Activity Report 2019

Team LINKMEDIA

Creating and Exploiting Explicit Links between Multimedia Fragments

Joint team with Inria Rennes – Bretagne Atlantique

D6 – Media and Interactions

Table of contents

Project-Team LINKMEDIA

Creation of the Project-Team: 2014 July 01

Keywords:

Computer Science and Digital Science:

A3.3.2. - Data mining
A3.3.3. - Big data analysis
A5.3.3. - Pattern recognition
A5.4.1. - Object recognition
A5.4.3. - Content retrieval
A5.7. - Audio modeling and processing
A5.7.1. - Sound
A5.7.3. - Speech
A5.8. - Natural language processing
A9.2. - Machine learning
A9.3. - Signal analysis
A9.4. - Natural language processing

Other Research Topics and Application Domains:

B9. - Society and Knowledge
B9.3. - Medias
B9.6.10. - Digital humanities
B9.10. - Privacy

1. Team, Visitors, External Collaborators

Research Scientists

Laurent Amsaleg [Team leader, CNRS, Senior Researcher, HDR]
Guillaume Gravier [CNRS, Senior Researcher, HDR]
Yannis Avrithis [Inria, Advanced Research Position]
Vincent Claveau [CNRS, Researcher]
Teddy Furon [Inria, Researcher, HDR]
Miaojing Shi [Inria, Starting Research Position, until Nov 2019]

Faculty Members

Ewa Kijak [Univ de Rennes I, Associate Professor]
Simon Malinowski [Univ de Rennes I, Associate Professor]
Pascale Sébillot [INSA Rennes, Professor, HDR]

Post-Doctoral Fellow

Suresh Kirthi Kumaraswamy [Univ Le Mans, from Mar 2019]

PhD Students

Benoit Bonnet [Inria, from Nov 2019]
Ricardo Carlini Sperandio [Univ Federal de Minas Gerais, until Jul 2019]
Clément Dalloux [CNRS, until Nov 2019]
Mikail Demirdelen [Univ de Rennes I]
Cheikh Brahim El Vaigh [Inria]

Marzieh Gheisari Khorasgani [Inria]
Yann Lifchitz [Safran Tech, CIFRE]
Cyrielle Mallart [SIPA-Ouest France, CIFRE]
Oriane Siméoni [Inria]
François Torregrossa [Pages Jaunes, CIFRE]
Hanwei Zhang [Chinese Scholarship Council]

Technical staff

Mateusz Budnik [Inria, Engineer]
Jean Hany [Inria, Engineer, until Nov 2019]
Guillaume Le Noe-Bienvenu [CNRS, Engineer, from Nov 2019]
Cédric Maigrot [Inria, Engineer, from May 2019]
Florent Michel [Inria, Engineer, from May 2019]

Administrative Assistant

Aurélie Patier [Univ de Rennes I]

Visiting Scientist

Raquel Pereira de Almeida [Univ Federal de Minas Gerais, from Feb 2019]

2. Overall Objectives

2.1. Context

LINKMEDIA is concerned with the processing of extremely large collections of multimedia material. The material we refer to are collections of documents that are created by humans and intended for humans. It is material that is typically created by media players such as TV channels, radios, newspapers, archivists (BBC, INA, ...), as well as the multimedia material that goes through social-networks. It also includes material that includes images, videos and pathology reports for e-health applications, or that is in relation with e-learning which typically includes a fair amount of texts, graphics, images and videos associating in new ways teachers and students. It also includes material in relation with humanities that study societies through the multimedia material that has been produced across the centuries, from early books and paintings to the latest digitally native multimedia artifacts. Some other multimedia material are out of the scope of LINKMEDIA, such as the ones created by cameras or sensors in the broad areas of video-surveillance or satellite images.

Multimedia collections are rich in contents and potential, that richness being in part within the documents themselves, in part within the relationships between the documents, in part within what humans can discover and understand from the collections before materializing its potential into new applications, new services, new societal discoveries, ... That richness, however, remains today hardly accessible due to the conjunction of several factors originating from the inherent nature of the collections, the complexity of bridging the semantic gap or the current practices and the (limited) technology:

- *Multimodal:* multimedia collections are composed of very diverse material (images, texts, videos, audio, ...), which require sophisticated approaches at analysis time. Scientific contributions from past decades mostly focused on analyzing each media in isolation one from the other, using modality-specific algorithms. However, revealing the full richness of collections calls for jointly taking into account these multiple modalities, as they are obviously semantically connected. Furthermore, involving resources that are external to collections, such as knowledge bases, can only improve gaining insight into the collections. Knowledge bases form, in a way, another type of modality with specific characteristics that also need to be part of the analysis of media collections. Note that determining what a document is about possibly mobilizes a lot of resources, and this is especially costly and time consuming for audio and video. Multimodality is a great source of richness, but causes major difficulties for the algorithms running analysis;

- *Intertwined*: documents do not exist in isolation one from the other. There is more knowledge in a collection than carried by the sum of its individual documents and the relationships between documents also carry a lot of meaningful information. (Hyper)Links are a good support for materializing the relationships between documents, between parts of documents, and having analytic processes creating them automatically is challenging. Creating semantically rich typed links, linking elements at very different granularities is very hard to achieve. Furthermore, in addition to being disconnected, there is often no strong structure into each document, which makes even more difficult their analysis;
- *Collections are very large:* the scale of collections challenges any algorithm that runs analysis tasks, increasing the duration of the analysis processes, impacting quality as more irrelevant multimedia material gets in the way of relevant ones. Overall, scale challenges the complexity of algorithms as well as the quality of the result they produce;
- *Hard to visualize*: It is very difficult to facilitate humans getting insight on collections of multimedia documents because we hardly know how to display them due to their multimodal nature, or due to their number. We also do not know how to well present the complex relationships linking documents together: granularity matters here, as full documents can be linked with small parts from others. Furthermore, visualizing time-varying relationships is not straightforward. Data visualization for multimedia collections remains quite unexplored.

2.2. Scientific objectives

The ambition of LINKMEDIA is to propose **foundations, methods, techniques and tools to help humans make sense of extremely large collections of multimedia material**. Getting useful insight from multimedia is only possible if tools and users interact tightly. Accountability of the analysis processes is paramount in order to allow users understanding their outcome, to understand why some multimedia material was classified this way, why two fragments of documents are now linked. It is key for the acceptance of these tools, or for correcting errors that will exist. Interactions with users, facilitating analytics processes, taking into account the trust in the information and the possible adversarial behaviors are topics LINKMEDIA addresses.

3. Research Program

3.1. Scientific background

LINKMEDIA is de facto a multidisciplinary research team in order to gather the multiple skills needed to enable humans to gain insight into extremely large collections of multimedia material. It is *multimedia data* which is at the core of the team and which drives the design of our scientific contributions, backed-up with solid experimental validations. *Multimedia data*, again, is the rationale for selecting problems, applicative fields and partners.

Our activities therefore include studying the following scientific fields:

- multimedia: content-based analysis; multimodal processing and fusion; multimedia applications;
- computer vision: compact description of images; object and event detection;
- machine learning: deep architectures; structured learning; adversarial learning;
- natural language processing: topic segmentation; information extraction;
- information retrieval: high-dimensional indexing; approximate k-nn search; embeddings;
- data mining: time series mining; knowledge extraction.

3.2. Workplan

Overall, LINKMEDIA follows two main directions of research that are (i) extracting and representing information from the documents in collections, from the relationships between the documents and from what user build from these documents, and (ii) facilitating the access to documents and to the information that has been elaborated from their processing.

3.2.1. Research Direction 1: Extracting and Representing Information

LINKMEDIA follows several research tracks for *extracting* knowledge from the collections and *representing* that knowledge to facilitate users acquiring gradual, long term, constructive insights. Automatically processing documents makes it crucial to consider the accountability of the algorithms, as well as understanding when and why algorithms make errors, and possibly invent techniques that compensate or reduce the impact of errors. It also includes dealing with malicious adversaries carefully manipulating the data in order to compromise the whole knowledge extraction effort. In other words, LINKMEDIA also investigates various aspects related to the *security* of the algorithms analyzing multimedia material for knowledge extraction and representation.

Knowledge is not solely extracted by algorithms, but also by humans as they gradually get insight. This human knowledge can be materialized in computer-friendly formats, allowing algorithms to use this knowledge. For example, humans can create or update ontologies and knowledge bases that are in relation with a particular collection, they can manually label specific data samples to facilitate their disambiguation, they can manually correct errors, etc. In turn, knowledge provided by humans may help algorithms to then better process the data collections, which provides higher quality knowledge to humans, which in turn can provide some better feedback to the system, and so on. This virtuous cycle where algorithms and humans cooperate in order to make the most of multimedia collections requires specific support and techniques, as detailed below.

