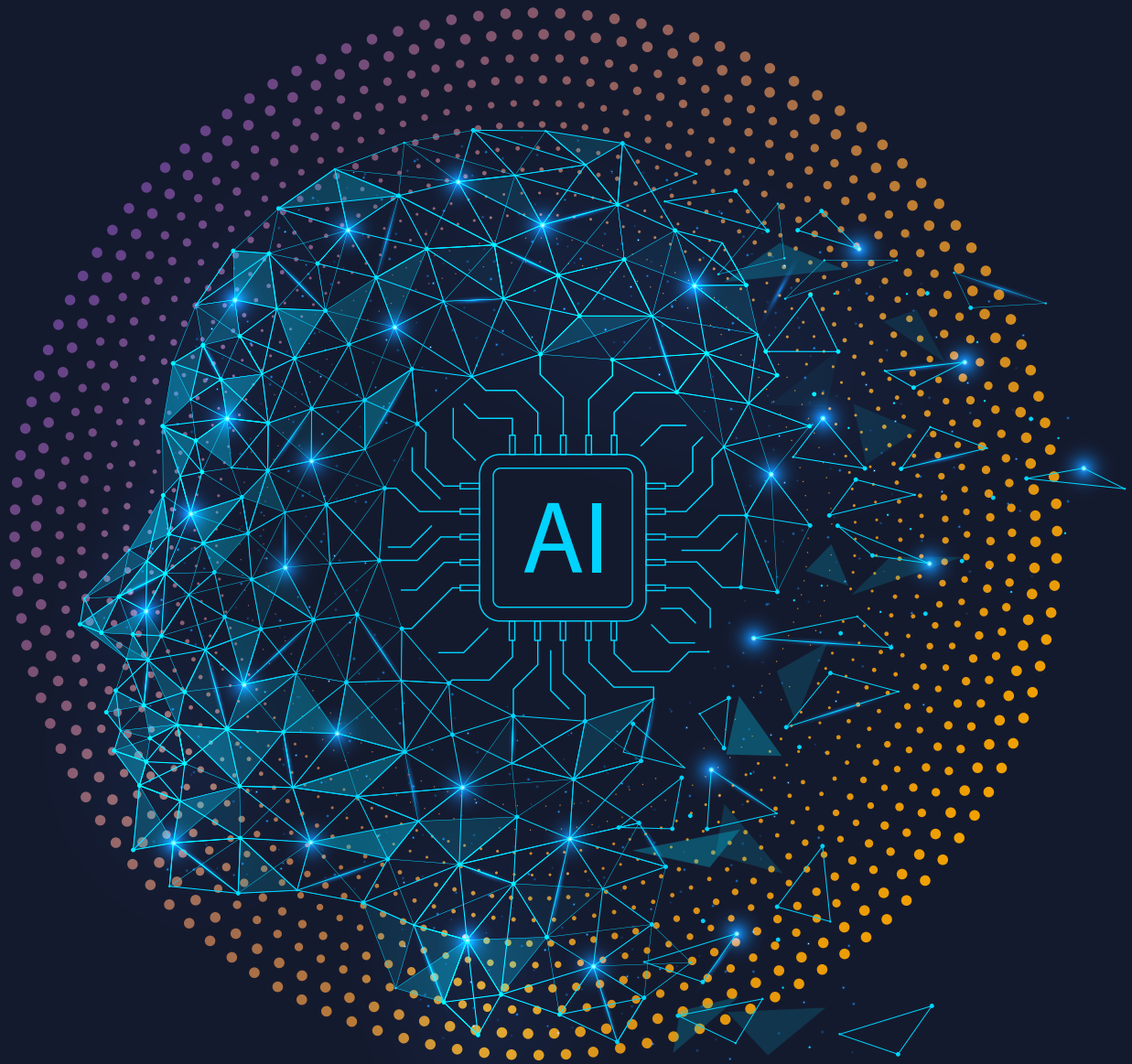
AI4media
ARTIFICIAL INTELLIGENCE FOR THE MEDIA AND SOCIETY

www.ai4media.eu

@ai4mediaproject

AI4Media Technological Highlights

Discover AI4Media's Key Research Outcomes on New Learning Paradigms & Distributed AI



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911

info@ai4media.eu www.ai4media.eu

Introduction

This booklet illustrates AI4Media contributions in investigating new learning paradigms tailored to media content and applications, which are pushing the borders of the state of the art in deep learning.

More information on these and other research outcomes of WP3 "New Learning Paradigms & Distributed AI" can be found in relevant public deliverables available here.

To this end, it discusses advances in lifelong and continuous learning, manifold and transfer learning, neural architecture search, quantum-assisted learning, and the fusion of evolutionary algorithms and deep learning. It also delves into decentralised and distributed computation solutions. The assets presented below are presented with an emphasis on their significance, showcasing the research highlights developed by collaborating partners.

For each contribution, the booklet specifies the target user groups, outlines the impact and added value to the media industry, and suggests potential future directions.

#1

FeTrIL: Feature Translation for Class-Incremental Learning

Partner organisations involved

CEA

→ **Code**

A few words about this technology

Continual learning enables artificial agents to learn from streamed data. This task is challenging when retraining with historical data is impossible when new classes occur. We propose a new method that combines the use of pre-trained deep models, class prototypes, the representation of past classes with data from new ones, and the training of an increasingly large classifier for each incremental state. The key contribution is the geometric translation of features toward the region of the representation space occupied by past classes to represent them. This translation is possible because the feature extractor is fixed. The method ensures a good balance between effectiveness and efficiency since it only requires the retraining of a linear layer. We show that an approximate version of the method can be deployed with very small performance loss.

