, b](\mathcal{I})$ whereas a non-diagonal term of $\Theta[b, b](\mathcal{I})$ represents the weight of an assignation of $\mathcal{I}$ to a joint category. Similarly, when $b \neq d$, then $\Theta[b, d](\mathcal{I})$ represents the weights of a cross-assessment assignation of $\mathcal{I}$ either to a given category when considering diag $(\Theta[b, d](\mathcal{I}))$ or to a joint category when focusing on a non-diagonal element of $\Theta[b, d](\mathcal{I})$. The output of layer #2 is thus a double indexed GCM sequence $\left\{ (\Theta_{k,\ell}[b, d](\mathcal{I}))_{k,\ell \in \mathbb{N}_K} : 1 \leqslant b, d \leqslant B \right\}$ that is reshaped as a B^2 multichannel image:

$$\left\{ (\Theta_{k,\ell}[b\prime](\mathcal{I}))_{k,\ell \in \mathbb{N}_K} : 1 \leqslant b\prime \leqslant B^2 \right\}$$

Layer #3 is composed by M convolution operators $\{h_m : m = 1, 2, \ldots, M\}$ that compute a series:

$$\mathcal{C}_m[k, \ell, b] = \sum_{q \in \mathcal{E}_Q} h_m[q] \Theta_{k,\ell}[\tau b - q](\mathcal{I}) \tag{1}$$

associated with MQ learnable linear parameters $(h_m[q])_{m,q}$ and stride parameter τ.

Layer #4 is designed to learn neural transfer features through:

$$
\begin{aligned}
&\Upsilon_{\alpha_m, \beta_m, \gamma_m} (\mathcal{C}_m[k, \ell, b]) \\
&= \begin{cases} \frac{\mathcal{C}_m[k,\ell,b] - \alpha_m}{1 + e^{-\beta(\mathcal{C}_m[k,\ell,b] - \gamma_m)}} & \text{if} \quad \mathcal{C}_m[k, \ell, b] \geqslant \alpha_m \\ 0 & \text{if} \quad \mathcal{C}_m[k, \ell, b] < \alpha_m \end{cases}
\end{aligned} \tag{2}
$$

where $(\alpha_m, \beta_m, \gamma_m)$ are $3M$ non-linear parameters for $m = 1, 2, \ldots, M$.

We can repeat convolutions and additional non-linear transfer functions, however, the above $3M$ non-linearities have accelerated performance in practice and we keep this lightweight CNN design complemented by three standard final layers being respectively associated with (i) a densely connected $K \times M$ linear operator:

$$Z_n = \sum_{m=1}^{M} \sum_{k,\ell=1}^{K} \lambda_{n,m} \sum_{b=1}^{B} \Upsilon_{\alpha_m, \beta_m, \gamma_m} (\mathcal{C}_m[k, \ell, b]) \tag{3}$$

(ii) a softmax operator that computes the new (aggregated) GAI assessment process: $\mathbf{Z} = (Z_n)_{n \in \mathbb{N}_K} \mapsto \mathbf{S}^* = \left(S_n^* = e^{Z_n} / \sum_{n \in \mathbb{N}_K} e^{Z_n} \right)_{n \in \mathbb{N}_K}$ and (iii) a classification layer based on the cross entropy loss for K mutually exclusive classes when considering the standard[3] classification for instance.

3 ASSESSMENT GAI AND STANDARD CLASSIFICATION

A GAI has to show outstanding performance in standard classification tasks. Standard classification issues are those associated with mono-class assignment with respect to a small number of output categories (up to 1 000 classes at present time in ImageNet[4] challenge). Since we will use SAIs

---

[3] An adaptation of this GAI will also be provided later for the extreme classification issue.

[4] ImageNet: large-scale database Russakovsky et al. (2015).

trained on ImageNet, we need an external and complex dataset that does not interact with ImageNet as, on the contrary, testing can be biased when the SAI has already seen the query during its training to classify ImageNet. We consider the Describable Textures Dataset[5] (DTD) as one of the rare complex case studies that is not biased by the ImageNet challenge.

