



- (51) **International Patent Classification:**
G06N 3/04 (2006.01) *G06N 3/08* (2006.01)
- (21) **International Application Number:**
PCT/EP2020/051127
- (22) **International Filing Date:**
17 January 2020 (17.01.2020)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
19154667.0 31 January 2019 (31.01.2019) EP
- (71) **Applicant:** **SIEMENS AKTIENGESELLSCHAFT** [DE/DE]; Werner-von-Siemens-Straße 1, 80333 München (DE).
- (72) **Inventors:** **KROMPASS, Denis**; Balanstr. 226, 81549 München (DE). **KÖPKEN, Hans-Georg**; Hausäckerweg 35, 91056 Erlangen (DE). **TOMOV, Tomislav**; Bismarckstraße 26, 80803 München (DE).

KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

- (84) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

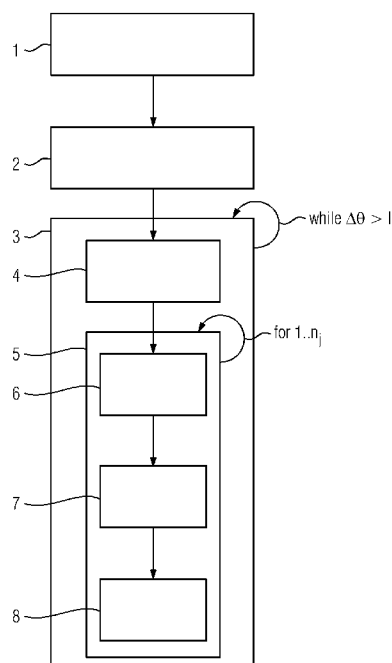
Published:

— with international search report (Art. 21(3))

- (81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP,

(54) **Title:** A METHOD FOR BUILDING A DEEP LATENT FEATURE EXTRACTOR FOR INDUSTRIAL SENSOR DATA

FIG 1



(57) **Abstract:** The present invention relates to a method of and a system for building a latent feature extractor as well as a neural network (NN) comprising latent feature extractor built by said method and/or with said system. The method comprises providing non-uniform training data for a multitude of tasks and optimising parameters (θ) of a neural network (NN) of the latent feature extractor based on the multitude of tasks.

WO 2020/156835 A1

Description

A method for building a deep latent feature extractor for industrial sensor data

5

The present invention relates to a method of and a system for building a latent feature extractor as well as a neural network (NN) comprising latent feature extractor built by said method and/or with said system.

10

Today, labelled data is one of the if not even the most important ingredient for enabling machine learning based artificial intelligence (AI) solutions to solve new and challenging automation tasks. Labelled data describes datasets that have been specifically annotated by domain knowledge with a specific problem setting in mind such that the parameters of a machine learning model can be optimised by examples.

15

For example in an image classification problem a method should automatically detect whether a person is present in an image recorded by a camera. The current state of the art in solving this problem with methods from the domain of machine learning requires to build a labelled dataset that contains images and corresponding labels that indicate whether a person is present or not in each image. Given this labelled dataset the machine learning algorithm can learn a function that maps input pixels to a classification by repeatedly viewing examples from the labelled dataset. Even though building such a dataset sounds quite straight forward it can be incredibly labour and cost intensive if not even infeasible since, depending on the complexity of the target problem, thousands to millions of these labelled data instances are required.

25

30

35

In the industrial domain, labelling a sufficient amount of data is hard if not even infeasible. The reason for this lies in the fact that in contrast to the common domain, where ba-

sically anybody can label data (e.g. determine whether an image shows a cat or a dog), in the industrial domain only a very small amount of highly specialized domain experts are able to label data (e.g. annotate the condition of a gas turbine). Unfortunately, the most advanced and promising methods from the field of AI, especially Deep Learning, require huge amounts of labelled data for learning complex relationships between inputs and the target task to solve.

10 There are basically two ways to solve this problem. The first is to invest large amounts of money in training a workforce of highly specialized data labellers that build the required labelled datasets for every single target problem for which an AI solution based on machine learning is pursued. Obviously this approach is not very attractive since it is very labour intensive and expensive and also requires access to a large pool of (external) data which is not feasible in many cases. The second approach relies on classic domain specific feature engineering based on established signal processing methods like Fast Fourier Transform (FFT), Wavelets, moving statistics to name a few. If these features significantly facilitate the learning task, also the amount of required labelled data will be reduced. Applying this approach can be seen as manual feature extraction where the input sensor signals are represented (embedded) in a new interpretable space. However, such feature engineering is often not possible as not always a suitable signal processing method exists/is applicable.

30 Consequently, there is a need for a way to reduce the amount of labelled data necessary for training neural networks for AI solutions in the industrial field where non-uniform and multidimensional data (e.g. different sensor data) is the input.

35

Therefore, the present invention provides a method for building a latent feature extractor according to independent claim

1 and a system of building a latent feature extractor as well as a neural network according to the further independent claims. Further refinements of the present invention are subject of the dependent claims.

5

According to a first aspect of the present invention a method of building a latent feature extractor comprises the steps of:

10 Providing non-uniform training data comprising k_T training examples for each single task T of a multitude of T_i tasks. Optimising parameters θ of a neural network (NN) of the latent feature extractor based on the multitude of T_i tasks. Thereto, the following steps are executed iteratively until the parameters θ have converged:

15 Randomly selecting a single tasks T from the multitude of T_i tasks for optimising the parameters θ .

20 Sampling the current randomly selected task T for n_j iterations with the respective k_T training examples of the non-uniform training data for optimising the parameters θ based on the current randomly selected task T .

In each iteration n of the n_j iterations the following steps are executed:

25 Extracting latent features from a training signal of the current training example k of the respective training examples k_T with the latent feature extractor.

Deriving an output signal o_i for the current randomly selected task T from the extracted latent features with a respective task module of the current randomly selected task T .

30 According to a second aspect of the present invention a system for building a latent feature extractor arranged and configured for executing the method as described above comprises the latent feature extractor and task modules. The latent feature extractor is utilised for extracting the latent features. The latent feature extractor is shared for each task T
35 of the multitude of T_i tasks. The task modules for each single task T of the multitude of T_i tasks are located down-

stream the latent feature extractor and utilised for deriving the output signal o_i for the respective task T.