Who can benefit

FeTrIL can be used by organisations that need continual learning algorithms applicable to data streams composed of new classes. It is useful when strong pre-trained models are already available and new classes need to be added to a visual recognition tool in a timely manner or with limited computational resources.

Impact and added value for the media industry

FeTrIL is flexible and can be deployed on top of any pre-trained feature extractor. Extensive experiments show that FeT compares favourably with more complex methods that use full model retraining when new data are ingested. The method is particularly useful when the number of classes is large, a setting in which competing methods relying on full model tuning become inoperant. Moreover, it can be combined with Asset #2, which focuses on novel class discovery to learn directly from unlabeled data.

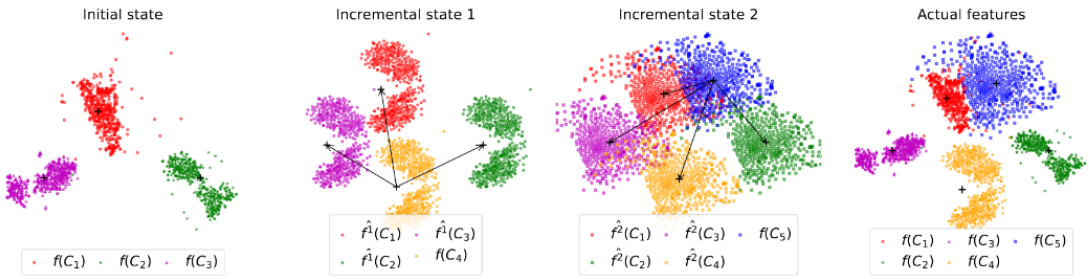
It is usable by media organisations with limited AI expertise since the method relies on the reuse of pre-trained models and the deployment of standard linear classifiers. For instance, media organisations could use the tool to extend their face recognition tools when novelty occurs in the news.

This is the case of new personalities, places, or events whose recognition can be integrated swiftly into existing tools. Such a capability is needed to keep pace with the fast-changing news landscape. It would allow these stakeholders to maintain the accuracy and coherence of their analyses over time. The method could also be applied to archives to enrich the associated metadata when new recognition capabilities are added.

Future developments on the technology

FeTrIL relies on an approximate representation of past classes. First, the current approximation is suboptimal because it hypothesises that the distributions of similar classes in the features space are themselves similar. We will research ways to improve this approximation by exploiting distributional data in addition to class prototypes.

Second, the obtained accuracy is correlated with the number of new classes available in each step. We will test the use of generative adversarial networks to produce pseudo-features. This component would improve results in cases when only a few classes are added.



#2

Novel Class Discovery

Partner organisations involved
UNITN

→ **Code here and here**

A few words about this technology

We study the new task of class-incremental Novel Class Discovery (class-iNCD), which refers to the problem of discovering novel categories in an unlabelled data set by leveraging a pre-trained model that has been trained on a labelled data set containing disjoint yet related categories. Apart from discovering novel classes, we also aim at preserving the ability of the model to recognize previously seen base categories. Inspired by rehearsal-based incremental learning methods, we propose a novel approach for class-iNCD which prevents forgetting of past information about the base classes by jointly exploiting base class feature prototypes and feature-level knowledge distillation. We also propose a self-training clustering strategy that simultaneously clusters novel categories and trains a joint classifier for both the base and novel classes. This makes our method able to operate in a class-incremental setting.
Our experiments, conducted on three common benchmarks, demonstrate that our method significantly outperforms state-of-the-art approaches.

Who can benefit

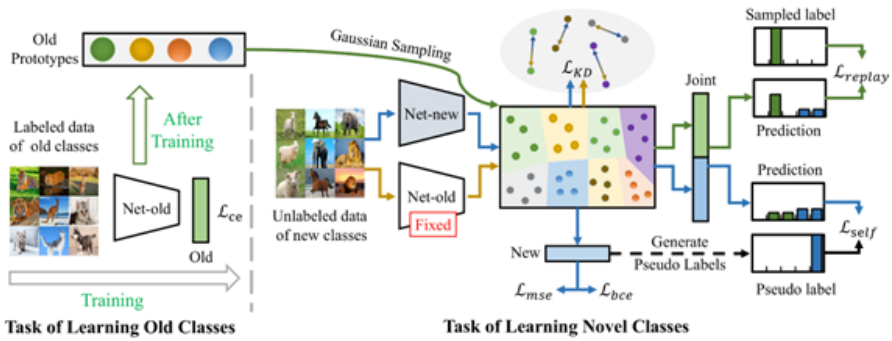
The tool could be useful to all the users (both professionals and amateurs) that need to employ a classification process in an unsupervised setting.

Impact and added value for the media industry

The important feature of our approach is that it is rather generic and as such can be applied to a large variety of applications. Additionally, our approach differs from the traditional NCD since we are not only interested in discovering novel classes but also we also aim to prevent forgetting the old classes. To prevent this forgetting phenomenon we proposed feature replay and feature-level distillation that is well suited for the class-iNCD. Moreover, to make inference task-agnostic, we propose to maintain a joint classifier that can classify any of the previously seen classes. We train this joint classifier using the pseudo-labels generated by the novel classifier head that is trained with a clustering loss. We compared our method to many relevant works and obtained superior performance on various benchmarks.