For the standard classification problem, performance of SIFT (Scale-Invariant Feature Transform, Lowe (1999)) based methods given in Cimpoi et al. (2015) are limited to 58%: this confirms that DTD is a complex case study. In addition, the limited number of training samples and the possible confusion between classes (superimposition of elementary perceptual fields) make the most performant CNN associated with a learning from scratch limited to 69% accuracy, see Zhang et al. (2017) (specific design inspired by CNN and GMM called Texture Encoding Network, DeepTEN). Transfer learning has shown the most relevant performance for standard classification of DTD, see Table 1, respectively by using: (i) Histogram based ResNet CNN feature extractor (HistRes) Peeples et al. (2020), (ii) dual texture and average pooling features based ResNet CNN (DEP) Xue et al. (2018), (iii) Fisher Vector CNN (FV-CNN) based on dense and local pooling per region and AlexNet Cimpoi et al. (2015) and (iv) a Local Fisher Vector CNN variant (LFV-CNN) based on VGG Song et al. (2017).

Table 2: Performance of SAIs XCEPTION, GOOGLENET and DENSENET and the GAI given by Figure 1 on DTD standard classification issue: mean accuracy and the maximum absolute deviation (number after the symbol $\pm$) of the accuracy sequences are computed over the 10 splits of the DTD.

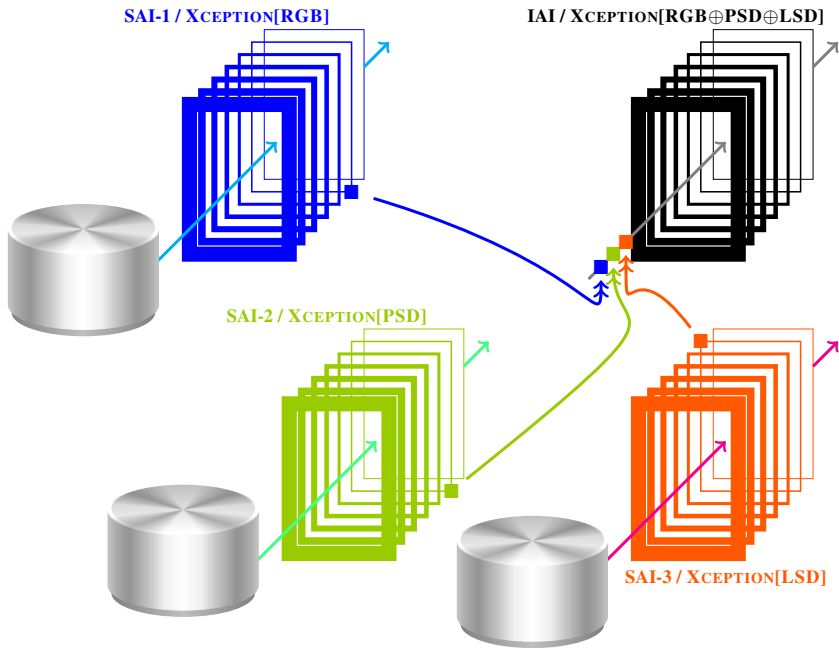| Accuracies of SAI models and the corresponding GAI			
RGB based SAI models			GAI (multimodel)
XCEPTION	GOOGLENET	DENSENET	
72 ± 2	64 ± 3	67 ± 4	74 ± 2

Figure 2: Multimodal GAI: the upstream part of the network learns to efficiently integrate assessment processes issued from the downstream SAIs, the latter have been trained with respect to several image observation modalities. Figure 1

[5]DTD Cimpoi et al. (2015) is collection of 5640 wild textural images annotated with a series of perceptual description characteristics. DTD dataset consists of 47 texture categories, with at least 90% of the image surface covered by the category description characteristics. DTD has been divided in 10 splits, each split being associated with train, validation and test images.