3.2.1.1. Machine Learning for Multimedia Material.

Many approaches are used to extract relevant information from multimedia material, ranging from very low-level to higher-level descriptions (classes, captions, ...). That diversity of information is produced by algorithms that have varying degrees of supervision. Lately, fully supervised approaches based on deep learning proved to outperform most older techniques. This is particularly true for the latest developments of Recurrent Neural Networkds (RNN, such as LSTMs) or convolutional neural network (CNNs) for images that reach excellent performance [62]. LINKMEDIA contributes to advancing the state of the art in computing representations for multimedia material by investigating the topics listed below. Some of them go beyond the very processing of multimedia material as they also question the fundamentals of machine learning procedures when applied to multimedia.

- *Learning from few samples/weak supervisions.* CNNs and RNNs need large collections of carefully annotated data. They are not fitted for analyzing datasets where few examples per category are available or only cheap image-level labels are provided. LINKMEDIA investigates low-shot, semi-supervised and weakly supervised learning processes: Augmenting scarce training data by automatically propagating labels [65], or transferring what was learned on few very well annotated samples to allow the precise processing of poorly annotated data [74]. Note that this context also applies to the processing of heritage collections (paintings, illuminated manuscripts, ...) that strongly differ from contemporary natural images. Not only annotations are scarce, but the learning processes must cope with material departing from what standard CNNs deal with, as classes such as "planes", "cars", etc, are irrelevant in this case.
- *Ubiquitous Training.* NN (CNNs, LSTMs) are mainstream for producing representations suited for high-quality classification. Their training phase is ubiquitous because the same representations can be used for tasks that go beyond classification, such as retrieval, few-shot, meta- and incremental learning, all boiling down to some form of metric learning. We demonstrated that this ubiquitous training is relatively simpler [65] yet as powerful as ad-hoc strategies fitting specific tasks [79]. We study the properties and the limitations of this ubiquitous training by casting metric learning as a classification problem.
- *Beyond static learning*. Multimedia collections are by nature continuously growing, and ML processes must adapt. It is not conceivable to re-train a full new model at every change, but rather to support continuous training and/or allowing categories to evolve as the time goes by. New classes may be defined from only very few samples, which links this need for dynamicity to the low-shot learning problem discussed here. Furthermore, active learning strategies determining which is the next sample to use to best improve classification must be considered to alleviate the annotation cost and the re-training process [69]. Eventually, the learning process may need to manage an extremely

large number of classes, up to millions. In this case, there is a unique opportunity of blending the expertise of LINKMEDIA on large scale indexing and retrieval with deep learning. Base classes can either be "summarized" e.g. as a multi-modal distribution, or their entire training set can be made accessible as an external associative memory [86].

- *Learning and lightweight architectures.* Multimedia is everywhere, it can be captured and processed on the mobile devices of users. It is necessary to study the design of lightweight ML architectures for mobile and embedded vision applications. Inspired by [90], we study the savings from quantizing hyper-parameters, pruning connections or other approximations, observing the trade-off between the footprint of the learning and the quality of the inference. Once strategy of choice is progressive learning which early aborts when confident enough [70].
- *Multimodal embeddings.* We pursue pioneering work of LINKMEDIA on multimodal embedding, i.e., representing multiple modalities or information sources in a single embedded space [83], [85], [84]. Two main directions are explored: exploiting adversarial architectures (GANs) for embedding via translation from one modality to another, extending initial work in [84] to highly heterogeneous content; combining and constraining word and RDF graph embeddings to facilitate entity linking and explanation of lexical co-occurrences [81].
- *Accountability of ML processes.* ML processes achieve excellent results but it is mandatory to verify that accuracy results from having determined an adequate problem representation, and not from being abused by artifacts in the data. LINKMEDIA designs procedures for at least explaining and possibly interpreting and understanding what the models have learned. We consider heat-maps materializing which input (pixels, words) have the most importance in the decisions [77], Taylor decompositions to observe the individual contributions of each relevance scores or estimating LID [47] as a surrogate for accounting for the smoothness of the space.
- *Extracting information.* ML is good at extracting features from multimedia material, facilitating subsequent classification, indexing, or mining procedures. LINKMEDIA designs extraction processes for identifying parts in the images [75], [76], relationships between the various objects that are represented in images [53], learning to localizing objects in images with only weak, image-level supervision [78] or fine-grained semantic information in texts [58]. One technique of choice is to rely on generative adversarial networks (GAN) for learning low-level representations. These representations can e.g. be based on the analysis of density [89], shading, albedo, depth, etc.
- *Learning representations for time evolving multimedia material.* Video and audio are time evolving material, and processing them requests to take their time line into account. In [71], [57] we demonstrated how shapelets can be used to transform time series into time-free high-dimensional vectors, preserving however similarities between time series. Representing time series in a metric space improves clustering, retrieval, indexing, metric learning, semi-supervised learning and many other machine learning related tasks. Research directions include adding localization information to the shapelets, fine-tuning them to best fit the task in which they are used as well as designing hierarchical representations.

3.2.1.2. Adversarial Machine Learning.

Systems based on ML take more and more decisions on our behalf, and maliciously influencing these decisions by crafting adversarial multimedia material is a potential source of dangers: a small amount of carefully crafted noise imperceptibly added to images corrupts classification and/or recognition. This can naturally impact the insight users get on the multimedia collection they work with, leading to taking erroneous decisions e.g.

This adversarial phenomenon is not particular to deep learning, and can be observed even when using other ML approaches [52]. Furthermore, it has been demonstrated that adversarial samples generalize very well across classifiers, architectures, training sets. The reasons explaining why such tiny content modifications succeed in producing severe errors are still not well understood.

We are left with little choice: we must gain a better understanding of the weaknesses of ML processes, and in particular of deep learning. We must understand why attacks are possible as well as discover mechanisms protecting ML against adversarial attacks (with a special emphasis on convolutional neural networks). Some initial contributions have started exploring such research directions, mainly focusing on images and computer vision problems. Very little has been done for understanding adversarial ML from a *multimedia* perspective [56].

LINKMEDIA is in a unique position to throw at this problem new perspectives, by experimenting with other modalities, used in isolation one another, as well as experimenting with true multimodal inputs. This is very challenging, and far more complicated and interesting than just observing adversarial ML from a computer vision perspective. No one clearly knows what is at stake with adversarial audio samples, adversarial video sequences, adversarial ASR, adversarial NLP, adversarial OCR, all this being often part of a sophisticated multimedia processing pipeline.

Our ambition is to lead the way for initiating investigations where the full diversity of modalities we are used to work with in multimedia are considered from a perspective of adversarial attacks and defenses, both at learning and test time. In addition to what is described above, and in order to trust the multimedia material we analyze and/or the algorithms that are at play, LINKMEDIA investigates the following topics:

- *Beyond classification.* Most contributions in relation with adversarial ML focus on classification tasks. We started investigating the impact of adversarial techniques on more diverse tasks such as retrieval [46]. This problem is related to the very nature of euclidean spaces where distances and neighborhoods can all be altered. Designing defensive mechanisms is a natural companion work.
- *Detecting false information.* We carry-on with earlier pioneering work of LINKMEDIA on false information detection in social media. Unlike traditional approaches in image forensics [60], we build on our expertise in content-based information retrieval to take advantage of the contextual information available in databases or on the web to identify out-of-context use of text or images which contributed to creating a false information [72].
- *Deep fakes.* Progress in deep ML and GANs allow systems to generate realistic images and are able to craft audio and video of existing people saying or doing things they never said or did [68]. Gaining in sophistication, these machine learning-based "deep fakes" will eventually be almost indistinguishable from real documents, making their detection/rebutting very hard. LINKMEDIA develops deep learning based counter-measures to identify such modern forgeries. We also carry on with making use of external data in a provenance filtering perspective [91] in order to debunk such deep fakes.
- *Distributions, frontiers, smoothness, outliers.* Many factors that can possibly explain the adversarial nature of some samples are in relation with their distribution in space which strongly differs from the distribution of natural, genuine, non adversarial samples. We are investigating the use of various information theoretical tools that facilitate observing distributions, how they differ, how far adversarial samples are from benign manifolds, how smooth is the feature space, etc. In addition, we are designing original adversarial attacks and develop detection and curating mechanisms [47].

3.2.1.3. Multimedia Knowledge Extraction.

Information obtained from collections via computer ran processes is not the only thing that needs to be represented. Humans are in the loop, and they gradually improve their level of understanding of the content and nature of the multimedia collection. Discovering knowledge and getting insight is involving multiple people across a long period of time, and what each understands, concludes and discovers must be recorded and made available to others. Collaboratively inspecting collections is crucial. Ontologies are an often preferred mechanism for modeling what is inside a collection, but this is probably limitative and narrow.