Future developments on the technology

The goal is to move towards Generalised category discovery (GCD) which aims at grouping unlabeled samples from known and unknown classes, given labelled data of known classes. Also, to meet the recent decentralisation trend in the community, we will consider a practical yet challenging task, namely Federated GCD (Fed-GCD), where the training data are distributively stored in local clients and cannot be shared among clients.



#3
ManifoldMixMS

Partner organisations involved

JR

→ **Code**

A few words about this technology

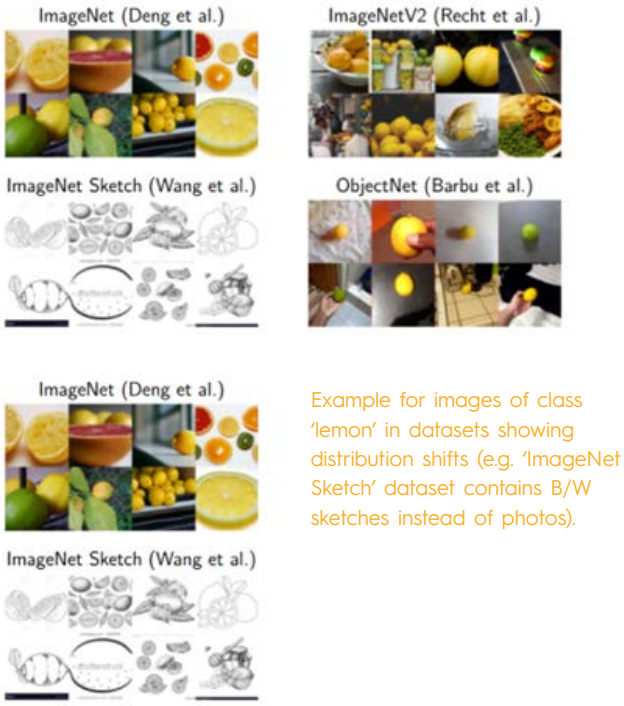
The standard recipe applied in transfer learning is to finetune a pretrained model on the task-specific dataset with different hyperparameter settings and pick the model with the highest accuracy on the validation dataset. Unfortunately, this leads to models which do not perform well under distribution shifts, e.g. when the model is given graphical sketches of the object as input instead of photos. In order to address this, we propose the manifold mixing model soup, an algorithm which mixes together the latent space manifolds of multiple finetuned models in an optimal way in order to generate a fused model. We show that the fused model gives significantly better out-of-distribution performance (+3.5 % compared to best individual model) when finetuning a CLIP model for image classification. In addition, it also provides better accuracy on the original dataset where the finetuning has been done.

Who can benefit

The presented fusing algorithm is quite general and therefore can help to obtain better performing AI models for a variety of computer vision tasks, especially when the input data differs drastically from the data used for training the model. These tasks can be e.g. image classification, object detection, segmentation, action recognition, image/video enhancement and many others.

Impact and added value for the media industry

The media industry employs a lot of computer vision technology in various stages of the chain, from content production, post-production to providing the content via different channels. Better AI models result in better quality of the content and ultimately a better user experience when consuming the content.



Example for images of class 'lemon' in datasets showing distribution shifts (e.g. 'ImageNet Sketch' dataset contains B/W sketches instead of photos).

Future developments on the technology

The proposed technology is not specific to any neuronal network architecture or input modality. Having demonstrated its usability for images, we therefore intend to employ this fusing algorithm also for Large Language Models (LLMs) and multimodal foundation models.

#4

Finding non-linear RBF Generative Paths in GAN Latent Space

Partner organisations involved
QMUL

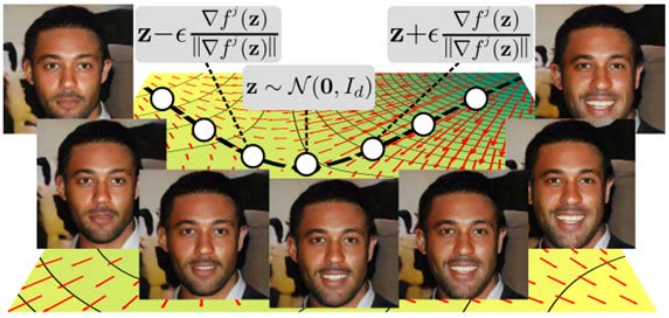
→ **Code**

A few words about this technology

We study the problem of discovering interpretable generative paths in the latent space of a pre-trained GAN. That is, paths sampling across which is expected to lead to image generation where only one (or a few) generative factors are activated. For instance, we seek for paths that change only the facial attributes of synthetic faces (preserving the identity and other irrelevant characteristics). We do so in a model-agnostic (i.e., our method is applicable to any GAN architecture) and totally unsupervised way (no additional manual annotation is needed).

Who can benefit

The tool could be useful to all the users (both professionals and amateurs) that need to use a pre-trained GAN generator for controllable content (images) creation.



Impact and added value for the media industry

The media industry can benefit from this methodology since it allows for the interpretable and controllable generation of synthetic content, which can be crucial in cases where privacy is important (i.e., using synthetic faces instead of real ones).

Future developments on the technology

The goal is to move towards generation with finer-grained control over the content based on various conditioning modalities (such as natural language) and beyond the scope of GANs (i.e., using Diffusion Models).