3.1 MULTIMODEL ASSESSMENTS AND STANDARD CLASSIFICATION

Figure 1 provides a block diagram summarizing the multimodel framework for assessment integration: the GAI is composed by a set of downstream CNNs (SAI-1, SAI-2, ..., SAI-B) and an upstream integration CNN (IAI). The later aggregates the assessments issued from the B SAIs. For the sake of illustration, we will assume that $B = 3$ and that SAI-1, SAI-2 and SAI-3 are respectively XCEPTION Chollet (2017), GOOGLENET Szegedy et al. (2015) and DENSENET Huang et al. (2017).

The processing steps for training the multimodel GAI on a standard classification paradigm are given below.

Step 1 (specialization): individual CNNs that are chosen with respect to the diversity criteria on both their diagram configurations and their layer compositions learn independently to classify the training database. Transfer learning[6] is used here since the training examples are very limited[7] for the DTD dataset. At the end of this step, these CNNs are specialized with respect to the DTD standard classification issue and thus, they deserve their calling name of SAIs: performance of some SAIs considered as case studies are given for the DTD test dataset in Table 2. The IAI part of the GAI is turned off in this learning stage.

Step 2 (learning optimal IAI aggregation): SAIs have now fixed weights and IAI is turned on. Any training example $\mathcal{I}$ is sent to the different SAIs and the output SAI assessment sequence, for instance

$$\Lambda_1(\mathcal{I}) = \left\{ \mathbf{S}^{\text{XCEPTION}}(\mathcal{I}), \mathbf{S}^{\text{GOOGLENET}}(\mathcal{I}), \mathbf{S}^{\text{DENSENET}}(\mathcal{I}) \right\}$$

is transmitted to the IAI for integration learning as highlighted in Figure 1: IAI updates its linear and nonlinear weights (see equation 1, equation 2 and equation 3) with respect to a cross-entropy loss.

As it can be seen in Table 2, the GAI obtained outperforms any of the corresponding SAI. Moreover, its performance is on the top-highest reached on DTD and recalled in Table 1 and this, without any resort to expert based assistance for feature extraction as in FV methods given in Table 1.

Table 3: Performance of SAIs XCEPTION$_{\text{RGB}}$, XCEPTION$_{\text{PSD}}$, XCEPTION$_{\text{LSD}}$ and GAI XCEPTION$_{\text{RGB}\oplus\text{PSD}\oplus\text{LSD}}$ (Figure 2) for standard classification: mean accuracy and the maximum absolute deviation (number after the symbol $\pm$) of the accuracy sequences are computed over the 10 splits of the DTD.

Accuracies on DTD for modality based SAIs and the corresponding GAI			GAI (multimodal)
SAI models			
XCEPTION$_{\text{RGB}}$	XCEPTION$_{\text{PSD}}$	XCEPTION$_{\text{LSD}}$	
72 ± 2	71 ± 3	68 ± 5	76 ± 4

3.2 MULTIMODAL ASSESSMENTS AND STANDARD CLASSIFICATION

We now consider the derivation of a GAI adapted to a multimodal image observation frameworks. A main domain where such a GAI can be relevant is *information fusion from different sensors* [Option 1]. However, since any transform of an image by using a projective kernel being the creation of a new observation modality [Option 2], then the results provided here are not limited to standard sensor information fusion. Furthermore, since annotation is usually intricate when considering other observations than standard optical modality[8], then we will use Option 2 and, more specifically, the DTD classification issue under transforms over RGB domain. Figure 2 summarizes the multimodal GAI framework proposed.

We have assumed in Figure 2, a simple case study where the same CNN model is used for analyzing the observation modalities RGB, PSD and LSD selected: RGB is associated by default to DTD

---

[6]The transfer involves learning optimal fully connected layer.

[7]Otherwise, training from scratch must be preferred.

[8]Our eyes are naturally configured only for the spectral band corresponding to visible wavelength, the other bands of the electromagnetic spectrum such as those corresponding to radar or infrared are therefore more difficult for us to unscramble and annotate.

whereas PSD and LSD are respectively Power Spectral Density[9] and Local Standard Deviation[10] transformations on the original RGB DTD. The processing steps for training such a multimodal GAI are given below.