LINKMEDIA is concerned with making use of existing strategies in relation with ontologies and knowledge bases. In addition, LINKMEDIA uses mechanisms allowing to materialize the knowledge gradually acquired by humans and that might be subsequently used either by other humans or by computers in order to better and more precisely analyze collections. This line of work is instantiated at the core of the iCODA project LINKMEDIA coordinates. We are therefore concerned with:

- *Multimedia analysis and ontologies.* We develop approaches for linking multimedia content to entities in ontologies for text and images, building on results in multimodal embedding to cast entity linking into a nearest neighbor search problem in a high-dimensional joint embedding of content and entities [85]. We also investigate the use of ontological knowledge to facilitate information extraction from content [9].
- *Explainability and accountability in information extraction.* In relation with ontologies and entity linking, we develop innovative approaches to explain statistical relations found in data, in particular lexical or entity co-occurrences in textual data, for example using embeddings constrained with translation properties of RDF knowledge or path-based explanation within RDF graphs. We also work on confidence measures in entity linking and information extraction, studying how the notions of confidence and information source can be accounted for in knowledge basis and used in human-centric collaborative exploration of collections.
- *Dynamic evolution of models for information extraction.* In interactive exploration and information extraction, e.g., on cultural or educational material, knowledge progressively evolves as the process goes on, requiring on-the-fly design of new models for content-based information extractors from very few examples, as well as continuous adaptation of the models. Combining in a seamless way low-shot, active and incremental learning techniques is a key issue that we investigate to enable this dynamic mechanisms on selected applications.

3.2.1.4. Research Direction 2: Accessing Information

LINKMEDIA centers its activities on enabling humans to make good use of vast multimedia collections. This material takes all its cultural and economic value, all its artistic wonder when it can be accessed, watched, searched, browsed, visualized, summarized, classified, shared, ... This allows users to fully enjoy the incalculable richness of the collections. It also makes it possible for companies to create business rooted in this multimedia material.

Accessing the multimedia data that is inside a collection is complicated by the various type of data, their volume, their length, etc. But it is even more complicated to access the information that is not materialized in documents, such as the relationships between parts of different documents that however share some similarity. LINKMEDIA in its first four years of existence established itself as one of the leading teams in the field of multimedia analytics, contributing to the establishment of a dedicated community (refer to the various special sessions we organized with MMM, the iCODA and the LIMAH projects, as well as [66], [67], [63]).

Overall, facilitating the access to the multimedia material, to the relevant information and the corresponding knowledge asks for algorithms that efficiently *search* collections in order to identify the elements of collections or of the acquired knowledge that are matching a query, or that efficiently allow *navigating* the collections or the acquired knowledge. Navigation is likely facilitated if techniques are able to handle information and knowledge according to hierarchical perspectives, that is, allow to reveal data according to various levels of details. Aggregating or *summarizing* multimedia elements is not trivial.

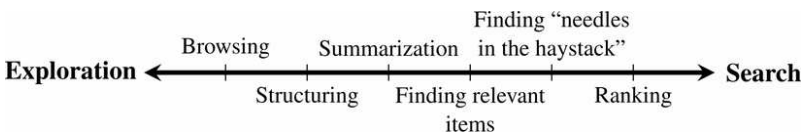

Figure 1. Exploration-search axis with example tasks

Three topics are therefore in relation with this second research direction. LINKMEDIA tackles the issues in relation to searching, to navigating and to summarizing multimedia information. Information needs when discovering the content of a multimedia collection can be conveniently mapped to the exploration-search axis, as first proposed by Zahálka and Worring in [88], and illustrated by Figure 1 where expert users typically work near the right end because their tasks involve precise queries probing search engines. In contrast, lay-users start near the exploration end of the axis. Overall, users may alternate searches and explorations by going back and forth along the axis. The underlying model and system must therefore be highly dynamic, support interactions with the users and propose means for easy refinements. LINKMEDIA contributes to advancing the state of the art in searching operations, in navigating operations (also referred to as browsing), and in summarizing operations.

3.2.1.4.1. Searching.

Search engines must run similarity searches very efficiently. High-dimensional indexing techniques therefore play a central role. Yet, recent contributions in ML suggest to revisit indexing in order to adapt to the specific properties of modern features describing contents.

- *Advanced scalable indexing.* High-dimensional indexing is one of the foundations of LINKMEDIA. Modern features extracted from the multimedia material with the most recent ML techniques shall be indexed as well. This, however, poses a series of difficulties due to the dimensionality of these features, their possible sparsity, the complex metrics in use, the task in which they are involved (instance search, k-nn, class prototype identification, manifold search [65], time series retrieval, ...). Furthermore, truly large datasets require involving sketching [50], secondary storage and/or distribution [49], [48], alleviating the explosion of the number of features to consider due to their local nature or other innovative methods [64], all introducing complexities. Last, indexing multimodal embedded spaces poses a new series of challenges.
- *Improving quality.* Scalable indexing techniques are approximate, and what they return typically includes a fair amount of false positives. LINKMEDIA works on improving the quality of the results returned by indexing techniques. Approaches taking into account neighborhoods [59], manifold structures instead of pure distance based similarities [65] must be extended to cope with advanced indexing in order to enhance quality. This includes feature selection based on intrinsic dimensionality estimation [47].
- *Dynamic indexing.* Feature collections grow, and it is not an option to fully reindex from scratch an updated collection. This trivially applies to the features directly extracted from the media items, but also to the base class prototypes that can evolve due to the non-static nature of learning processes. LINKMEDIA will continue investigating what is at stake when designing dynamic indexing strategies.

3.2.1.4.2. Navigating.

Navigating a multimedia collection is very central to its understanding. It differs from searching as navigation is not driven by any specific query. Rather, it is mostly driven by the relationships that various documents have one another. Relationships are supported by the links between documents and/or parts of documents. Links rely on semantic similarity, depicting the fact that two documents share information on the same topic. But other aspects than semantics are also at stake, e.g., time with the dates of creation of the documents or geography with mentions or appearance in documents of some geographical landmarks or with geo-tagged data.

In multimedia collections, links can be either implicit or explicit, the latter being much easier to use for navigation. An example of an implicit link can be the name of someone existing in several different news articles; we, as humans, create a mental link between them. In some cases, the computer misses such configurations, leaving such links implicit. Implicit links are subject to human interpretation, hence they are sometimes hard to identify for any automatic analysis process. Implicit links not being materialized, they can therefore hardly be used for navigation or faceted search. Explicit links can typically be seen as hyperlinks, established either by content providers or, more aligned with LINKMEDIA, automatically determined from

content analysis. Entity linking (linking content to an entity referenced in a knowledge base) is a good example of the creation of explicit links. Semantic similarity links, as investigated in the LIMAH project and as considered in the search and hyperlinking task at MediaEval and TRECVid, are also prototypical links that can be made explicit for navigation. Pursuing work, we investigate two main issues:

- *Improving multimodal content-based linking.* We exploit achievements in entity linking to go beyond lexical or lexico-visual similarity and to provide semantic links that are easy to interpret for humans; carrying on, we work on link characterization, in search of mechanisms addressing link explainability (i.e., what is the nature of the link), for instance using attention models so as to focus on the common parts of two documents or using natural language generation; a final topic that we address is that of linking textual content to external data sources in the field of journalism, e.g., leveraging topic models and cue phrases along with a short description of the external sources.
- *Dynamicity and user-adaptation.* One difficulty for explicit link creation is that links are often suited for one particular usage but not for another, thus requiring creating new links for each intended use; whereas link creation cannot be done online because of its computational cost, the alternative is to generate (almost) all possible links and provide users with selection mechanisms enabling personalization and user-adaptation in the exploration process; we design such strategies and investigate their impact on exploration tasks in search of a good trade-off between performance (few high-quality links) and genericity.

3.2.1.4.3. Summarizing.

Multimedia collections contain far too much information to allow any easy comprehension. It is mandatory to have facilities to aggregate and summarize a large body on information into a compact, concise and meaningful representation facilitating getting insight. Current technology suggests that multimedia content aggregation and story-telling are two complementary ways to provide users with such higher-level views. Yet, very few studies already investigated these issues. Recently, video or image captioning [87], [82] have been seen as a way to summarize visual content, opening the door to state-of-the-art multi-document text summarization [61] with text as a pivot modality. Automatic story-telling has been addressed for highly specific types of content, namely TV series [54] and news [73], [80], but still need a leap forward to be mostly automated, e.g., using constraint-based approaches for summarization [51], [80].

Furthermore, not only the original multimedia material has to be summarized, but the knowledge acquired from its analysis is also to summarize. It is important to be able to produce high-level views of the relationships between documents, emphasizing some structural distinguishing qualities. Graphs establishing such relationships need to be constructed at various level of granularity, providing some support for summarizing structural traits.

Summarizing multimedia information poses several scientific challenges that are:

- *Choosing the most relevant multimedia aggregation type*: Taking a multimedia collection into account, a same piece of information can be present in several modalities. The issue of selecting the most suitable one to express a given concept has thus to be considered together with the way to mix the various modalities into an acceptable production. Standard summarization algorithms have to be revisited so that they can handle continuous representation spaces, allowing them to benefit from the various modalities [55].
- *Expressing user's preferences*: Different users may appreciate quite different forms of multimedia summaries, and convenient ways to express their preferences have to be proposed. We for example focus on the opportunities offered by the constraint-based framework.
- *Evaluating multimedia summaries*: Finding criteria to characterize what a good summary is remains challenging, e.g., how to measure the global relevance of a multimodal summary and how to compare information between and across two modalities. We tackle this issue particularly via a collaboration with A. Smeaton at DCU, comparing the automatic measures we will develop to human judgments obtained by crowd-sourcing;

- *Taking into account structuring and dynamicity*: Typed links between multimedia fragments, and hierarchical topical structures of documents obtained via work previously developed within the team are two types of knowledge which have seldom been considered as long as summarization is concerned. Knowing that the event present in a document is causally related to another event described in another document can however modify the ways summarization algorithms have to consider information. Moreover the question of producing coarse-to-fine grain summaries exploiting the topical structure of documents is still an open issue. Summarizing dynamic collections is also challenging and it is one of the questions we consider.