#5

PandA: Unsupervised Learning of Parts and Appearances in the Feature Maps of GANs

Partner organisations involved

QMUL

→ Code

A few words about this technology

We develop an architecture-agnostic method for improving local editing using Generative Adversarial Networks (GANs). Unlike existing methods, which are often specific to particular GAN architectures and either lack localised control or require manual input, the proposed approach discovers spatial parts and their appearances in an unsupervised manner. This method enables precise, context-aware local image editing at the pixel level. The discovered factors also serve as saliency maps, highlighting key concepts without any labels. Experimental results across various GAN architectures and datasets demonstrate that this method is more efficient and offers significantly better-localised control than current state-of-the-art techniques.

Who can benefit

Our face anonymization framework can be useful to all end-users (both professionals and amateurs) that need to obfuscate the identity of a given face, but at the same time retain characteristics/attributes that are useful, such as hair style, skin colour, facial expression, etc.

Impact and added value for the media industry

The method facilitates more precise spatial control in synthetic image generation. Any media industry generating synthetic content might find the method or its ideas useful for synthetic content.

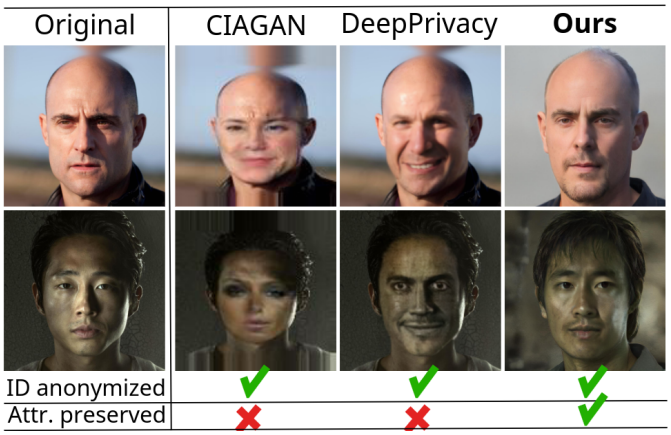


Illustration of anonymized results using the proposed anonymization framework (Ours) in comparison with state-of-the-art methods (CIAGAN and DeepPrivacy). The anonymized counterpart produced by our method cannot only de-identify the original identity effectively, but also faithfully retain its facial attributes.

#6

solo-learn: A Library of Self-supervised Methods for Visual Representation Learning

Partner organisations involved
UNITN

→ **Code**

A few words about this technology

We have put considerable effort into putting together a library of self-supervised methods for unsupervised visual representation learning powered by PyTorch Lightning. We aim at providing SOTA self-supervised methods in a comparable environment while, at the same time, implementing training tricks. The library is self-contained, but it is possible to use the models outside of solo-learn. solo-learn opens up avenues for exploiting large-budget SSL solutions on inexpensive smaller infrastructures and seeks to democratise SSL by making it accessible to all.

Who can benefit

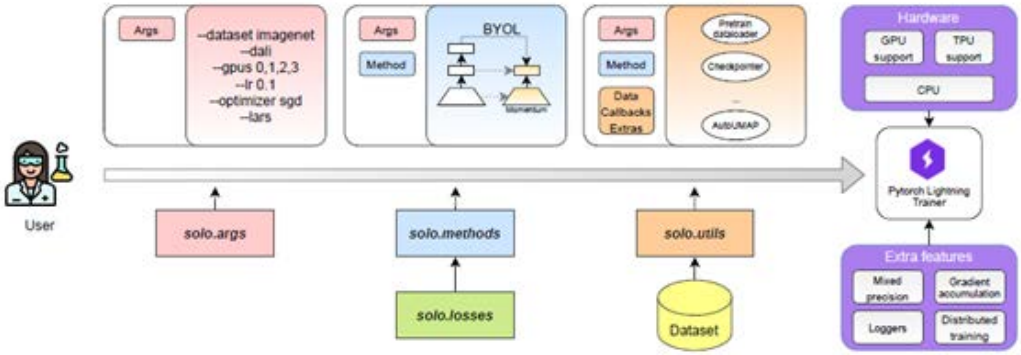
The tool could be useful to all the users (both professionals and amateurs) working on self-supervised methods for unsupervised visual representation learning.

Impact and added value for the media industry

Deep networks trained with large annotated datasets have shown stunning capabilities in the context of computer vision and in media applications. However, the need for human supervision is a strong limiting factor. Unsupervised learning aims to mitigate this issue by training models from unlabeled datasets. The most prominent paradigm for unsupervised visual representation learning is Self-supervised Learning (SSL), where the intrinsic structure of the data provides supervision for the model. Recently, the scientific community devised increasingly effective SSL methods that match or surpass the performance of supervised methods. Nonetheless, implementing and reproducing such works turns out to be complicated. Official repositories of state-of-the-art SSL methods have very heterogeneous implementations or no implementation at all. Solo-learn is meant to cover these problems and to be an aggregation platform for many of the existing technologies.

Future developments on the technology

We are continuously adding new SSL methods, improving usability, documents, and tutorials.