Step 0 (pre-processing, multimodality concern): this step aims at the generation or reorganization of the dataset associated with all the observation modalities selected. For the chosen Option 2, it suffices to generate PSD and LSD since the original DTD dataset is provided an annotated RGB dataset. One can note that this pre-processing step is also required if one chooses Option 1 mentioned above: the sensor acquisitions performed in optical, radar and infrared modalities for instance have to be jointly annotated with respect to the same and pre-specified output categories.

Step 1 (specialization): it consists in training any instance of the selected network architecture with respect to the selected modalities. This step generates modality-specialized SAIs: classification performance of the SAIs specialized with respect to modalities RGB, PSD and LSD of DTD are given when using transfer learning and the XCEPTION network in Table 3.

Step 2 (generalization): the IAI (head of the GAI) takes a mutimodal set of assessment processes:

$$\Lambda_2(\mathcal{I}^{\text{RGB}}, \mathcal{I}^{\text{PSD}}, \mathcal{I}^{\text{LSD}}) = \left\{ \mathbf{S}^{\text{XCEPTION}_{\text{RGB}}}(\mathcal{I}^{\text{RGB}}), \mathbf{S}^{\text{XCEPTION}_{\text{PSD}}}(\mathcal{I}^{\text{PSD}}), \mathbf{S}^{\text{XCEPTION}_{\text{LSD}}}(\mathcal{I}^{\text{LSD}}) \right\}$$

issued from the downstream SAIs (that are trained in Step 1) and this IAI learns to aggregate efficiently the assessment processes from the calculus detailed in Section 2.

Table 3 shows that the GAI obtained outperforms any of the corresponding monomodal SAI. In addition, one can note, by comparing Table 2 and Table 3 that the multimodal GAI outperforms the multimodel one. Modality based diversity integrated in the best relevant network architecture (XCEPTION) is a good catalyst in boosting accuracy. Moreover, we obtain the highest performance ever reached on DTD, up-to-date.

The next section addresses some additional extensions of the GAI framework when considering a decentralized learning scheme and extreme classification tasks.

Table 4: Performance of the GAI described by Figure 3 in extreme classification issues associated with 2 and 3 category interactions. Every GAI involved in this Figure has a form given by Figure 1.

Accuracies of GAIs and DGAIs in extreme classification					
	multimodel GAIs			multimodal DGAI	
	GAI-1	GAI-2	GAI-3	DGAI-1	DGAI-2
2 characteristics	41.86	48.64	39.59	57.69	55.43
3 characteristics	30.39	31.86	27.45	33.82	32.84

4 DISTRIBUTED GAI AND EXTREME CLASSIFICATION

Extreme classification involves testing multiple assignments to non-necessarily independent categories. It is associated with a very large number of output categories, for example when seeking the best representative powerset element, where the powerset applies on the possible combinations of interacting elementary categories. Since an overcomplete expert annotation with respect to millions of categories is unthinkable[11] in a reasonable delay, extreme classification has to deserve a learning paradigm that is different from the one used in standard classification: training has to be complemented by additional rules for testing multiple hypotheses on a given observation.

---

[9]Power Spectral Density: Fourier transform of the 2D autocorrelation function. This transform is used because Fourier transform and its variants are known to be relevant for texture representation.

[10]Local Standard Deviation: replacing a pixel value by the standard deviation computed with respect to data included in a small window associated with the neighborhood of this pixel.