4. Application Domains

4.1. Asset management in the entertainment business

Media asset management—archiving, describing and retrieving multimedia content—has turned into a key factor and a huge business for content and service providers. Most content providers, with television channels at the forefront, rely on multimedia asset management systems to annotate, describe, archive and search for content. So do archivists such as the Institut National de l'Audiovisuel, the bibliothèque Nationale de France, the Nederlands Instituut voor Beeld en Geluid or the British Broadcast Corporation, as well as media monitoring companies, such as Yacast in France. Protecting copyrighted content is another aspect of media asset management.

4.2. Multimedia Internet

One of the most visible application domains of linked multimedia content is that of multimedia portals on the Internet. Search engines now offer many features for image and video search. Video sharing sites also feature search engines as well as recommendation capabilities. All news sites provide multimedia content with links between related items. News sites also implement content aggregation, enriching proprietary content with user-generated content and reactions from social networks. Most public search engines and Internet service providers offer news aggregation portals. This also concerns TV on-demand and replay services as well as social TV services and multi-screen applications. Enriching multimedia content, with explicit links targeting either multimedia material or knowledge databases is central here.

4.3. Data journalism

Data journalism forms an application domain where most of the technology developed by LINKMEDIA can be used. On the one hand, data journalists often need to inspect multiple heterogeneous information sources, some being well structured, some other being fully unstructured. They need to access (possibly their own) archives with either searching or navigational means. To gradually construct insight, they need collaborative multimedia analytics processes as well as elements of trust in the information they use as foundations for their investigations. Trust in the information, watching for adversarial and/or (deep) fake material, accountability are all crucial here.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Highlights of the year

- Our activities in relation with fake news were extensively highlighted in 2019. Ewa Kijak and Vincent Claveau gave a few interviews in newspapers, in a nationwide radio broadcast as well as in several TV shows.

- A chaire position in Artificial Intelligence for Defense has been granted to Teddy Furon. This chaire is supported by the national Defense Innovation Agency. The chaire will last 4 years, starting early 2020.
- Laurent Amsaleg (General Chair), Guillaume Gravier (Program Committee Chair), Yannis Avrithis (Workshops Chair) as well as almost all students of LINKMEDIA (as volunteers) were involved in running the 27th ACM Multimedia conference in Nice. This edition, very successful, was attended by close to 800 people.

5.1.2. Awards

Oriane Siméoni received the best presentation award from the International Computer Vision Summer School (ICVSS) 2019[1].

6. New Software and Platforms

6.1. Lookinlabs-Global

KEYWORD: Search Engine

FUNCTIONAL DESCRIPTION: Lookinlabs allows you to find, among teams/individuals/publications, those best matching your query.

- Authors: William Kokou Dedzoe and Jean Hany
- Contact: Vincent Claveau

6.2. TagEx

Yet another Part-of-Speech Tagger for French

KEYWORD: Natural language processing

FUNCTIONAL DESCRIPTION: TagEx is available as a web-service on https://allgo.inria.fr . Refer to Allgo for its usage.

- Contact: Vincent Claveau
- URL: https://allgo.inria.fr/app/tagex

6.3. NegDetect

Negation Detection

KEYWORD: Natural language processing

FUNCTIONAL DESCRIPTION: NegDetect relies on several layers of machine learning techniques (CRF, neural networks).

- Contact: Vincent Claveau

7. New Results

7.1. Extracting and Representing Information

7.1.1. Text Mining in the Clinical Domain

Participants: Clément Dalloux, Vincent Claveau.

[1] https://iplab.dmi.unict.it/icvss2019/

Clinical records cannot be shared, which is a real hurdle to develop and compare information extraction techniques. In the framework of the BigClin Project we have developed annotated corpora, that share the same linguistic properties than records, but can be freely distributed for research purposes. Several corpora and several types of annotation were proposed for French, Portuguese and English. They are made freely available for research purposes and are described in [27], [25]. These corpora will foster reproducible research on clinical text mining.

Thanks to these datasets, we have organized the DeFT text-mining competition in 2019. Several NLP techniques and tools have been developed within the project in order to identify relevant medical or linguistic information [30], [26]. They are all chiefly based on machine learning approaches, and for most of them, more specifically, on deep learning. For instance, we have developed a new Part-of-Speech tagger and lemmatizer for French, especially suited to handle medical texts; it is freely available as a web-service at https://allgo.inria.fr . The identification of negation and uncertainty is important to precisely understand the clinical texts. Thus, we have continued our work on neural techniques to find the negation/uncertainty cues and their scope (part of sentence concerned by the negation or uncertainty). It achieves state-of-the-art results on English, and is pioneer work for French and Portuguese for which it sets a new standard [4], [21]; it is available at https://allgo.inria.fr . Other achievements in text-mining include: numerical value extraction (finding concepts that are measured, such as lab results, numerical expressions, their units) in French, English and Portuguese, the identification of gender, age, outcome and admission reasons in French clinical texts, ...

7.1.2. Embedding in hyperbolic spaces

Participants: François Torregrossa, Vincent Claveau, Guillaume Gravier.

During this year, we have studied non-Euclidean spaces into which one can embed data (for instance, words). We have developed the HierarX tool which projects multiple datasources into hyperbolic manifolds: Lorentz or Poincaré. From similarities between word pairs or continuous word representations in high dimensional spaces, HierarX is able to embed knowledge in hyperbolic geometries with small dimensionality. Those shape information into continuous hierarchies. The source code is available on the Inria's GitLab.

7.1.3. Aggregation and embedding for group membership verification

Participants: Marzieh Gheisari Khorasgani, Teddy Furon, Laurent Amsaleg.

This paper proposes a group membership verification protocol preventing the curious but honest server from reconstructing the enrolled signatures and inferring the identity of querying clients [24]. The protocol quantizes the signatures into discrete embeddings, making reconstruction difficult. It also aggregates multiple embeddings into representative values, impeding identification. Theoretical and experimental results show the trade-off between the security and error rates.

7.1.4. Group Membership Verification with Privacy: Sparse or Dense?

Participants: Marzieh Gheisari Khorasgani, Teddy Furon, Laurent Amsaleg.

Group membership verification checks if a biometric trait corresponds to one member of a group without revealing the identity of that member. Recent contributions provide privacy for group membership protocols through the joint use of two mechanisms: quantizing templates into discrete embeddings, and aggregating several templates into one group representation. However, this scheme has one drawback: the data structure representing the group has a limited size and cannot recognize noisy query when many templates are aggregated. Moreover, the sparsity of the embeddings seemingly plays a crucial role on the performance verification. This contribution proposes a mathematical model for group membership verification allowing to reveal the impact of sparsity on both security, compactness, and verification performances [23]. This models bridges the gap towards a Bloom filter robust to noisy queries. It shows that a dense solution is more competitive unless the queries are almost noiseless.

7.1.5. Privacy Preserving Group Membership Verification and Identification

Participants: Marzieh Gheisari Khorasgani, Teddy Furon, Laurent Amsaleg.

When convoking privacy, group membership verification checks if a biometric trait corresponds to one member of a group without revealing the identity of that member. Similarly, group membership identification states which group the individual belongs to, without knowing his/her identity. A recent contribution provides privacy and security for group membership protocols through the joint use of two mechanisms: quantizing biometric templates into discrete embeddings, and aggregating several templates into one group representation. This paper significantly improves that contribution because it jointly learns how to embed and aggregate instead of imposing fixed and hard coded rules [10]. This is demonstrated by exposing the mathematical underpinnings of the learning stage before showing the improvements through an extensive series of experiments targeting face recognition. Overall, experiments show that learning yields an excellent trade-off between security/privacy and the verification/identification performances.

7.1.6. Intrinsic Dimensionality Estimation within Tight Localities

Participants: Laurent Amsaleg, Oussama Chelly [Microsoft Germany], Michael Houle [National Institute of Informatics, Japan], Ken-Ichi Kawarabayashi [National Institute of Informatics, Japan], Miloš Radovanović [Univ. Novi Sad, Serbia], Weeris Treeratanajaru [Chulalongkorn University, Thailand].

Accurate estimation of Intrinsic Dimensionality (ID) is of crucial importance in many data mining and machine learning tasks, including dimensionality reduction, outlier detection, similarity search and subspace clustering. However, since their convergence generally requires sample sizes (that is, neighborhood sizes) on the order of hundreds of points, existing ID estimation methods may have only limited usefulness for applications in which the data consists of many natural groups of small size. In this paper, we propose a local ID estimation strategy stable even for 'tight' localities consisting of as few as 20 sample points [31]. The estimator applies MLE techniques over all available pairwise distances among the members of the sample, based on a recent extreme-value-theoretic model of intrinsic dimensionality, the Local Intrinsic Dimension (LID). Our experimental results show that our proposed estimation technique can achieve notably smaller variance, while maintaining comparable levels of bias, at much smaller sample sizes than state-of-the-art estimators.