#7

Generalised Funnelling: A Method Based on Ensemble Learning and Heterogeneous Document Embeddings for Cross-Lingual Text Classification

Partner organisations involved

CNR

→ Code

→ Paper

A few words about this technology

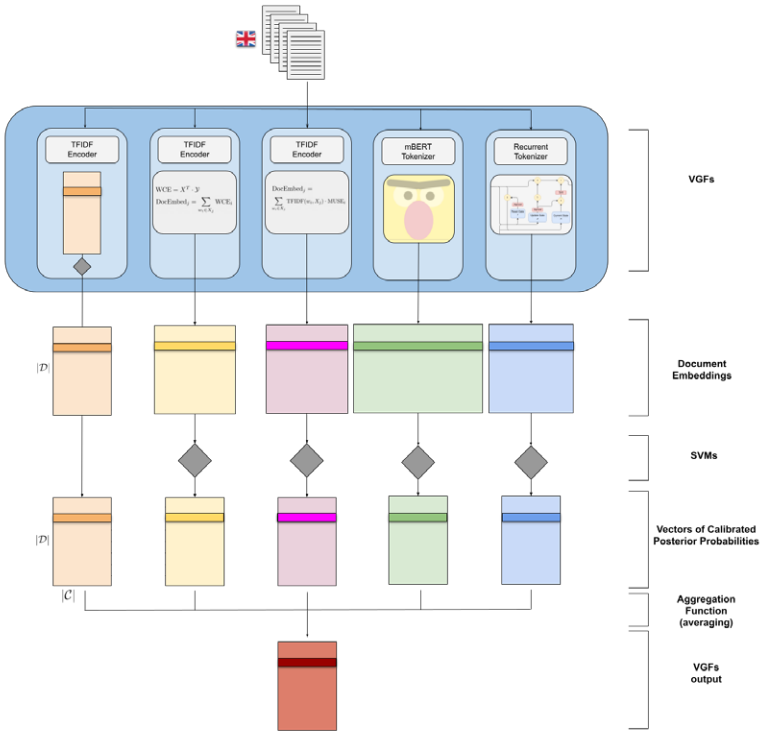
Generalised Funnelling (gFun) is a generalisation of the Funnelling heterogeneous transfer learning architecture for cross-lingual text classification (CLTC) in which 1st-tier components can be arbitrary view-generating functions, i.e., language-dependent functions that each produce a language-independent representation ("view") of the (monolingual) document. We have released an instance of gFun in which the meta-classifier receives as input a vector of calibrated posterior probabilities (as in Fun) aggregated to other embedded representations that embody other types of correlations, such as word-class correlations (as encoded by Word-Class Embeddings), word-word correlations (as encoded by Multilingual Unsupervised or Supervised Embeddings), and word-context correlations (as encoded by multilingual BERT). We have shown that this instance of gFun substantially improves over Fun and over state-of-the-art baselines, by reporting experimental results obtained on two large, standard datasets for multilingual multilabel text classification. Our code that implements gFun is publicly available.

Who can benefit

Immediate beneficiaries of the library are developers of cross-lingual text classification (CLTC) applications that need to generate text classifiers for a language in which training documents are scarce (few-shot) or absent (zero-shot).

Impact and added value for the media industry

Within the AI4Media project, the technology has been taken up by two industrial partners, RAI and ATC. RAI has developed, with CNR's help, a cross-lingual text classification application for classifying news in languages (such as French, German, and Slovenian) spoken in regions of (northern) Italy inhabited by significant numbers of speakers of languages other than Italian. ATC has developed, with CNR's help, a cross-lingual text classification application for performing sentiment classification in three different languages.



Future developments on the technology

We are currently testing the technology in the cross-modal scenario, where different modalities (e.g., text, image) are used instead of different languages. A future step might consist in tackling cross-lingual and cross-modal at the same time.

#8

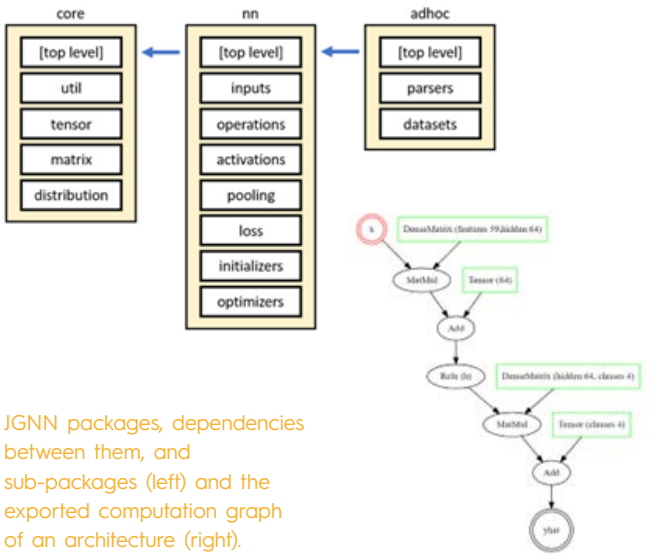
JGNN: A native Java library for Graph Neural Networks

Partner organisations involved
CERTH

→ **Code**

A few words about this technology

This is a library to define, train, and run Graph Neural Networks (GNNs) in native Java. GNNs are an emerging machine learning paradigm that organises relational data into graphs and takes advantage of their structure to improve the predictive ability of traditional neural networks. This paradigm is used in predictive and recommender systems to perform tasks like node classification, link prediction, and graph classification (classifying the types of graphs). This library implements GNN and other machine learning components in native Java, without external dependencies. Implementations are cross-platform and efficient both in terms of memory and data transfers, and support threaded parallelization. GNN architectures can be built either with a programmatic combination of components or by parsing Python-like expressions. Components include interoperable dense and sparse matrix operations and inline parameter definitions. The library can perform inference on pretrained architectures, but also supports training and fine-tuning without needing GPUs while keeping training times tractable (the bottleneck in running and training GNNs is typically sparse matrix operations, where GPU parallelization yields small running time gains). The library supports logical debugging with named tensor dimensions, visual and algorithmic validation of computation graphs, and comprehensive error messages that expose the internal states of computations.