5 According to a third aspect of the present invention a neural network (NN), preferably a deep NN and in particular an (deep) industrial NN, comprises a latent feature extractor, a combiner module in case multiple input signals are provided to the neural network and a specific task module. The latent feature extractor is built by the method and/or with the system as described above. The combiner module is located downstream the latent feature extractor. The specific task module is located downstream the latent feature extractor and (if present) the combiner module, respectively.

15 In order to reduce the amount of labelled training data the specific target task is not learned from scratch but model parameters θ of the latent feature extractor are pre-initialised to a reasonable starting point. This can be seen as equipping the (deep) (industrial) NN (= model) with some amount of prior domain knowledge beforehand which lowers the amount of required labelled examples to solve the target task tremendously. As mentioned above, in the industrial domain labelled data is a major bottleneck for implementing AI solutions based on machine learning. Having pre-trained models comprising general latent feature extractors (pre-trained (deep) NN) that automatically extract latent features from general industrial sensor data such as temperatures, torques, current and voltage profiles do reduce the amount of labelled training data needed to train an industrial neuronal network for a specific task.

35 In the following first various (deep) (industrial) neural network architecture concepts that are of interest for learning a latent feature extractor for industrial sensor data are described.

In the simplest scenario the input to the (deep) NN/model consists of a single signal. For example, the input to the model consists of a single sequence of fixed length like 1s [second] of torque measurements of the spindle in a milling machine measured with a 500hz [Hertz] frequency, which results in 500 values per input. Given the context of the data the data has to be properly normalized such that the values are not too large or too low. Normally, min-max normalization or centring the data and scaling to unit variance or just scaling the values by some other constant (e.g. observed standard deviation within the complete data) can be utilised. The normalized signal will be fed into the subsequent (deep) NN. The (deep) NN consumes the normalized input signal and produces outputs optimised for a specific target task. For example, for an anomaly detection scenario the model might output the probability that the input signal contains an anomaly or in an activity recognition task class probabilities for various activities that might have generated the input signal. Generally, the (deep) NN architecture can be subdivided into two modules (which are optimised end-to-end not independently later on), the (general) latent feature extractor ((deep) "sub-NN") and a specific task module ((deep) "sub-NN"). The latent feature extractor consumes the normalized sensor signal data and outputs a latent feature/embedding of that input signal. The latent feature extractor represents a function that encodes the input sensor signal(s) into a latent feature space where each feature represents meaningful information extracted from the input data that is useful for the downstream specific task module. Here, the latent feature extraction is performed by a (deep) NN that consists of multiple 1D convolutional neural network layers that are stacked on top of each other. With a deeper architecture the model can extract locally detected latent features and further combine them in a hierarchical fashion in the deeper layers. The (deep) NN architecture might further contain some other components like batch normalization layers or skip connections to stabilize the training of the

components. The specific task module consumes the latent features produced by the latent feature extractor and produces outputs required for the specified target task. The specific task module is a (deep) neural network that combines the latent features and produces values meaningful for the specific target problem (e.g. an anomaly probability, class probabilities or any other value).

In a more complex scenario the input to the (deep) (industrial) NN/model consists of multiple input signals. Each of the input signals to the approach consists of a single sequence of fixed length, for example 1s of torque measurements of the spindle and the x, y, z coordinates of the spindle in a milling machine measured with a 500hz frequency, which results in 500 values per input signal. Like described above, the input data needs to be normalised. In contrast to the single input scenario each of the types of the input signals have to be normalised individually if they were not measured by a related sensor. For example a torque measurement needs to be normalised differently than a coordinate measurement. The resulting normalized signal will be fed into the (deep) (industrial) NN. The (deep) NN consumes the normalized input signal and produces outputs optimised for the specific target task (e.g. anomaly detection scenario or activity recognition task). Generally the (deep) NN architecture can be subdivided into three modules (which are optimised end-to-end not independently later on). The latent feature extractor ((deep) "sub"-NN) consumes each of the normalized sensor signals individually and outputs latent features/embeddings for each of the input signals. The latent feature extractor represents a function that encodes each single input sensor signal into a latent feature space where each features represent meaningful information extracted from the input data that is useful for the downstream modules. As described before, the latent feature extraction is performed by a (deep) NN that consists of multiple 1D convolutional neural network layers that are stacked on top of each other. With a deeper architecture the

model can extract locally detected latent features and further combine them in a hierarchical fashion in the deeper layers. The deep neural network architecture might further contain some components like batch normalization layers or skip connections to stabilize the training of the (deep) NN. 5 The combiner module combines the latent features produced by the latent feature extractor for every input signal in a reasonable way. In the simplest case, they get simply concatenated, added or multiplied. In more complex scenarios, the combiner is also a (deep) "sub-NN" that learns meaningful 10 combinations of the various latent features extracted from each input signal. The output of the combiner module is fed into the specific task module. The specific task module consumes the combined latent features produced by the combiner module and produces outputs required for the specific target 15 task. The specific task model is a (deep) "sub-NN" that combines latent features and produces values meaningful for the target problem (e.g. an anomaly probability, class probabilities or any other value). The (deep) (industrial) NN is hence 20 a combination of the (deep) "sub-NN" of the latent feature extractor and the combiner module (if present) and the (deep) "sub-NN" of the specific task module.

The general latent feature extractor is built or pre-trained 25 in a multi-task training setting. In order to train the latent feature extractor on a general level multiple different tasks T_i are optimised in the pre-training while the latent feature extractor is shared between all the different tasks T_i . The combiner modules and of course the task modules are 30 not shared. This means that during pre-training of the general latent feature extractor all tasks T_i optimise the same (deep) NN of the general latent feature extractor, which is responsible for extracting the useful features from the various non-uniform input industrial sensor signals.

35

Based on the non-uniform training data the (general) latent feature extractor is built or rather pre-trained. For each of

the T_i different tasks, which are used for building/pre-training the latent feature extractor, the non-uniform training data comprises a predefined number of k_T training examples. The T_i tasks may include several tasks with multiple
5 input signals and some tasks with only one input signal. The different tasks may cover different problems in the industrial field, for example, error/fault detection, state detection or prediction, security issues etc. For example, a task with multiple input signals could be detection of anomalies in the
10 operation of a gas turbine (probability thereof) based on different sensor signals having different dimensions including rotational speed, torque, current load, vibrations of the drive shaft, state of stress of the turbine blades and the like. An exemplary task with only one single sensor input
15 signal could be the detection of faulty electrical components based on a current profile at a power source. The more different the tasks are (underlying problem and used input signals) the more general the resulting latent feature extractor will be.