7.1.7. Selective Biogeography-Based Optimizer Considering Resource Allocation for Large-Scale Global Optimization

Participants: Meiji Cui [Tongji University, China], Li Li [Tongji University, China], Miaojing Shi.

Biogeography-based optimization (BBO), a recent proposed meta-heuristic algorithm, has been successfully applied to many optimization problems due to its simplicity and efficiency. However, BBO is sensitive to the curse of dimensionality; its performance degrades rapidly as the dimensionality of the search space increases. In [3], a selective migration operator is proposed to scale up the performance of BBO and we name it selective BBO (SBBO). The differential migration operator is selected heuristically to explore the global area as far as possible whist the normal distributed migration operator is chosen to exploit the local area. By the means of heuristic selection, an appropriate migration operator can be used to search the global optimum efficiently. Moreover, the strategy of cooperative co-evolution (CC) is adopted to solve large-scale global optimization problems (LSOPs). To deal with subgroup imbalance contribution to the whole solution in the context of CC, a more efficient computing resource allocation is proposed. Extensive experiments are conducted on the CEC 2010 benchmark suite for large-scale global optimization, and the results show the effectiveness and efficiency of SBBO compared with BBO variants and other representative algorithms for LSOPs. Also, the results confirm that the proposed computing resource allocation is vital to the large-scale optimization within the limited computation budget.

7.1.8. Friend recommendation for cross marketing in online brand community based on intelligent attention allocation link prediction algorithm

Participants: Shugang Li [Shanghai University, China], Xuewei Song [Shanghai University, China], Hanyu Lu [Shanghai University, China], Linyi Zeng [Shanghai University, Industrial and Commercial Bank of China, China], Miaojing Shi, Fang Liu [Shanghai University, China].

Circle structure of online brand communities allows companies to conduct cross-marketing activities by the influence of friends in different circles and build strong and lasting relationships with customers. However, existing works on the friend recommendation in social network do not consider establishing friendships between users in different circles, which has the problems of network sparsity, neither do they study the adaptive generation of appropriate link prediction algorithms for different circle features. In order to fill the gaps in previous works, the intelligent attention allocation link prediction algorithm is proposed to adaptively build attention allocation index (AAI) according to the sparseness of the network and predict the possible friendships between users in different circles. The AAI reflects the amount of attention allocated to the user pair by their common friend in the triadic closure structure, which is decided by the friend count of the common friend. Specifically, for the purpose of overcoming the problem of network sparsity, the AAIs of both the direct common friends and indirect ones are developed. Next, the decision tree (DT) method is constructed to adaptively select the suitable AAIs for the circle structure based on the density of common friends and the dispersion level of common friends' attention. In addition, for the sake of further improving the accuracy of the selected AAI, its complementary AAIs are identified with support vector machine model according to their similarity in value, direction, and ranking. Finally, the mutually complementary indices are combined into a composite one to comprehensively portray the attention distribution of common friends of users in different circles and predict their possible friendships for cross-marketing activities. Experimental results on Twitter and Google+ show that the model has highly reliable prediction performance [5].

7.1.9. Revisiting the medial axis for planar shape decomposition

Participants: Nikos Papanelopoulos [NTUA, Greece], Yannis Avrithis, Stefanos Kollias [U. of Lincoln, UK].

We present a simple computational model for planar shape decomposition that naturally captures most of the rules and salience measures suggested by psychophysical studies, including the minima and short-cut rules, convexity, and symmetry. It is based on a medial axis representation in ways that have not been explored before and sheds more light into the connection between existing rules like minima and convexity. In particular, vertices of the exterior medial axis directly provide the position and extent of negative minima of curvature, while a traversal of the interior medial axis directly provides a small set of candidate endpoints for part-cuts. The final selection follows a prioritized processing of candidate part-cuts according to a local convexity rule that can incorporate arbitrary salience measures. Neither global optimization nor differentiation is involved. We provide qualitative and quantitative evaluation and comparisons on ground-truth data from psycho-physical experiments. With our single computational model, we outperform even an ensemble method on several other competing models [6].

7.1.10. Graph-based Particular Object Discovery

Participants: Oriane Siméoni, Ahmet Iscen [Univ. Prague], Giorgos Tolias [Univ. Prague], Yannis Avrithis, Ondra Chum [Univ. Prague].

Severe background clutter is challenging in many computer vision tasks, including large-scale image retrieval. Global descriptors, that are popular due to their memory and search efficiency, are especially prone to corruption by such a clutter. Eliminating the impact of the clutter on the image descriptor increases the chance of retrieving relevant images and prevents topic drift due to actually retrieving the clutter in the case of query expansion. In this work, we propose a novel salient region detection method. It captures, in an unsupervised manner, patterns that are both discriminative and common in the dataset. Saliency is based on a centrality measure of a nearest neighbor graph constructed from regional CNN representations of dataset images. The proposed method exploits recent CNN architectures trained for object retrieval to construct the image representation from the salient regions. We improve particular object retrieval on challenging datasets containing small objects [7].

7.1.11. Label Propagation for Deep Semi-supervised Learning

Participants: Ahmet Iscen [Univ. Prague], Giorgos Tolias [Univ. Prague], Yannis Avrithis, Ondra Chum [Univ. Prague].

Semi-supervised learning is becoming increasingly important because it can combine data carefully labeled by humans with abundant unlabeled data to train deep neural networks. Classic methods on semi-supervised learning that have focused on transductive learning have not been fully exploited in the inductive framework followed by modern deep learning. The same holds for the manifold assumption—that similar examples should get the same prediction. In this work, we employ a transductive label propagation method that is based on the manifold assumption to make predictions on the entire dataset and use these predictions to generate pseudo-labels for the unlabeled data and train a deep neural network. At the core of the transductive method lies a nearest neighbor graph of the dataset that we create based on the embeddings of the same network. Therefore our learning process iterates between these two steps. We improve performance on several datasets especially in the few labels regime and show that our work is complementary to current state of the art [12], [38].

7.1.12. Dense Classification and Implanting for Few-Shot Learning

Participants: Yann Lifchitz, Yannis Avrithis, Sylvaine Picard [SAFRAN Group], Andrei Bursuc [Valéo].

Few-shot learning for deep neural networks is a highly challenging and key problem in many computer vision tasks. In this context, we are targeting knowledge transfer from a set with abundant data to other sets with few available examples. We propose in [14], [40] two simple and effective solutions: (i) dense classification over feature maps, which for the first time studies local activations in the domain of few- shot learning, and (ii) implanting, that is, attaching new neurons to a previously trained network to learn new, task-specific features. Implanting enables training of multiple layers in the few-shot regime, departing from most related methods derived from metric learning that train only the final layer. Both contributions show consistent gains when used individually or jointly and we report state of the art performance on few-shot classification on miniImageNet.

7.1.13. Point in, Box out: Beyond Counting Persons in Crowds

Participants: Yuting Liu [Sichuan University, China], Miaojing Shi, Qijun Zhao [Sichuan University, China], Xiaofang Wang [RAINBOW Team, IRISA].

Modern crowd counting methods usually employ deep neural networks (DNN) to estimate crowd counts via density regression. Despite their significant improvements, the regression-based methods are incapable of providing the detection of individuals in crowds. The detection-based methods, on the other hand, have not been largely explored in recent trends of crowd counting due to the needs for ex- pensive bounding box annotations. In this work, we instead propose a new deep detection network with only point supervision required [15]. It can simultaneously detect the size and location of human heads and count them in crowds. We first mine useful person size information from point-level annotations and initialize the pseudo ground truth bounding boxes. An online updating scheme is introduced to refine the pseudo ground truth during training; while a locally- constrained regression loss is designed to provide additional constraints on the size of the predicted boxes in a local neighborhood. In the end, we propose a curriculum learning strategy to train the network from images of relatively accurate and easy pseudo ground truth first. Extensive experiments are conducted in both detection and counting tasks on several standard benchmarks, e.g. ShanghaiTech, UCF CC 50, WiderFace, and TRANCOS datasets, and the results show the superiority of our method over the state-of-the-art.

7.1.14. Revisiting Perspective Information for Efficient Crowd Counting

Participants: Miaojing Shi, Zhaohui Yang [Peking University, China], Chao Xu [Peking University, China], Qijun Chen [Tongji University, China].