JGNN packages, dependencies between them, and sub-packages (left) and the exported computation graph of an architecture (right).

Who can benefit

Immediate beneficiaries of the library are software developers that need to deploy cross-platform AI solutions on edge devices, especially if these need to adapt to local personalised content. In particular, JGNN serves as a prototyping tool for graph analysis and supports AI training and fine-tuning on the edge devices themselves.

Impact and added value for the media industry

JGNN facilitates the lightweight adoption of AI on edge devices, for example by converting computationally intensive local data to graphs (e.g., images can be converted to scene graphs) and then processing or learning on these with GNN architectures that are significantly more lightweight than other technologies. Training or refinement of architectures within user devices enables the creation of systems that are privacy-aware (they do not need to communicate with a central service to learn) but also offer a high degree of personalization. This enables the creation of "smarter" applications for the media industry that can improve the experience of both content creators and audiences.
One example application would be learning to recommend media items based on device user actions in social media (e.g., which accounts, posts/news articles, or images they view). This could include local -and therefore private- monitoring of user activity to gather more information for accurate personalization. It could also take into account media stored in users' devices by converting local media to graphs and using those to learn what content the user may be interested in. The AI making recommendations could either be trained from scratch on each user device or be trained beforehand in some dataset and fine-tuned on each device. To safeguard user data while personalising the recommendations, this application involves some amount of training within edge devices, which is not supported by other machine learning technologies.

Future developments on the technology

We are continuously extending and optimising in terms of speed and memory footprint the library's capabilities, based on new use cases arising from practical usage and input from the open source community.

#9

Filter-Pruning of Lightweight Face Detectors Using a Geometric Median Criterion

Partner organisations involved

CERTH

→ Code

A few words about this technology

This asset serves the purpose of enabling face detection in images/videos in the most computationally-efficient manner. The asset comprises a method for making existing lightweight face detectors even more compact and lightweight, and a collection of materials that resulted from the application of this method, specifically: a collection of pre-pruned model weights for two already small and compact face detectors, named EXTD (Extremely Tiny Face Detector) and EResFD (Efficient ResNet Face Detector); and all the required files for Android deployment of the most efficient among the two (i.e., the EResFD model and its pruned versions) using the Torchscript framework.

The devised method for making lightweight face detectors even more compact and lightweight belongs to a class of techniques known as filter pruning. The main pruning algorithm that we utilise is Filter Pruning via Geometric Median (FPGM), combined with the Soft Filter Pruning (SFP) iterative procedure. We also apply L1 Norm pruning, as a baseline to compare with the devised method. Our experiments showed that the devised method not only has the potential to further reduce the model size of already lightweight face detectors with limited accuracy loss, but can even lead to small accuracy gains for low pruning rates.

Who can benefit

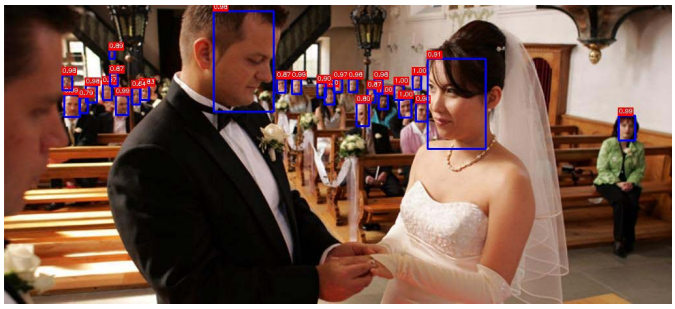
Immediate beneficiaries of the pruning network and the resulting lightweight face detection models are mobile application developers, who need to introduce face detection functionalities in their applications. The users of such applications benefit from shorter response times and lower computation and energy consumption when using face detection functionalities.

Impact and added value for the media industry

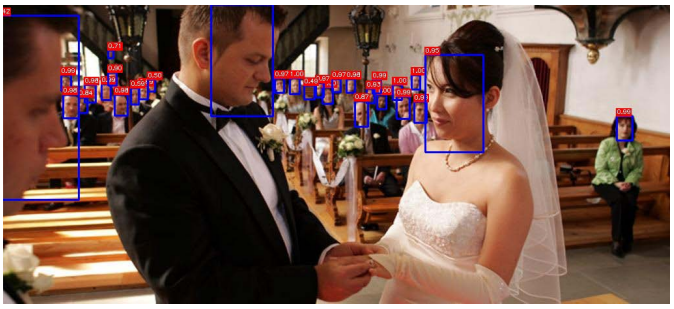
Face detectors are an important component of many applications, which often have to run on edge devices with limited processing power and memory. Therefore, there is a pressing demand for compact face detection models that can function efficiently across resource-constrained devices.

One example application in the media industry is the detection of faces in images/videos captured with a resource-constrained mobile device for blurring them, before these media items are sent via a communication network to any third party, to protect the identity of the depicted individuals (who may be witnesses of a crime, for example). Similarly, the protection of children from exposure in online social networks often calls for covering their faces in media items that are posted (e.g., by replacing their faces with face icons), and this can be easily performed immediately after the capture of the media item (on the camera itself), by detecting their faces and producing an additional online-safe version of the captured image/video.