20

For each of the different tasks used for training the (general) latent feature extractor corresponding labelled training data has to be provided. Due to the different tasks the provided training data is non-uniform (different number of
25 input signals as well as different values and dimensions). The non-uniform training data used for building/pre-training the (general) latent feature extractor comprises a predefined number of training examples for each of the different tasks.

30 The (general) latent feature extractor is based on its (deep) NN. The (deep) NN comprises the internal parameters θ . During pre-training of the (deep) NN of the latent feature extractor the parameters θ are optimised.

35 The optimising is done by first randomly selecting one single task T of the multitude of T_i different tasks. The selection of the tasks is random and based on an equal probability $p(T)$

for all tasks such that during the process of optimising the parameters θ each of the tasks T_i is selected nearly equally often.

5 Subsequently, the randomly selected task T is sampled based on the corresponding k_T training examples of the non-uniform training data for n_j iterations. The number of iterations n_j is predefined. Sampling means feeding the current training signals of the current training example to the latent feature
10 latent feature extractor and based on latent features generated by the latent feature extractor deriving the current output o_i by means of the respective task module for the current randomly selected task T . Based on the generated output o_i of the current task T the parameters θ are then adjusted such that the
15 next output o_i will as closely as possible approximate to a given result of the current training example. In other words, (latent) features are extracted from the training signal(s) of the current training example k by the latent feature extractor based on the current parameters θ . Then the extracted
20 latent features are processed by the respective task module for the current task T and the output o_i is generated. The output o_i is compared to the given result of the current training example. Based on the difference between the current generated output o_i and the result the parameters θ are ad-
25 justed such that the next output o_i will more closely approximate to the given result of the next training example. The parameters θ may be adjusted with a stochastic gradient descent based optimiser. In each iteration of the n_j iterations the current randomly selected task T is sampled with all k
30 training examples.

After one randomly selected task T has been sampled for the n_j iterations the next task T of the multitude of tasks T_i is randomly selected and subsequently sampled for n_j iterations.
35 The random selection and subsequent sampling of tasks is continued until the parameters θ of the (general) latent feature extractor converge. Converging means that after one task T

has been sampled in the n_j iterations the parameters do not have changed by more than a predefined threshold.

5 During optimising the parameters θ of the latent feature extractor, parameters of each of the T_i tasks are also optimised. However, these parameters are not part of the latent feature extractor and can thus be disregarded.

10 For tasks having more than one input signal, after the (general) latent feature extractor has extracted the (latent) features from the current training example k for the current randomly selected task T a combiner module for the respective task T (as described above) may combine the extracted features (e.g. by concatenating them or in a meaningful way) before the extracted features are provided to the respective
15 task module of the current task T .

An exemplary pre-training algorithm is outlined in pseudo code below:

20

1. Pre-process the data (normalizing)
2. Randomly initialize the parameters θ
3. While θ not converged
4. Sample a random task T of T_i ($\sim p(T)$)
- 25 5. For n_j iterations:
 6. Sample k_T training examples for Task T
 7. Use the task module (model) for task T to evaluate and update the parameters θ (with a stochastic gradient descent based optimiser)
- 30 8. End for
9. End while

35 With the described pre-training scheme a general or common feature extractor is built that produces useful latent features or signal embeddings that are valuable to all downstream tasks. Thus, the latent feature extractor is of general nature and also valuable to other tasks not used during

the pre-training. Note that the tasks T_i are not related, meaning the first task could be classifying activities from sensor signals produced by sensors attached to the human body while in movement (smart watch/fitness tracker), while the
5 next task consists of an anomaly classification problem on daily gas turbine sensor data or a tool break detector in a milling machine based on machining sensor data. During pre-training the latent feature extractor is embedded in various architectures (T_i different tasks) which try to solve a di-
10 verse set of tasks independently, yet they have to use and adapt the same latent feature extractor. This way the latent feature extractor must produce meaningful and general latent features that are beneficial for all T_i tasks. Thereby, the latent feature extractor is directly optimised to serve the
15 desired general purpose.

A (deep) (industrial) NN according to the present invention (as described above) comprises the (general) latent feature extractor ((deep) "sub-NN") pre-trained as described above.
20 Said NN further comprises the specific task module for (model of) the specific task the industrial NN is intended for ((deep) "sub-NN") and, in case more than one input signal is processed by the NN, a combiner module between the latent feature extractor and the specific task module. The NN has to
25 be trained for the specific task before it can be utilized. Due to the pre-trained (general) latent feature extractor significantly less training examples need to be provided for training of the industrial NN. During the training of the NN, the parameters θ of the latent feature extractor ((deep)
30 "sub-NN") are not changed anymore. However, in order to further fine tune the whole NN the parameters θ of the latent feature extractor ((deep) "sub-NN") may be further adjusted.

Having access to a general latent feature extractor for industrial sensor data is a very powerful asset as. Given such
35 feature extractor one can integrate it in an architecture of an industrial NN as feature extraction component. For new

(industrial) tasks, the developer has to only train the specific task module or the combiner module and task module, respectively, which are of significantly lower complexity than the latent feature extractor. Of course, the developer has the freedom to further tune the latent feature extractor if needed. For this reason, only a relatively low amount of labelled training data is required to optimise the complete (deep) (industrial) NN/model for the specific (industrial) task. A decrease in the required amount of labelled training data by a factor of up to 1000 can be achieved. Further, such pre-trained (deep) (industrial) NNs/machine learning models enable a larger pool of developers to transfer quickly to new and complex industrial problems in a shorter period of time without the requirement to build and maintain large labelled datasets or to set up an infrastructure to train and evaluate large and complex models on these datasets. It greatly reduces the need for domain specific feature engineering since the feature extraction is performed automatically. In addition, the time for training will also be significantly reduced, resulting in faster development iterations and a greater chance of success by reducing both, labour cost and time at the same time. Hence, the (general) latent feature extractor described above has the potential to dramatically increase the speed and scale of AI based solution development based on non-uniform input signals like industrial sensor data together with a decrease in labour and costs.

According to a refinement of the present invention the non-uniform training data are sensor data from different sensors.

The sensor data comprises signals that are based on different physical variables like temperature, humidity, pressure, tension, strain, force, acceleration, velocity, distance, voltage, current, flow etc. These signals have different dimensions and value ranges. Thus, the resulting training data is very non-uniform.