[11]The 1 000 categories of ImageNet are the result of several years of evaluations by a large community of experts and users.
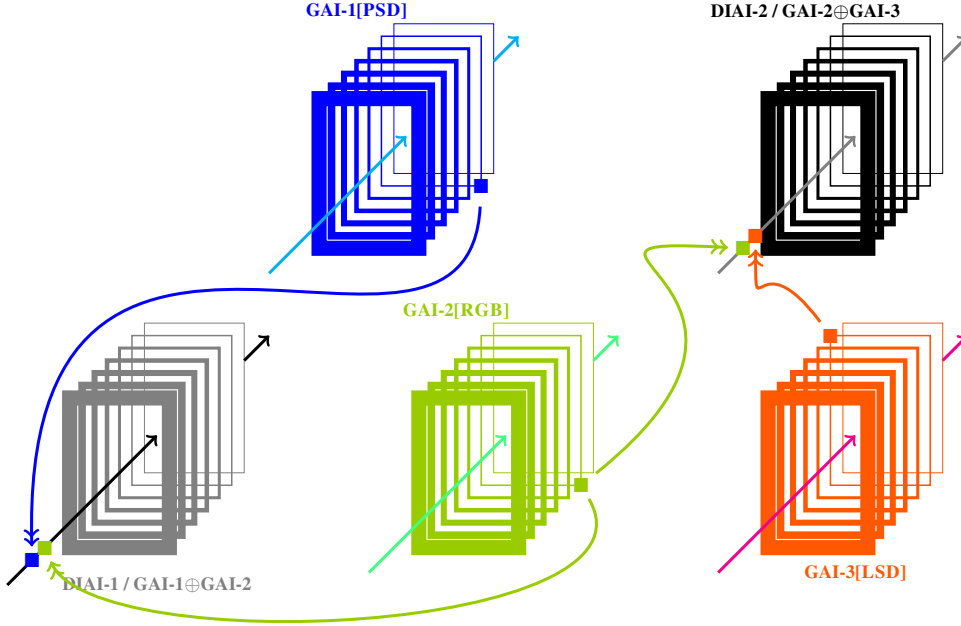
Figure 3: Distributed multimodal system with cooperative GAIs. DGAI = (DGAI-1, DGAI-2), where DGAI-1 = (DIAI-1, GAI-1, GAI-2) and DGAI-2 = (DIAI-2, GAI-2, GAI-3). Any GAI has the form XCEPTION ⊕ GOOGLENET ⊕ DENSENET given in Figure 1.

4.1 ASSESSMENT AND DISTRIBUTED GAIS : EXTREME CLASSIFICATION

The extreme classification paradigm developed here is incremental from the standard classification described in Section 3. We consider first a monomodal multimodel GAI defined as in Section 3.1 and after its training stage, thus with fixed optimal weights for both its IAI and SAIs. In addition, we consider the case of a multinomodal distributed set of GAIs (DGAI) defined as in Section 3.2, any of these GAI operating on multimodel SAIs as in Section 3.1. Figure 3 illustrates the distributed integration sub-network (DIAI for *Distributed IAI*) associated with this system. Any DIAI is assumed to have the same design as an IAI (see layer composition given by Section 2), but a DIAI aggregates assessment processes that directly issue from local multimodel GAI.

The assessment delivered by a local DGAI for considering ν as pertaining to the category k is a probability $S_k^*(\nu)$. Let $\mathbb{T}_k$ denotes the union of the training and validation examples associated with the k-th label in the standard mono-label classification issue described in Section 3. Let $\mathbb{V}_k \subset \mathbb{T}_k$ be the set of examples associated with dominant[12] assessment k that are correctly classified by the GAI in this standard classification issue.

Define the *admissibility bound* for being considered by a GAI as an element of the category k:

$$\mathbf{A}_k^* = \min_\nu \{S_k^*(\nu) : \nu \in \mathbb{V}_k\} \tag{4}$$

Note that $\mathbf{A}_k$ represents the assessment of a <u>correct</u> classification for the training/validation example having the weakest SAI features. The multi-output categorization is hereafter related to multiple decision testing, where ν is questioned independently to belong to any of the different categories indexed by $\mathbb{N}_K$ from the test:

$$d(\nu) = \left(\{S_k^*(\nu) : k \in \mathbb{N}_K\} \underset{0}{\overset{1}{\gtrless}} \{\mathbf{A}_k^* : k \in \mathbb{N}_K\} \right) \tag{5}$$

where $\gtrless$ is the *greater or less than* statistical assignment operator. The comparison is element wise and $d(\nu)$ is a set of carnality K. For instance, $d = (1, 0, 0, 0, \ldots, 0)$ refers to category #1 assignment solely, whereas $d = (1, 1, 0, 0, \ldots, 0)$ refers to both categories #1 and #2 assignment.