Media archives can also benefit from reducing the computational and energy consumption requirements of face detection, which is a processing step often used for facilitating the annotation and retrieval of archival content. While this processing is typically performed on powerful servers rather than on resource-constrained mobile or edge devices, energy consumption is becoming more and more of a major consideration. Using more lightweight models for performing this task directly contributes to reduced energy consumption and lower environmental footprint for this task.



Original model.



Pruned model.

Future developments on the technology

We continue to work on improving our pruning method, with the purpose of identifying with even higher accuracy which bits and pieces of a given face detector network contribute the most to the good results of the network. This will help us to further optimise the pruned models, further improving the accuracy vs. efficiency balance of the pruned face detectors that we can produce for use in resource-constrained devices.

#10

Open-Ended Evolution for Minecraft Building Generation

Partner organisations involved
UM

→ **Paper**
→ **Video**

A few words about this technology

This asset introduces a procedural content generator which evolves Minecraft buildings according to an open-ended and intrinsic definition of novelty. To accomplish this goal, we evaluate individuals' novelty in the latent space using a 3D autoencoder, and alternate between phases of exploration and transformation. During exploration, the system evolves multiple populations of CPPNs through CPPN-NEAT and constrained novelty search in the latent space (defined by the current autoencoder). We apply a set of repair and constraint functions to ensure candidates adhere to basic structural rules and constraints during evolution. During transformation, we reshape the boundaries of the latent space to identify new interesting areas of the solution space by retraining the autoencoder with novel content. We validate the proposed system using five different approaches for training the autoencoder during transformation and its impact on populations' quality and diversity during evolution. Our results show that by retraining the autoencoder we can achieve better open-ended complexity compared to a static model, which is further improved when retraining using larger datasets of individuals with diverse complexities.

Who can benefit

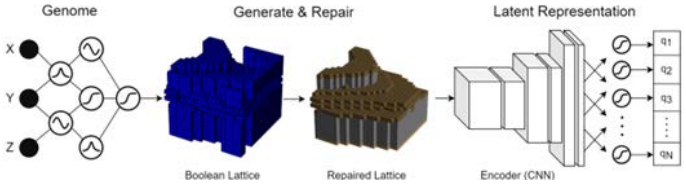
The current tool can be useful for game designers for generating novel game content. With modifications to the underlying models, this framework can also be useful for designers looking to generate other artefacts such as visuals (images) and audio.

Impact and added value for the media industry

As it achieves some level of open-ended novelty/complexity, this technology can help train generative AI models which are less susceptible to convergence and can produce new interesting outputs over time. This asset extends an established algorithm called DeLeNoX (which was introduced to improve the open-ended creativity of content generators) into a 3D application. This can be helpful for generative media in domains such as games, images, audio where generating streams of novel content for the end user is desirable (i.e. avoiding repetition in the output). It can function both as an autonomous generator, or as a design assistant for designers to draw inspiration from during their design process.

Future developments on the technology

The current technology has been tested with limited constraints on the output, and therefore does not produce realistic results (buildings, in our example case). Therefore, in future work, testing with more constrained generation is crucial to ensure open-ended novelty/complexity is preserved in more realistic outputs.



Generating 3D structures using CPPNs and compressing them into latent vectors using an encoder.

#11

MAP-Elites with Transverse Assessment (MEliTa) for Multimodal Problems in Creative Domains

Partner organisations involved

UM

→ **Paper**

A few words about this technology

The recent advances in language-based generative models have paved the way for the orchestration of multiple generators of different artefact types (text, image, audio, etc.) into one system. Presently, many open-source pre-trained models combine text with other modalities, thus enabling shared vector embeddings to be compared across different generators. As a test bed for this asset, we generate text descriptions and cover images for a hypothetical video game and assign each artefact a unique modality-specific behavioural characteristic. Results indicate that MEliTA can improve text-to-image mappings within the solution space, compared to a baseline MAP-Elites algorithm that strictly treats each image-text pair as one solution. Our approach represents a significant step forward in multimodal bottom-up orchestration and lays the groundwork for more complex systems coordinating multimodal creative agents in the future.

Who can benefit

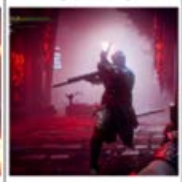
This technology can be useful for media creators, both beginner and professional, across multiple domains. It can function as part of a design assistant system or autonomously generate content for writers, game developers, content creators, and more.

Impact and added value for the media industry

We propose a novel approach to handle multimodal creative tasks using Quality Diversity evolution. This takes the form of a variation of the MAP-Elites algorithm, called MEliTA, which is tailored for multimodal creative tasks and leverages deep-learned models that assess coherence across modalities. Thanks to this, MEliTA has the potential to help in many areas of the media industry. In our test case, we show its capability in generating text and visuals for fictional games. This can also be extended to other domains in media, such as video and audio generation. It could also be useful for designers in other use cases, such as generating text and visuals for stories and articles. This can function as an autonomous generator, or as part of a design assistant system for writers/designers to take inspiration from as they work.

Future developments on the technology

One of the more interesting areas for future development for MEliTA involves generating more complex artefact modalities and more than two at once, such as generating audio, visuals and story for an interactive reading experience. The algorithm can also be tested with more behavioural characteristics to improve upon the diversity of the generated content.