Consequently, the resulting (general) latent feature extractor provides meaningful latent features for a wide variety of industrial applications or problems.

- 5 According to a refinement of the present invention T_i is between 2 and 1000 and preferably between 2 and 100 and/or wherein n_j is between 1 and 10000 and/or wherein each k_T is equal to one of 32, 64 and 128.
- 10 A robust general latent feature extractor may be generated by pre-training with about $T_i = 100$ different tasks, wherein each randomly selected task T is sampled in about $n_j = 10000$ iterations with $k_T \in [32, 64, 128]$ corresponding training examples.
- 15 According to a refinement of the present invention the method further comprises the step of:
Normalising the non-uniform training data before the step of optimising.
- 20 According to a refinement of the present invention the (deep) (industrial) neural network further comprises a normaliser module upstream the latent feature extractor.
- 25 Before the training signal(s) of the non-uniform training data are provided to the latent feature extractor, either during pre-training of the latent feature extractor or in the (deep) (latent) NN, the signal(s) are normalised by a normaliser module. The normalising can be done by a min-max normalisation or by centring the data and scaling to unit variance
30 or by just scaling the values by some other constant (e.g. observed standard deviation within the complete data). Thus, the latent feature extractor is provided with values that are all in the same range.

According to a refinement of the present invention the k_T training examples for each task T comprise one or multiple training signals from one or multiple sensors.

5 By pre-training the latent feature extractor with non-uniform training data comprising training examples, some of which having one training signal and several of which having more than one training signal, for the T_i tasks generality of the resulting latent feature extractor can be ensured. Here some
10 of the T_i tasks correspondingly consume one input signal from one sensor and several of the T_i tasks consume multiple input signals from multiple sensors.

According to a further refinement of the present invention in
15 each iteration n of the n_j iterations where the k_T training examples for the current randomly selected task T comprise multiple training signals in the step of extracting features are extracted for every training signal of the multiple training signals and wherein in each iteration n of the n_j
20 iterations where the k_T training examples for the current randomly selected task T comprise multiple training signals the following step is additionally executed:
Combining the extracted latent features of each training signal of the multiple training signals into combined extracted
25 features with a respective combiner module of the current randomly selected task T before the step of deriving.

According to a further refinement of the present invention the system further comprises downstream the latent feature
30 extractor and upstream the respective task module for each task T where the respective k_T training examples comprise multiple training signals a combiner module for combining the extracted latent features of each training signal of the multiple training signals.

35

For each task T having more than one input signal, the latent features extracted from the corresponding multiple training

signals by the latent feature extractor have to be combined by the combiner module before they are provided to the respective task module in order to produce an meaningful output O_i .

5

According to a refinement of the present invention the latent feature extractor comprises between 10 and 30 layers and each layer of the latent feature extractor comprises between 10 and 1000 neurons.

10

According to a refinement of the present invention each task module comprises between 1 and 2 layers and each layer of the task modules comprises between 10 and 1000 neurons.

15

According to a further refinement of the present invention each combiner module comprises between 1 and 2 layers and each layer of the combiner modules comprises between 10 and 1000 neurons.

20

According to a refinement of the present invention the latent feature extractor of the (deep) (industrial) NN comprises between 10 and 30 layers comprising between 10 and 1000 neurons and alternatively or additionally the specific task module of the NN comprises between 1 and 2 layers comprising between 10 and 1000 neurons and alternatively or additionally the combiner module comprises between 1 and 2 layers comprising between 10 and 1000 neurons.

25

Due to the reduced complexity of the "sub-NN" of the specific task module and of the combiner module, respectively, compared to the pre-trained latent feature extractor (that does not have to be trained anymore) the amount of labelled training data as well as the time needed for training the (deep) (industrial) NN can be significantly reduced.

35

The present invention and its technical field are subsequently explained in further detail by exemplary embodiments shown

in the drawings. The exemplary embodiments only conduce better understanding of the present invention and in no case are to be construed as limiting for the scope of the present invention. Particularly, it is possible to extract aspects of the subject-matter described in the figures and to combine it with other components and findings of the present description or figures, if not explicitly described differently. Equal reference signs refer to the same objects, such that explanations from other figures may be supplementally used.

10

Fig. 1 schematically shows a flow chart of a method of building a latent feature extractor.

15

Fig. 2 schematically shows a system for building a latent feature extractor.

20

Fig. 3-4 schematically shows two embodiments of a deep industrial network comprising a latent feature extractor.

25

In Fig. 1 a flow chart of a method of building a latent feature extractor is schematically depicted. In a first step of providing 1 non-uniform training data is provided. The non-uniform training data comprises $k_T \in [32, 64, 128]$ training examples for each single task T of a multitude of $T_i = 100$ tasks. The tasks comprise industrial applications or problems like deriving a probability of an anomaly in the operation of a gas turbine. Each training example comprises one or more sensor signals (e.g. rotational speed, temperature, current load etc.) and is labelled with a result for the respective task T (e.g. a probability of an anomaly). In an optional second step of normalising 2 the non-uniform training data is normalised. The non-uniform training data is normalised by a min-max normalisation or by centring the data and scaling to unit variance or by just scaling the values by some other constant (e.g. observed standard deviation within the complete data). In a third iterative step of optimising 3 param-

35

eters θ of a neural network (NN) of the latent feature extractor are optimised based on the multitude of T_i tasks. Thereto a fourth step of randomly selecting 4 and a fifth step of sampling 5 are executed iteratively until the parameters θ have converged ($\Delta\theta > \epsilon$). In the fourth step of randomly selecting 4 a single tasks T from the multitude of T_i tasks is randomly selected for optimising the parameters θ . The selecting of the tasks T is based on an equal probability $p(T)$ for all tasks such that during the process of optimising the parameters θ each of the tasks T_i is selected nearly equally often. In the fifth step of sampling 5 the current randomly selected task T is sampled for n_j iterations. Thereby, the current randomly selected task T is sampled with the respective k_T training examples of the non-uniform training data. Thereby, the parameters θ are optimised based on the current randomly selected task T . In each iteration n of the n_j iterations a sixth step of extracting 6, an optional seventh step of combining 7 and an eighth step of deriving 8 are executed. In the sixth step of extracting 6 features from a training signal of the current training example k of the respective training examples k_T are extracted with the latent feature extractor. In the optional seventh step of combining 7 the extracted latent features of each training signal of the multiple training signals are combined into combined extracted features with a respective combiner module of the current randomly selected task T . The combiner module is a (deep) "sub-NN" that learns meaningful combinations of the various latent features extracted from each training signal. In the eighth step of deriving 8 an output signal o_i for the current randomly selected task T is derived from the (combined) extracted features with a respective task module of the current randomly selected task T . Based on the output o_i the parameters θ are adjusted such that the next output approximates the respective result of the next training example more closely.