Crowd counting is the task of estimating people numbers in crowd images. Modern crowd counting methods employ deep neural networks to estimate crowd counts via crowd density regressions. A major challenge of this task lies in the perspective distortion, which results in drastic person scale change in an image. Density regression on the small person area is in general very hard. In this work, we propose a perspective-aware convolutional neural network (PACNN) for efficient crowd counting, which integrates the perspective information into density regression to provide additional knowledge of the person scale change in an image [18]. Ground truth perspective maps are firstly generated for training; PACNN is then specifically designed to predict multi-scale perspective maps, and encode them as perspective-aware weighting layers

in the network to adaptively combine the outputs of multi-scale density maps. The weights are learned at every pixel of the maps such that the final density combination is robust to the perspective distortion. We conduct extensive experiments on the ShanghaiTech, WorldExpo'10, UCF CC 50, and UCSD datasets, and demonstrate the effectiveness and efficiency of PACNN over the state-of-the-art.

7.1.15. Local Features and Visual Words Emerge in Activations

Participants: Oriane Siméoni, Yannis Avrithis, Ondra Chum [Univ. Prague].

We propose a novel method of deep spatial matching (DSM) for image retrieval [19], [41]. Initial ranking is based on image descriptors extracted from convolutional neural network activations by global pooling, as in recent state-of- the-art work. However, the same sparse 3D activation tensor is also approximated by a collection of local features. These local features are then robustly matched to approximate the optimal alignment of the tensors. This hap- pens without any network modification, additional layers or training. No local feature detection happens on the original image. No local feature descriptors and no visual vocabulary are needed throughout the whole process. We experimentally show that the proposed method achieves the state-of-the-art performance on standard benchmarks across different network architectures and different global pooling methods. The highest gain in performance is achieved when diffusion on the nearest-neighbor graph of global descriptors is initiated from spatially verified images.

7.1.16. Combining convolutional side-outputs for road image segmentation

Participants: Raquel Almeida, Simon Malinowski, Ewa Kijak, Silvio Guimaraes [PUC Minas].

Image segmentation consists in creating partitions within an image into meaningful areas and objects. It can be used in scene understanding and recognition, in fields like biology, medicine, robotics, satellite imaging, amongst others. In this work [17], we take advantage of the learned model in a deep architecture, by extracting side-outputs at different layers of the network for the task of image segmentation. We study the impact of the amount of side-outputs and evaluate strategies to combine them. A post-processing filtering based on mathematical morphology idempotent functions is also used in order to remove some undesirable noises. Experiments were performed on the publicly available KITTI Road Dataset for image segmentation. Our comparison shows that the use of multiples side outputs can increase the overall performance of the network, making it easier to train and more stable when compared with a single output in the end of the network. Also, for a small number of training epochs (500), we achieved a competitive performance when compared to the best algorithm in KITTI Evaluation Server.

7.1.17. BRIEF-based mid-level representations for time series classification

Participants: Raquel Almeida, Simon Malinowski, Silvio Guimaraes [PUC Minas].

Time series classification has been widely explored over the last years. Amongst the best approaches for that task, many are based on the Bag-of-Words framework, in which time series are transformed into a histogram of word occurrences. These words represent quantized features that are extracted beforehand. In this work [20], we aim to evaluate the use of accurate mid-level representation called BossaNova in order to enhance the Bag-of-Words representation and to propose a new binary time series descriptor, called BRIEF-based descriptor. More precisely, this kind of representation enables to reduce the loss induced by feature quantization. Experiments show that this representation in conjunction to BRIEF-based descriptor is statistically equivalent to traditional Bag-of-Words, in terms time series classification accuracy, being about 4 times faster. Furthermore, it is very competitive when compared to the state-of-the-art.

7.1.18. Toward a Framework for Seasonal Time Series Forecasting Using Clustering

Participants: Simon Malinowski, Thomas Guyet [LACODAM Team], Colin Leverger [LACODAM Team], Alexandre Termier [LACODAM Team].

Seasonal behaviours are widely encountered in various applications. For instance, requests on web servers are highly influenced by our daily activities. Seasonal forecasting consists in forecasting the whole next season for a given seasonal time series. It may help a service provider to provision correctly the potentially required resources, avoiding critical situations of over- or under provision. In this article, we propose a generic framework to make seasonal time series forecasting. The framework combines machine learning techniques (1) to identify the typical seasons and (2) to forecast the likelihood of having a season type in one season ahead. We study in[13] this framework by comparing the mean squared errors of forecasts for various settings and various datasets. The best setting is then compared to state-of-the-art time series forecasting methods. We show that it is competitive with them.

7.1.19. Smooth Adversarial Examples

Participants: Hanwei Zhang, Yannis Avrithis, Teddy Furon, Laurent Amsaleg.

This paper investigates the visual quality of the adversarial examples. Recent papers propose to smooth the perturbations to get rid of high frequency artefacts. In this work, smoothing has a different meaning as it perceptually shapes the perturbation according to the visual content of the image to be attacked [44]. The perturbation becomes locally smooth on the flat areas of the input image, but it may be noisy on its textured areas and sharp across its edges. This operation relies on Laplacian smoothing, well-known in graph signal processing, which we integrate in the attack pipeline. We benchmark several attacks with and without smoothing under a white-box scenario and evaluate their transferability. Despite the additional constraint of smoothness, our attack has the same probability of success at lower distortion.

7.1.20. Walking on the Edge: Fast, Low-Distortion Adversarial Examples

Participants: Hanwei Zhang, Yannis Avrithis, Teddy Furon, Laurent Amsaleg.

Adversarial examples of deep neural networks are receiving ever increasing attention because they help in understanding and reducing the sensitivity to their input. This is natural given the increasing applications of deep neural networks in our everyday lives. When white-box attacks are almost always successful, it is typically only the distortion of the perturbations that matters in their evaluation. In this work [45], we argue that speed is important as well, especially when considering that fast attacks are required by adversarial training. Given more time, iterative methods can always find better solutions. We investigate this speed-distortion trade-off in some depth and introduce a new attack called boundary projection (BP) that improves upon existing methods by a large margin. Our key idea is that the classification boundary is a manifold in the image space: we therefore quickly reach the boundary and then optimize distortion on this manifold.

7.1.21. Accessing watermarking information: Error exponents in the noisy case

Participant: Teddy Furon.

The study of the error exponents of zero-bit watermarking is addressed in the article by Comesana, Merhav, and Barni, under the assumption that the detector relies solely on second order joint empirical statistics of the received signal and the watermark. This restriction leads to the well-known dual hypercone detector, whose score function is the absolute value of the normalized correlation. They derive the false negative error exponent and the optimum embedding rule. However, they only focus on high SNR regime, i.e. the noiseless scenario. This work extends this theoretical study to the noisy scenario. It introduces a new definition of watermarking robustness based on the false negative error exponent, derives this quantity for the dual hypercone detector, and shows that its performances is almost equal to Costa's lower bound [22].

7.1.22. Detecting fake news and image forgeries

Participants: Cédric Maigrot, Vincent Claveau, Ewa Kijak.

Social networks make it possible to share information rapidly and massively. Yet, one of their major drawback comes from the absence of verification of the piece of information, especially with viral messages. Based on the work already presented in the previous years, C. Maigrot defended his thesis on the detection of image forgeries, classification of reinformation websites, and on the late fusion of models based on the text, image and source analysis [1]. This work was also given a large visibility thanks to numerous interviews in Press and TV (see the dedicated section about popularization).

7.1.23. Learning Interpretable Shapelets for Time Series Classification through Adversarial Regularization

Times series classification can be successfully tackled by jointly learning a shapelet-based representation of the series in the dataset and classifying the series according to this representation. However, although the learned shapelets are discriminative, they are not always similar to pieces of a real series in the dataset. This makes it difficult to interpret the decision, i.e. difficult to analyze if there are particular behaviors in a series that triggered the decision. In this work [29], we make use of a simple convolutional network to tackle the time series classification task and we introduce an adversarial regularization to constrain the model to learn more interpretable shapelets. Our classification results on all the usual time series benchmarks are comparable with the results obtained by similar state-of-the-art algorithms but our adversarially regularized method learns shapelets that are, by design, interpretable.

7.1.24. Using Knowledge Base Semantics in Context-Aware Entity Linking

Participants: Cheikh Brahim El Vaigh, Guillaume Gravier, Pascale Sébillot.

Done as part of the IPL iCODA, in collaboration with CEDAR Inria team.

Entity linking is a core task in textual document processing, which consists in identifying the entities of a knowledge base (KB) that are mentioned in a text. Approaches in the literature consider either independent linking of individual mentions or collective linking of all mentions. Regardless of this distinction, most approaches rely on the Wikipedia encyclopedic KB in order to improve the linking quality, by exploiting its entity descriptions (web pages) or its entity interconnections (hyperlink graph of web pages). We devised a novel collective linking technique which departs from most approaches in the literature by relying on a structured RDF KB [9]. This allows exploiting the semantics of the interrelationships that candidate entities may have at disambiguation time rather than relying on raw structural approximation based on Wikipedia's hyperlink, graph. The few approaches that also use an RDF KB simply rely on the existence of a relation between the candidate entities to which mentions may be linked. Instead, we weight such relations based on the RDF KB structure and propose an efficient decoding strategy for collective linking. Experiments on standard benchmarks show significant improvement over the state of the art.

7.1.25. Neural-based lexico-syntactic relation extraction in news archives

Participants: Guillaume Gravier, Cyrielle Mallart, Pascale Sébillot.