---

[12]The k-th assessment is a dominant one if $S_k^*(\nu) \geqslant S_\ell^*(\nu)$ for every ℓ such that $1 \leqslant \ell \leqslant K$.

The metric proposed for performance evaluation is associated with the number of positions at which the set of query's decisions $d(\nu)$ corresponds to the true annotation sequence $d^\dagger(\nu)$, relatively to the true number of the active categories:

$$a(\nu) = \frac{\sum_m d_m(\nu) \otimes d_m^\dagger(\nu)}{\sum_m d_m^\dagger(\nu)}$$

where $\otimes$ is the AND operator. Accuracy is then an average over $a(\nu)$ for the different ν tested.

4.2 GAI AND DGAI EXTREME CLASSIFICATION: THE DTD CASE STUDY

If we consider the 47 texture classes associated with DTD in the standard classification issue (mutually exclusive categories, see Section 3), then the corresponding extreme classification for seeking all possible interacting elementary phenomena on a given texture observation encompasses $\sum_{k=1}^{47} \binom{47}{k}$ categories, thus more than 10^{14} categories: such a multiclass annotation by human experts is highly demanding since one has to decide if a query texture $\mathcal{I}_0$ pertains to K_0 subcategories with $1 \leqslant K_0 \leqslant 47$ and also point out these K_0 categories. At large, this amounts to admit that there are **multiple correct nested decisions** that can allow the description of a given texture. An overcomplete annotation to decide which texture pertains exactly to one of the (a) 47 mutually exclusive elementary characteristics, (b) 1 081 categories corresponding to interactions between 2 elementary characteristics (superimposition on the same observation), (c) 16 215 categories for 3 characteristics in interaction, (d) 178 365 categories for exactly 4 characteristics in interaction, *etc.*, is naturally difficult and the notion of correct answers will remain imprecise in general.

Specifically, for DTD, ground truth annotation is available for 821, 68 and 5 examples having respectively 2, 3 and 4 characteristics. All examples possessing 4 characteristics, as well as 600 among examples possessing 2 characteristics, are transferred to the GAI training dataset $\mathbb{V}$ in the category corresponding to their dominant characteristics. Testing GAI/DGAI concerns the remaining 221 double category and 68 triple category textures. Table 4 presents accuracies of the 3 GAIs and the 2 DGAI presented in Figure 3 thanks to the metrics defined in Section 4.1, respectively for 2 and 3 correct answers.

5 CONCLUSION

The paper has provided assessment integration frameworks for multimodel, multimodal and distributed convolutional neural network frameworks. It has also addressed the difficult and open extreme classification issue in image analysis. The DTD dataset has been used as a case study thanks to its multiple output category annotations. Performance obtained in both standard and extreme classification allows to state that assessment aggregation is a consistent direction for building generalized artificial intelligence.

The paper has considered CNNs as case study of the generalization by learning from assessment processes. It is worth emphasizing that the generalization framework is not limited to CNNs. Indeed, assessment measures can also be requested/provided in an agent/expert-based learning system where one focuses, not on experts' decision, but on experts' interpretation of the facts. The latter are expected to lead to a more efficient aggregation-learning because expert assessments and interpretations of a complex scene is a less biased process than a final expert decision on the same scene.

REFERENCES

D. Chae, J. A. Shin, and S. Kim. Collaborative adversarial autoencoders: An effective collaborative filtering model under the gan framework. *IEEE Access*, 7:37650–37663, 2019. ISSN 2169-3536. doi: 10.1109/ACCESS.2019.2905876.

Z. Chen, B. Subagdja, and A. Tan. End-to-end deep reinforcement learning for multi-agent collaborative exploration. In *2019 IEEE International Conference on Agents (ICA)*, pp. 99–102, Oct 2019. doi: 10.1109/AGENTS.2019.8929192.

F. Chollet. Xception: Deep learning with depthwise separable convolutions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1800–1807, 2017.

M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, , and A. Vedaldi. Describing textures in the wild. In *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015.