In Fig. 2 a system 10 for building a latent feature extractor is schematically depicted. The system 10 is arranged and configured for executing the method of Fig. 1. The system 10 comprises a multitude of (deep) NNs one for each of $T_i = 100$ tasks ($T = 1..100$). The NNs of the tasks T share one latent feature extractor 11 ("sub-NN") having 30 layers with 1000 neurons each. Each NN of the tasks T_i comprises a respective task module 12.i ("sub-NN") for its task T downstream the latent feature extractor 11, each task module 12.i having 2 layers with 100 neurons each. Further, each NN of the T_i tasks may comprise an optional combiner module 13.i ("sub-NN") between the latent feature extractor 11 and the respective task module 12.i, each combiner module having 2 layers with 100 neurons each. Further, the system may comprise optional normaliser modules 14.i for the NNs of the T_i tasks upstream the latent feature extractor 11.

The training signals or rather input signals $si.m_i$ for each of the tasks T_i are optionally normalised in the corresponding normaliser module 14.i. The latent feature extractor 11 extracts the latent features lf from the (normalised) input signals $si.m_i$ ($s_{ni}.m_i$). The latent feature extractor 11 is shared for each task T of the multitude of T_i tasks. The optional combiner modules 13.i combine the extracted latent features lf for tasks T that consume more than one (normalised) input signal $si.m_i$ ($s_{ni}.m_i$) (here exemplarily the task $T = 1$ and the task $T = T_i = 100$) into combined latent features. The task modules 12.i downstream the latent feature extractor 11 derive the corresponding output signal o_i for the respective task T from the corresponding (combined) latent features. based on the output o_i the parameters θ are adjusted and iteratively optimised.

In Figs. 3 and 4 two exemplary embodiments of a deep industrial NN are schematically depicted.

In Fig. 3 a deep industrial NN 20 for a task consuming only one input signal is schematically depicted. The NN 20 comprises a normaliser module 14, a latent feature extractor 11 with 30 layers having 100 neurons each downstream the normaliser module 14 and a specific task module 22 with two layers having 100 neurons each downstream the latent feature extractor 11.

The latent feature extractor 11 has been pre-trained with the method of Fig. 1 and/or the system of Fig. 2. Consequently, the latent feature extractor 11 is able to output latent features lf for various different input signals s_i . Here, one input signal s (e.g. a current profile of a power source) is provided to the normaliser module 14. The normaliser module 14 converts the input signal s into a normalised input signal s_n . The latent feature extractor 11 extracts a latent feature lf from the normalised input signal s_n . The specific task module derives an output o (e.g. presence of a faulty electrical component) from the latent feature lf .

As the latent feature extractor 11 is pre-trained and much more complex (30 layers with 1000 neurons each) than the specific task module 22 (2 layers with 100 neurons each) the training of the NN 20 for the specific task (e.g. deriving if a faulty electrical component is present based on the current profile of a power source) is much faster and requires significantly less labelled training data than a conventional NN without pre-trained feature extractor.

In Fig. 4 a deep industrial NN 30 for a task consuming more than one one input signal is schematically depicted. In the following only differences to the NN 20 of Fig. 3 are described. The NN 30 comprises a specific task module 32 with two layers having 100 neurons each downstream the latent feature extractor 11 and additionally a combiner module 33 with two layers having 100 neurons each between the latent feature extractor 11 and the specific task module 32.

Here, more than one input signal $s_{.1..s.m_i}$ ($m_i > 1$) (e.g. a rotational speed, a tension state of turbine blades and a current load of a gas turbine) is provided to the normaliser module 14. The normaliser module 14 converts the input signals $s_{.1..s.m_i}$ into normalised input signals $s_{n.1..s_{n.m_i}}$. The latent feature extractor 11 extracts latent features lf from the normalised input signals $s_{n.1..s_{n.m_i}}$. The combiner module 33 combines the latent features lf to a combined latent feature. The specific task module 32 derives an output o (e.g. probability of an abnormal operation state) from the combined latent feature provided by the combiner module 33.

As the latent feature extractor 11 is pre-trained and much more complex (30 layers with 1000 neurons each) than the specific task module 32 and the combiner module 33 (2 layers with 100 neurons each) the training of the NN 30 for the specific task (e.g. deriving if an abnormal operation state is present in a gas turbine) is much faster and requires significantly less labelled training data than a conventional NN without pre-trained feature extractor.

Although specific embodiments have been illustrated and described herein, it will be appreciated by those of ordinary skill in the art that a variety of alternate and/or equivalent implementations exist. It should be appreciated that the exemplary embodiment or exemplary embodiments are only examples, and are not intended to limit the scope, applicability, or configuration in any way. Rather, the foregoing summary and detailed description will provide those skilled in the art with a convenient road map for implementing at least one exemplary embodiment, it being understood that various changes may be made in the function and arrangement of elements described in an exemplary embodiment without departing from the scope as set forth in the appended claims and their legal equivalents. Generally, this application is intended to cover any adaptations or variations of the specific embodiments discussed herein.

In the foregoing detailed description, various features are grouped together in one or more examples for the purpose of streamlining the disclosure. It is understood that the above description is intended to be illustrative, and not restrictive. It is intended to cover all alternatives, modifications and equivalents as may be included within the scope of the invention. Many other examples will be apparent to one skilled in the art upon reviewing the above specification.

Specific nomenclature used in the foregoing specification is used to provide a thorough understanding of the invention. However, it will be apparent to one skilled in the art in light of the specification provided herein that the specific details are not required in order to practice the invention. Thus, the foregoing descriptions of specific embodiments of the present invention are presented for purposes of illustration and description. They are not intended to be exhaustive or to limit the invention to the precise forms disclosed; obviously many modifications and variations are possible in view of the above teachings. The embodiments were chosen and described in order to best explain the principles of the invention and its practical applications, to thereby enable others skilled in the art to best utilize the invention and various embodiments with various modifications as are suited to the particular use contemplated. Throughout the specification, the terms "including" and "in which" are used as the plain-English equivalents of the respective terms "comprising" and "wherein," respectively. Moreover, the terms "first," "second," and "third," etc., are used merely as labels, and are not intended to impose numerical requirements on or to establish a certain ranking of importance of their objects. In the context of the present description and claims the conjunction "or" is to be understood as including ("and/or") and not exclusive ("either ... or").