Done as part of the IPL iCODA, in collaboration with Ouest France

Relation extraction is the task of finding and classifying the relationship between two entities in a text. We pursued work on the detection of relations between entities, seen as a binary classification problem. In the context of large-scale news archives, we argue that detection is paramount before even considering classification, where most approaches consider the two tasks jointly with a null garbage class. This does hardly allow for the detection of relations for unseen categories, which are all considered as garbage. We designed a bi-LSTM sequence neural model acting on features extracted from the surface realization, the part-of-speech tags and the dependency parse tree and compared with a state-of-the-art relation detection LSTM-based approach. Experimental evaluations rely on a dataset derived from 200k Wikipedia articles in French containing 4M linked mentions of entities: 330k pairs of entities co-occur in the same sentence, of which 1 % are actual relations according to Wikidata. Results show the benefit of our binary detection approach over previous methods and over joint detection and classification.

7.1.26. Graph Convolutional Networks for Learning with Few Clean and Many Noisy Labels

Participants: Ahmet Iscen [Google Research], Giorgos Tolias [Univ. Prague], Yannis Avrithis, Ondra Chum [Univ. Prague], Cordelia Schmid [Google Research].

In this work we consider the problem of learning a classifier from noisy labels when a few clean labeled examples are given [39]. The structure of clean and noisy data is modeled by a graph per class and Graph Convolutional Networks (GCN) are used to predict class relevance of noisy examples. For each class, the GCN is treated as a binary classifier learning to discriminate clean from noisy examples using a weighted binary cross-entropy loss function, and then the GCN-inferred "clean" probability is exploited as a relevance measure. Each noisy example is weighted by its relevance when learning a classifier for the end task. We evaluate our method on an extended version of a few-shot learning problem, where the few clean examples of novel classes are supplemented with additional noisy data. Experimental results show that our GCN-based cleaning process significantly improves the classification accuracy over not cleaning the noisy data and standard few-shot classification where only few clean examples are used. The proposed GCN-based method outperforms the transductive approach (Douze et al., 2018) that is using the same additional data without labels.

7.1.27. Rethinking deep active learning: Using unlabeled data at model training

Participants: Oriane Siméoni, Mateusz Budnik, Yannis Avrithis, Guillaume Gravier.

Active learning typically focuses on training a model on few labeled examples alone, while unlabeled ones are only used for acquisition. In this work we depart from this setting by using both labeled and unlabeled data during model training across active learning cycles [42]. We do so by using unsupervised feature learning at the beginning of the active learning pipeline and semi-supervised learning at every active learning cycle, on all available data. The former has not been investigated before in active learning, while the study of latter in the context of deep learning is scarce and recent findings are not conclusive with respect to its benefit. Our idea is orthogonal to acquisition strategies by using more data, much like ensemble methods use more models. By systematically evaluating on a number of popular acquisition strategies and datasets, we find that the use of unlabeled data during model training brings a spectacular accuracy improvement in image classification, compared to the differences between acquisition strategies. We thus explore smaller label budgets, even one label per class.

7.1.28. Training Object Detectors from Few Weakly-Labeled and Many Unlabeled Images

Participants: Zhaohui Yang [Peking University], Miaojing Shi, Yannis Avrithis, Chao Xu [Peking University], Vittorio Ferrari [Google Research].

Weakly-supervised object detection attempts to limit the amount of supervision by dispensing the need for bounding boxes, but still assumes image-level labels on the entire training set are available. In this work, we study the problem of training an object detector from one or few clean images with image-level labels and a larger set of completely unlabeled images [43]. This is an extreme case of semi-supervised learning where the labeled data are not enough to bootstrap the learning of a classifier or detector. Our solution is to use a standard weakly-supervised pipeline to train a student model from image-level pseudo-labels generated on the unlabeled set by a teacher model, bootstrapped by region-level similarities to clean labeled images. By using the recent pipeline of PCL and more unlabeled images, we achieve performance competitive or superior to many state of the art weakly-supervised detection solutions.

7.2. Accessing Information

7.2.1. Ontological modeling of human reading experience

Participants: Guillaume Gravier, Pascale Sébillot.

Done as part of the JPI CH READ-IT projects, in collaboration with Open University (UK) and Univeersité Le Mans (FR)

Diaries, correspondence and authors' libraries provide important evidence into the evolution of ideas and society. Studying these phenomena is connected to understanding changes of perspective and values. Within the framework of the READ-IT project, we developed an ontological data approach modelling changes in the contents of diaries, correspondence and authors' libraries related to reading. By considering these three types of sources, we designed a conceptual data model to permit the study and increase the usability of sources containing evidence of reading experiences, highlighting common challenges and patterns related to changes to readers and to the medium of reading when confronting historical events [36], [8].

7.2.2. Integration of Exploration and Search: A Case Study of the M^3 Model

Participants: Snorri Gíslason [IT Univ. Copenhagen], Björn Þór Jónsson [IT Univ. Copenhagen], Laurent Amsaleg.

Effective support for multimedia analytics applications requires exploration and search to be integrated seamlessly into a single interaction model. Media metadata can be seen as defining a multidimensional media space, casting multimedia analytics tasks as exploration, manipulation and augmentation of that space. We present an initial case study of integrating exploration and search within this multidimensional media space [11]. We extend the M^3 model, initially proposed as a pure exploration tool, and show that it can be elegantly extended to allow searching within an exploration context and exploring within a search context. We then evaluate the suitability of relational database management systems, as representatives of today's data management technologies, for implementing the extended M^3 model. Based on our results, we finally propose some research directions for scalability of multimedia analytics.

7.2.3. Exquisitor: Breaking the Interaction Barrier for Exploration of 100 Million Images

Participants: Hanna Ragnarsdóttir [Reykjavik University], Þórhildur Þorleiksdóttir [Reykjavik University], Omar Shahbaz Khan [IT Univ. Copenhagen], Björn Þór Jónsson [IT Univ. Copenhagen], Gylfi Þór Gudmundsson [School of Computer Science, Reykjavik], Jan Zahálka [bohem.ai], Stevan Rudinac [University of Amsterdam], Laurent Amsaleg, Marcel Worring [University of Amsterdam].

We present Exquisitor, a media explorer capable of learning user preferences in real-time during interactions with the 99.2 million images of YFCC100M. Exquisitor owes its efficiency to innovations in data representation, compression, and indexing. Exquisitor can complete each interaction round, including learning preferences and presenting the most relevant results, in less than 30 ms using only a single CPU core and modest RAM. In short, Exquisitor can bring large-scale interactive learning to standard desktops and laptops, and even high-end mobile devices [16].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. CIFRE PhD: Incremental dynamic construction of knowledge bases from text mining

Participants: Guillaume Gravier, Cyrielle Mallart, Pascale Sébillot.

Duration: 3 years, started in Dec. 2018
Partner: Ouest France

In the context of a newspaper, the thesis explores the combination of text mining and knowledge representation techniques to assist the extraction, interpretation and validation of valuable pieces of information from the journal's content so as to incrementally build a full-scale knowledge base. This thesis is in close relation with the iCODA Inria Project Lab, with direct contribution to the project's results.

8.1.2. CIFRE PhD: Embedding heterogeneous data for directory search

Participants: Vincent Claveau, Guillaume Gravier, François Torregrossa.

Duration: 3 years, started in Dec. 2018
Partner: SoLocal

The thesis aims at learning how to jointly exploit heterogeneous sources of information (e.g., names, activity sector, user profiles, queries, etc.) in the design of neural network embeddings for information retrieval and language understanding. Applications cover natural language query analysis and personalized information retrieval in Pagesjaunes' directory.

8.1.3. CIFRE PhD: Few shot learning for object recognition in aerial images

Participants: Yannis Avrithis, Yann Lifchitz.

Duration: 3 years, started in March 2018
Partner: Safran Tech

This is a CIFRE PhD thesis project aiming to study architectures and learning techniques most suitable for object recognition from few samples and to validate these approaches on multiple recognition tasks and use-cases related to aerial images.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. Computer vision for smart phones (MobilAI)

Participants: Yannis Avrithis, Mateusz Budnik.

Duration: 2 years, started in September 2018
Partners: Lamark, Quai des Apps, AriadNext

The ability of our mobile devices to process visual information is currently not limited by their camera or computing power but by the network. Many mobile apps suffer from long latency due to data transmitted over the network for visual search. MobilAI aims to provide fast visual recognition on mobile devices, offering quality user experience whatever the network conditions. The idea is to transfer efficient deep learning solutions for image classification and retrieval onto embedded platforms such as smart phones. The intention is to use such solutions in B2B and B2C application contexts, for instance recognizing products and ordering online, accessing information about artifacts in exhibitions, or identifying identity documents. In all cases, visual recognition is performed on the device, with minimal or no access to the network.

9.1.2. CominLabs Project BigCLIN

Participants: Vincent Claveau, Ewa Kijak, Clément Dalloux.