Game 1	Game 2	Game 3	Game 4	Game 5
				
In this game, players have to use their sword and weapons wisely because there are many enemies who will try very hard for you. After many defeats in a long time, each one offers new challenges.	After a global catastrophe, you are left alone to face your past. The greatest evil - an alien race called Korda – is planning on tearing Earth apart as they do every few centuries in this new adventure from Arkane Studios and Infamous Games.	Play as a shadow warrior, taking on different threats in this third person shooter that's more brutal than ever before.	In this game, players have to use their swords and bows in order not only survive but also fight with others.	After years' long struggle, Dark Lord Arthur's dark legacy is coming to light, and a new evil seems lurking at every corner. TheShadow Warrior2 : Shadows OfthePast.

#12

QuaPy: A Python-Based Framework for Quantification

Partner organisations involved
CNR

→ **Code**

A few words about this technology

QuaPy is an open-source framework for performing quantification (a.k.a. supervised prevalence estimation), written in Python. Quantification is the task of training quantifiers via supervised learning, where a quantifier is a predictor that estimates the relative frequencies (a.k.a. prevalence values) of the classes of interest in a sample of unlabelled data. While quantification can be trivially performed by applying a standard classifier to each unlabelled data item and counting how many data items have been assigned to each class, it has been shown that this "classify and count" method is outperformed by methods specifically designed for quantification. QuaPy provides implementations of a number of baseline methods and advanced quantification methods, of routines for quantification-oriented model selection, of several broadly accepted evaluation measures, and of robust evaluation protocols routinely used in the field. QuaPy also makes available datasets commonly used for testing quantifiers, and offers visualisation tools for facilitating the analysis and interpretation of the results. The software is open-source and publicly available under a BSD-3 licence via GitHub, and can be installed via pip.

Who can benefit

Immediate beneficiaries of the library are developers of applications that require estimating the relative frequencies (a.k.a. prevalence values) of the classes of interest in a sample of unlabelled data. Examples include estimating the relative frequencies of different sentiments (e.g., Positive, Neutral, Negative) towards specific issues in corpora of news; an example application is gauging the bias of a news source towards specific political parties.

Impact and added value for the media industry

An example application for quantification technologies in the media industry may consist of estimating the relative frequencies of different sentiments (e.g., Positive, Neutral, Negative), or stances, towards specific issues in corpora of news, with the goal of gauging the bias of a news source towards specific political parties. Another application may consist of estimating the relative frequencies of different sentiments towards specific political candidates in the corpora of microblog (e.g., Twitter). More in general, quantification is helpful whenever estimating the relative frequencies of the classes of interest in samples of unlabelled data needs to be performed with high accuracy.

Future developments on the technology

We keep updating the library with new quantification methods as they are being developed. The library has become a reference for the quantification community as a whole, so that other researchers from other nations are also actively contributing.

#13

RL Optimization of Quantum Circuits

Partner organisations involved

BSC

→ **Paper**

A few words about this technology

Reinforcement Learning (RL) method for optimising quantum circuits using graph-theoretic simplification rules of ZX-diagrams. The agent, trained using the Proximal Policy Optimization (PPO) algorithm, employs Graph Neural Networks to approximate the policy and value functions.

Who can benefit

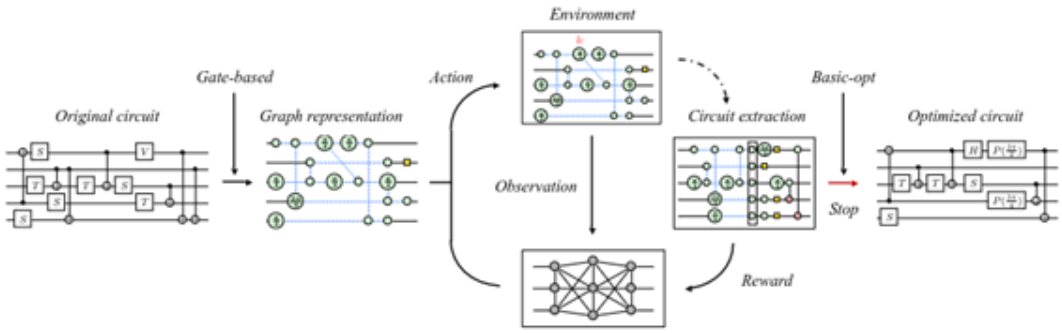
Users of Quantum computers with limited capabilities require optimization of Quantum circuits to maximise the exploitation of this technology. This approach allows the use of Quantum processors beyond their nominal specs.

Impact and added value for the media industry

This approach allows the optimization of any Quantum circuit. In particular, any media processing task involving Quantum circuits may benefit from this technology. In particular, computer vision tasks implemented as a Quantum workflow may reduce the Quantum device requirements. This may allow the use of these methods in a shorter time.

Future developments on the technology

It is expected that future versions will improve the results. In particular, the operations allowed by the agent, together with an improved training process, will allow to obtain even optimised smaller Quantum circuits.



Schematic view of the operations involved in the simplification of a Quantum circuit using an RL agent acting over manipulations formulated with ZX calculus.

AI4media

ARTIFICIAL INTELLIGENCE FOR THE MEDIA AND SOCIETY

Our Consortium

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911

info@ai4media.eu www.ai4media.eu