Patent Claims

1. Method of building a latent feature extractor, comprising the steps of:
 - 5 - providing non-uniform training data comprising k_T training examples for each single task T of a multitude of T_i tasks; and
 - optimising parameters θ of a neural network of the latent feature extractor based on the multitude of T_i tasks,
10 wherein the following steps are executed iteratively until the parameters θ have converged:
 - randomly selecting a single tasks T from the mul-
15 titude of T_i tasks for optimising the parameters θ ;
 - sampling the current randomly selected task T for n_j iterations with the respective k_T training ex-
20 amples of the non-uniform training data for opti-
mising the parameters θ based on the current ran-
domly selected task T ,
wherein in each iteration n of the n_j iterations
the following steps are executed:
 - extracting latent features from a training
25 signal of the current training example k of
the respective training examples k_T with the
latent feature extractor; and
 - deriving an output signal o_i for the current
30 randomly selected task T from the extracted
latent features with a respective task module
of the current randomly selected task T .
2. Method according to claim 1, wherein the non-uniform training data are sensor data from different sensors.
- 35 3. Method according to claim 1 or 2, wherein T_i is between 2 and 1000 and preferably between 2 and 100 and/or wherein

n_j is between 1 to 10000 and/or wherein each k_T is equal to one of 32, 64 and 128.

4. Method according to any preceding claim, further comprising the step of:
 - normalising the non-uniform training data before the step of optimising.
5. Method according to any preceding claim, wherein the k_T training examples for each task T comprise one or multiple training signals from one or multiple sensors.
6. Method according to claim 5, wherein in each iteration n of the n_j iterations where the k_T training examples for the current randomly selected task T comprise multiple training signals in the step of extracting features are extracted for every training signal of the multiple training signals and wherein in each iteration n of the n_j iterations where the k_T training examples for the current randomly selected task T comprise multiple training signals the following step is additionally executed:
 - combining the extracted latent features of each training signal of the multiple training signals into combined extracted features with a respective combiner module of the current randomly selected task T before the step of deriving.
7. System for building a latent feature extractor arranged and configured for executing the method according to claim 1, comprising:
 - the latent feature extractor for extracting the latent features, wherein the latent feature extractor is shared for each task T of the multitude of T_i tasks; and
 - task modules for each single task T of the multitude of T_i tasks downstream the latent feature extractor

for deriving the output signal o_i for the respective task T.

8. System according to claim 7, wherein the system is arranged and configured for executing the method according to any of claims 2 to 5.
9. System according to claim 7 or 8, wherein the latent feature extractor comprises between 10 and 30 layers and wherein each layer comprises between 10 and 1000 neurons.
10. System according to any of claims 7 to 9, wherein each task module comprises between 1 and 2 layers and wherein each layer comprises between 10 and 1000 neurons.
11. System according to any of claims 7 to 10 arranged and configured for executing the method according to claim 6, further comprising downstream the latent feature extractor and upstream the respective task module for each task T where the respective k_T training examples comprise multiple training signals a combiner module for combining the extracted latent features of each training signal of the multiple training signals.
12. System according to claim 11, wherein each combiner module comprises between 1 and 2 layers and wherein each layer comprises between 10 and 1000 neurons.
13. Neural network, comprising:
- a latent feature extractor built by the method according to any of claims 1 to 6 and/or with the system according to any of claims 7 to 12;
 - a combiner module downstream the latent feature extractor in case multiple input signals are provided to the neural network;
 - a specific task module downstream the latent feature extractor and the combiner module, respectively.

14. Neural network according to claim 13, further comprising a normaliser module upstream the latent feature extractor.

5 15. Neural network according to claim 13 or 14, wherein the latent feature extractor comprises between 10 and 30 layers and wherein each layer comprises between 10 and 1000 neurons and/or
10 wherein the specific task module comprises between 1 and 2 layers and wherein each layer comprises between 10 and 1000 neurons and/or the combiner module comprises between 1 and 2 layers and wherein each layer comprises between 10 and 1000 neurons.