S. Gehrmann, H. Strobelt, R. Kruger, H. Pfister, and A. M. Rush. Visual interaction with deep learning models through collaborative semantic inference. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):884–894, Jan 2020. ISSN 1941-0506. doi: 10.1109/TVCG.2019. 2934595.

G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, 2017.

Yanping Huang, Youlong Cheng, Ankur Bapna, Orhan Firat, Dehao Chen, Mia Chen, HyoukJoong Lee, Jiquan Ngiam, Quoc V Le, Yonghui Wu, and zhifeng Chen. Gpipe: Efficient training of giant neural networks using pipeline parallelism. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems 32*, pp. 103–112. Curran Associates, Inc., 2019.

Jakub Konecny, Brendan McMahan, and Daniel Ramage. Federated optimization:distributed optimization beyond the datacenter, 2015.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (eds.), *Advances in Neural Information Processing Systems 25*, pp. 1097–1105. Curran Associates, Inc., 2012.

Z. Li, J. Tang, and T. Mei. Deep collaborative embedding for social image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(9):2070–2083, Sep. 2019. ISSN 1939-3539. doi: 10.1109/TPAMI.2018.2852750.

D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 2, pp. 1150–1157 vol.2, 1999.

Y. Pan, H. Jiang, H. Yang, and J. Zhang. A novel method for improving the training efficiency of deep multi-agent reinforcement learning. *IEEE Access*, 7:137992–137999, 2019. ISSN 2169-3536. doi: 10.1109/ACCESS.2019.2942635.

Joshua Peeples, Weihuang Xu, and Alina Zare. Histogram layers for texture analysis, 2020.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vision*, 115(3):211–252, December 2015. ISSN 0920-5691. doi: 10.1007/s11263-015-0816-y.

S. Savazzi, M. Nicoli, and V. Rampa. Federated learning with cooperating devices: A consensus approach for massive iot networks. *IEEE Internet of Things Journal*, 7(5):4641–4654, 2020.

H. Shen, M. Jiang, J. Li, Q. Yuan, Y. Wei, and L. Zhang. Spatial-spectral fusion by combining deep learning and variational model. *IEEE Transactions on Geoscience and Remote Sensing*, 57(8): 6169–6181, Aug 2019. ISSN 1558-0644. doi: 10.1109/TGRS.2019.2904659.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

Yang Song, Fan Zhang, Qing Li, Heng Huang, Lauren J. O'Donnell, and Weidong Cai. Locally-transferred fisher vectors for texture classification. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

C. Szegedy, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, June 2015. doi: 10.1109/CVPR.2015.7298594.

Rohith Dwarakanath Vallam, Ramasuri Narayanam, Srikanth G. Tamilselvam, Nicholas Mattei, Sudhanshu S. Singh, Shweta Garg, and Gyana R. Parija. Deepaggregation: A new approach for aggregating incomplete ranked lists using multi-layer graph embedding. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS'19, pp. 2235–2237, Richland, SC, 2019. International Foundation for Autonomous Agents and Multiagent Systems. ISBN 9781450363099.

David H. Wolpert. Stacked generalization. *Neural Networks*, 5(2):241 – 259, 1992. ISSN 0893-6080. doi: https://doi.org/10.1016/S0893-6080(05)80023-1. URL http://www.sciencedirect.com/science/article/pii/S0893608005800231.

Jia Xue, Hang Zhang, and Kristin Dana. Deep texture manifold for ground terrain recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

H. Zhang, J. Xue, and K. Dana. Deep ten: Texture encoding network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2896–2905, Los Alamitos, CA, USA, jul 2017. IEEE Computer Society. doi: 10.1109/CVPR.2017.309. URL https://doi.ieeecomputersociety.org/10.1109/CVPR.2017.309.

L. Zhao, Q. Wang, Q. Zou, Y. Zhang, and Y. Chen. Privacy-preserving collaborative deep learning with unreliable participants. *IEEE Transactions on Information Forensics and Security*, 15:1486–1500, 2020. ISSN 1556-6021. doi: 10.1109/TIFS.2019.2939713.