15

FIG 1

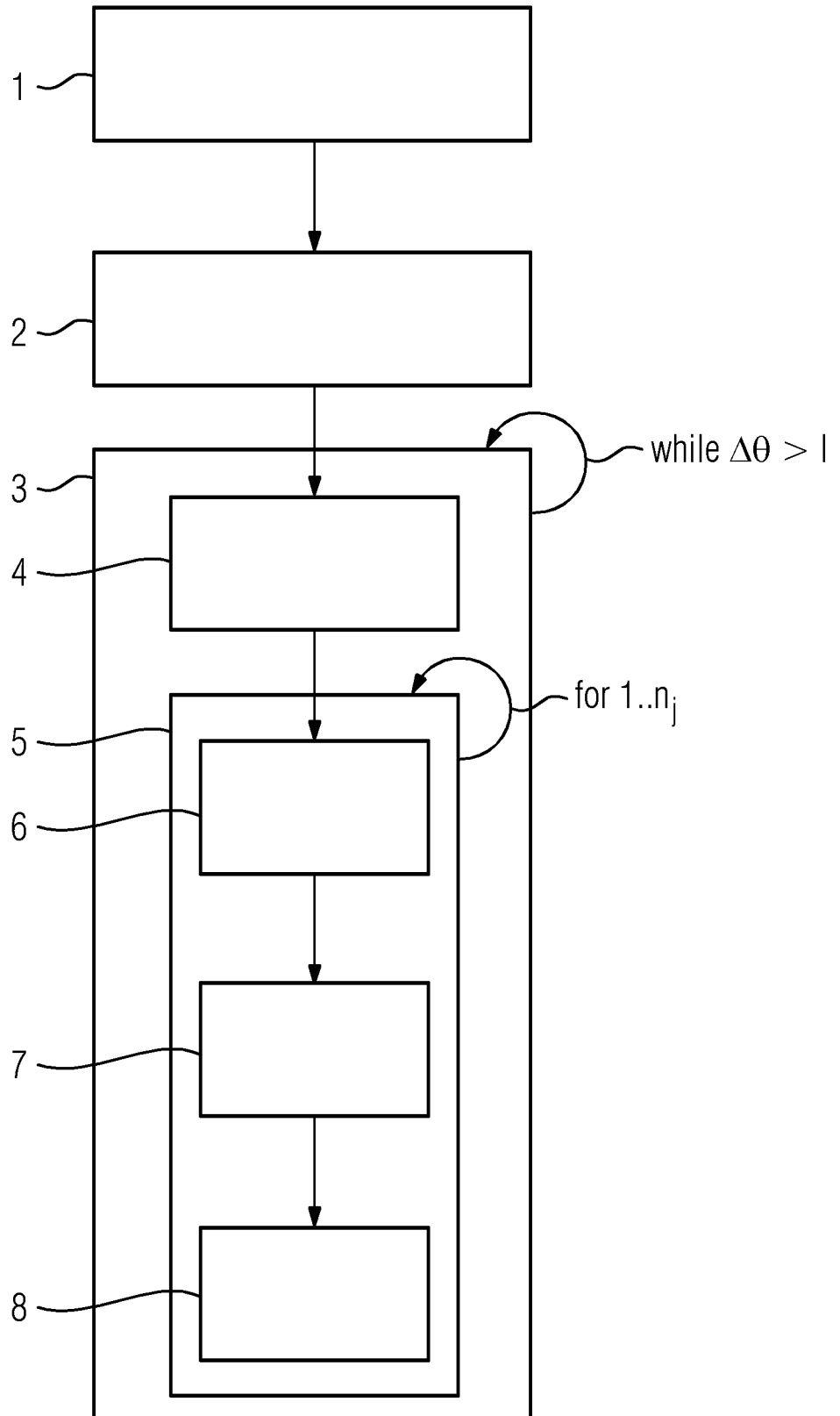


FIG 2

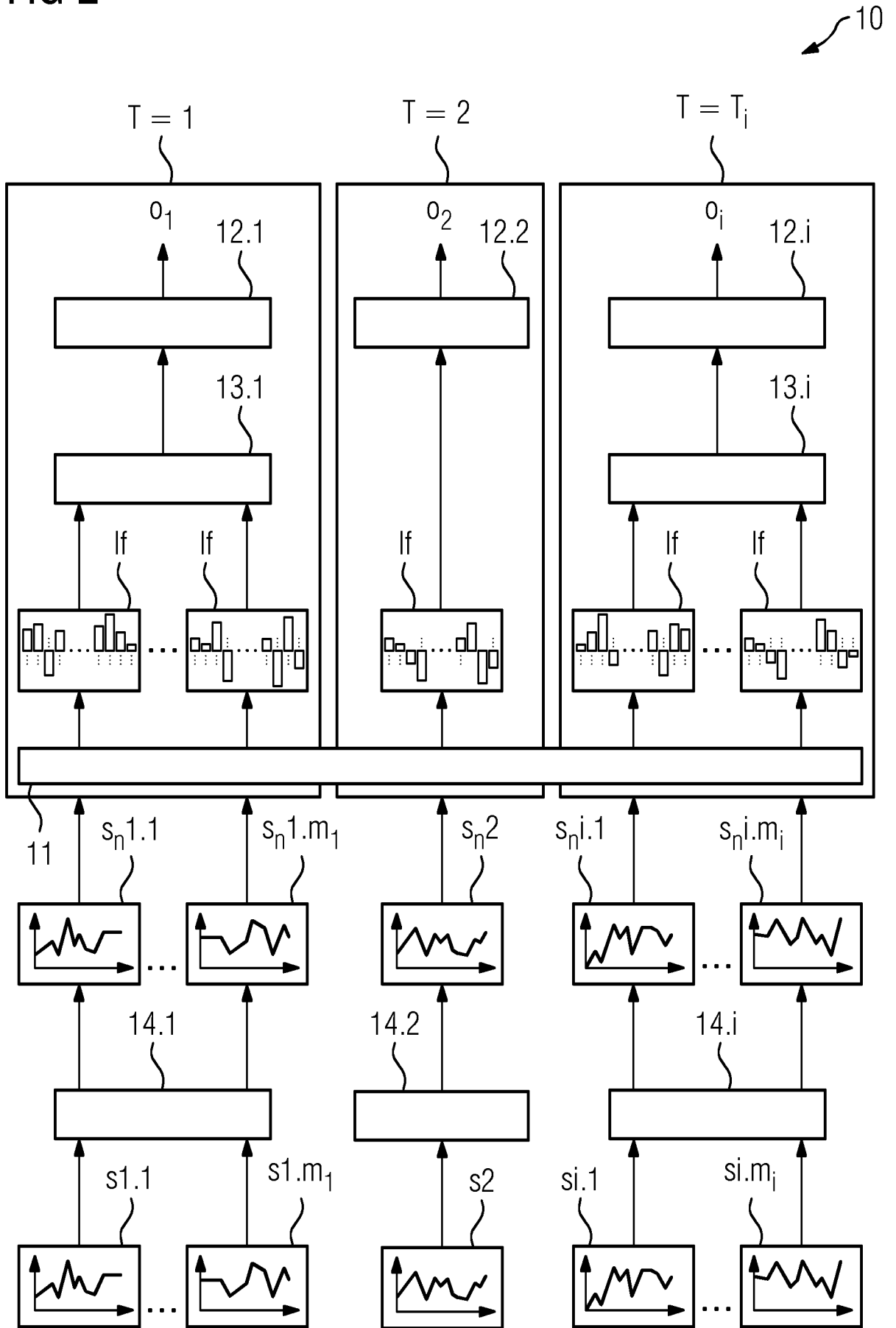


FIG 3

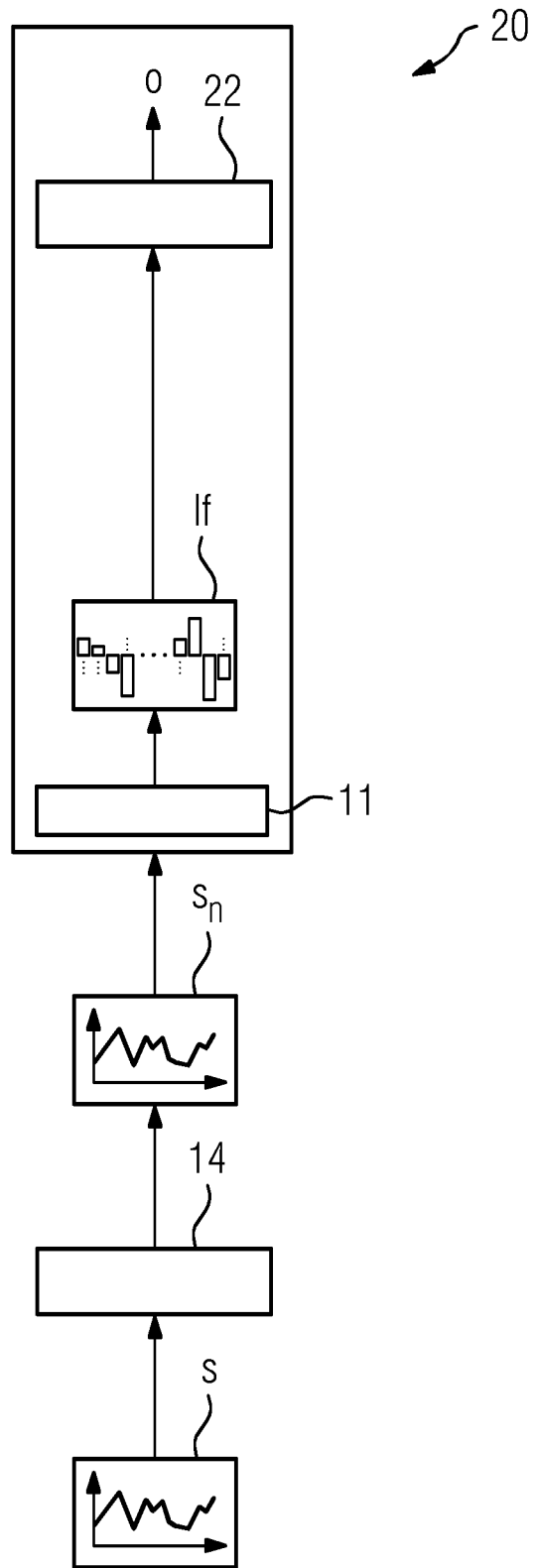
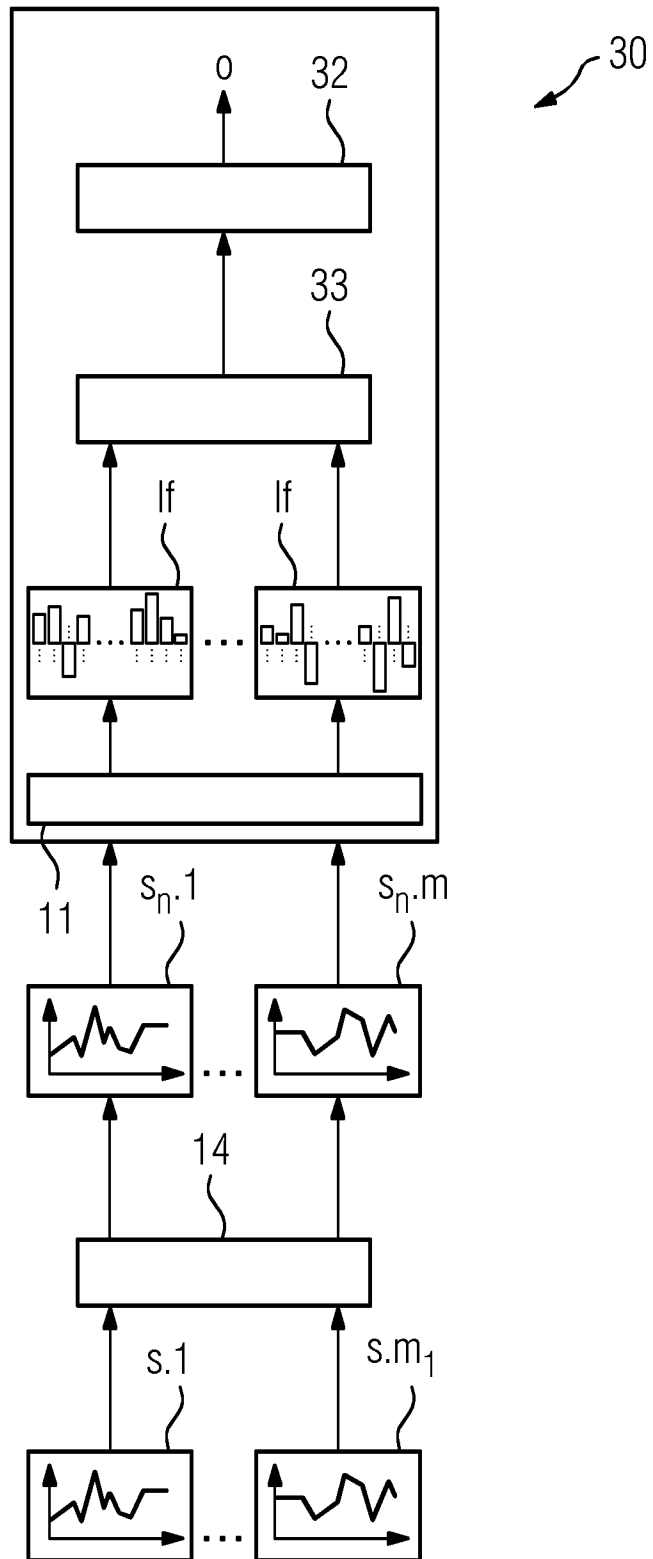


FIG 4



INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2020/051127

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06N3/04 G06N3/08
ADD.
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED
Minimum documentation searched (classification system followed by classification symbols)
G06N
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Karen Simonyan ET AL: "Two-Stream Convolutional Networks for Action Recognition in Videos", 9 June 2014 (2014-06-09), XP055324674, Retrieved from the Internet: URL:http://papers.nips.cc/paper/5353-two-stream-convolutional-networks-for-action-recognition-in-videos.pdf Sections 1, 4, 6 ----- -/--	1-15

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search 3 March 2020	Date of mailing of the international search report 12/03/2020
---	--

Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer Szymanski, Francois
--	---

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2020/051127

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	RONAN COLLOBERT ET AL: "A unified architecture for natural language processing", PROCEEDINGS OF THE 25TH INTERNATIONAL CONFERENCE ON MACHINE LEARNING, ICML '08, ACM PRESS, NEW YORK, NEW YORK, USA, 5 July 2008 (2008-07-05), pages 160-167, XP058106311, DOI: 10.1145/1390156.1390177 ISBN: 978-1-60558-205-4	1,7
A	Section 4.1 -----	2-6,8-15