Duration: 3 years, started in September 2016
Partners: STL-CNRS, Inserm/CHU Rennes, Inria
URL: https://bigclin.cominlabs.u-bretagneloire.fr/fr

Data collected or produced during clinical care process can be exploited at different levels and across different domains. Yet, a well-known challenge for secondary use of health big data is that much of detailed patient information is embedded in narrative text, mostly stored as unstructured data. The project proposes to address the essential needs when reusing unstructured clinical data at a large scale. We propose to develop new clinical records representation relying on fine-grained semantic annotation thanks to new NLP tools dedicated to French clinical narratives. To efficiently map this added semantic information to existing structured data for further analysis at big scale, the project also addresses distributed systems issues: scalability, management of uncertain data and privacy, stream processing at runtime, etc.

9.2. National Initiatives

9.2.1. Inria Project Lab Knowledge-driven data and content collaborative analytics (iCODA)

Participants: Laurent Amsaleg, Cheikh Brahim El Vaigh, Guillaume Gravier, Cyrielle Mallart, Pascale Sébillot.

Duration: 4.5 years, started in April 2017
Partners: Inria project-teams Linkmedia, CEDAR, GraphIK and ILDA, with Ouest-France, Le Monde and AFP

One of today's major issues in data science is the design of algorithms that allow analysts to efficiently infer useful information and knowledge by collaboratively inspecting heterogeneous information sources, from structured data to unstructured content. Taking data journalism as an emblematic use-case, the goal of the project is to develop the scientific and technological foundations for knowledge- mediated user-in-the-loop collaborative data analytics on heterogeneous information sources, and to demonstrate the effectiveness of the approach in realistic, high-visibility use-cases. The project stands at the crossroad of multiple research fields—content analysis, data management, knowledge representation, visualization—that span multiple Inria themes, and counts on a club of major press partners to define usage scenarios, provide data and demonstrate achievements.

9.2.2. Inria-BNF: Classification d'images patrimoniales (CIP)

Participants: Florent Michel, Laurent Amsaleg, Guillaume Gravier, Ewa Kijak, Yannis Avrithis.

Duration: 1 year, started in Dec 2018

This project is within the context of the collaborations between Inria and the French Ministry of Culture. In that context, we have started a collaboration with the French National Library (BNF) which collects, preserves and makes known the national documentary heritage. This collaboration aims at facilitating the automatic classification of heritage images through the use of recent deep-learning techniques. Such images are quite specific: they are not at all similar with what deep-learning techniques are used to work with, that is, the classification of heritage images does not target modern categories such as planes, cars, cats and dogs because this is irrelevant and because heritage collections do not include images of contemporary objects. Furthermore, heritage images come in vast quantities, but they are little annotated and deep-learning techniques can hardly rely on massive annotations to easily learn. Last, the learning has to be continuous as curators may need to add or modify existing classes, without re-learning everything from scratch.

The techniques of choice to reach that goal include the semi-supervised learning, low-shot learning techniques, knowledge transfer, fine tuning existing models, etc.

9.2.3. ANR Archival: Multimodal machine comprehension of language for new intelligent interfaces of scientific and cultural mediation

Participants: Laurent Amsaleg, Guillaume Gravier, Pascale Sébillot.

Duration: 3.5 year, started in Dec. 2019

The multidisciplinary and multi-actor ARCHIVAL project aims at yielding collaborations between researchers from the fields of Information and Communication Sciences as well as Computer Sciences around archive value enhancing and knowledge sharing for arts, culture and heritage. The project is structured around the following questionings: What part can machine comprehension methods play towards the reinterpretation of thematic archive collections? How can content mediation interfaces exploit results generated by current AI approaches?

ARCHIVAL teams will explore heterogeneous document collection structuration in order to explicitly reveal implicit links, to explain the nature of these links and to promote them in an intelligible way towards ergonomic mediation interfaces that will guarantee a successful appropriation of contents. A corpus has been delimited from the FMSH "self-management" collection, recently awarded as Collex, which will be completed from the large Canal-U academic audiovisual portal. The analysis and enhancement of this collection is of particular interest for Humanities and Social Sciences in a context where it becomes a necessity to structurally reconsider new models of socioeconomic development (democratic autonomy, social and solidarity-based economy, alternative development,...).

9.3. European Initiatives

9.3.1. EIT Digital CREEP2

Program: EIT Digital

Project acronym: CREEP 2

Project title: Cyberbullying effects prevention

Duration: 01/2019 - 12/2019

Coordinator: FBK, Italy

Other partners: Expert Systems (IT), Inria (FR), Engineering (IT)

Abstract: Project CREEP (Cyberbulling Effects Prevention) aims at identifying and preventing the possible negative impacts of cyberbullying on young people. It seeks to realise advanced technologies for the early detection of cyberbullying phenomena through the monitoring of social media and the communication of preventive advices and personalized recommendations tailored to adolescents' needs through a virtual coaching system (chatbot).

9.3.2. JPI CH READ-IT

Program: Joint Programming Initiative on Cultural Heritage

Project acronym: READ-IT

Project title: Reading Europe Advanced Data Investigation Tool

Duration: 05/2018 - 04/2021

Coordinator: Université Le Mans (FR)

Other partners: CNRS-IRISA (FR), Open University (UK), Universiteit Utrecht (NL), Institute of Czech Litterature (CZ)

Abstract: READ-IT is a transnational, interdisciplinary R&D project that will build a unique large-scale, user- friendly, open access, semantically-enriched investigation tool to identify and share groundbreaking evidence about 18th-21st century Cultural Heritage of reading in Europe. READ-IT will ensure the sustainable and reusable aggregation of qualitative data allowing an in-depth analysis of the Cultural Heritage of reading. State-of-the art technology in Semantic Web and information systems will provide a versatile, end-users oriented environment enabling scholars and ordinary readers to retrieve information from a vast amount of community-generated digital data leading to new understanding about the circumstances and effects of reading in Europe.

9.3.3. CHIST-ERA ID_IOT

Program: CHIST ERA

Project acronym: ID_IOT

Project title: Identification for the Internet of things

Duration: 3 years, started in Oct 2016.

Coordinator: Boris Skoric (Eindhoven Univ. of Technology (NL))

Other partners: Inria-RBA (Teddy Furon, Marzieh Gheisari Khorasgani), Univ. of Geneva (CH)

Abstract: The IoT will contain a huge number of devices and objects that have very low or nonexistent processing and communication resources, coupled to a small number of high-power devices. The weakest devices, which are most ubiquitous, will not be able to authenticate themselves using cryptographic methods. This project addresses these issues using physical unclonable functions (PUFs). PUFs, and especially quantum readout PUFs, are ideally suited to the IoT setting because they allow for the authentication and identification of physical objects without requiring any crypto or storage of secret information.

Furthermore, we foresee that back-end systems will not be able to provide security and privacy via cryptographic primitives due to the sheer number of IoT devices. Our plan is to address these problems using privacy preserving database structures and algorithms with good scaling behaviour. Approximate nearest neighbour (ANN) search algorithms, which have remarkably good scaling behaviour, have recently become highly efficient, but do not yet have the right security properties and have not yet been applied to PUF data. Summarised in a nutshell, the project aims to improve the theory and practice of technologies such as PUFs and ANN search in the context of generic IoT authentication and identification scenarios.

9.3.4. Collaborations with Major European Organizations

Program: ConFAP-CNRS Project

Project acronym: FIGTEM

Project title: FIne-Grain TExt Mining for clinical data

Duration: 01/2016 - 05/2019

Coordinator: CNRS-IRISA

Other partners: PUCPR, Curitiba, Brasil; CNRS-STL Lille; Inserm LTSI/CHU Rennes

Abstract: FIGTEM is a research project that involves STL-CNRS, CHU Rennes, PUC Parana, Curitiba and led by LINKMEDIAThis project aimed at developing natural language processing methods, including information extraction and indexing, dedicated to the clinical trial domain. The goal was to populate a formal representation of patients (via their electronic patient records) and clinical trial data in different languages (French, English, Portuguese). The main outcomes of the project was NLP tools for these 3 languages and annotated datasets made available for research purposes. It ended in May 2019.

9.4. International Initiatives

9.4.1. Inria International Partners

9.4.1.1. Informal International Partners

- Michael Houle, NII, Japan
- Marcel Worring, UvA, Netherlands
- Martha Larson, Radboud U., Netherlands

9.4.2. Participation in Other International Programs

9.4.2.1. STIC AmSud TRANSFORM

Program: STIC AmSud

Project acronym: TRANSFORM

Project title: Transforming multimedia data for indexing and retrieval purposes

Duration: 01/2018 - 31/2019

Partners: CNRS-IRISA (FR), PUC Minas (BR), UChile (CL)

9.4.2.2. CAPES/COFECUB HIMMD

Program: CAPES/COFECUB

Project acronym: HIMMD

Project title: Hierarchical graph-based analysis of image, video, and multimedia data

Duration: 01/2019 - 31/2022

Partners: LIGM (FR), IRISA (FR), INPG (FR), PUC Minas (BR), UNICamp (BR), UFMG (BR)

9.5. International Research Visitors

9.5.1. Visits of International Scientists

- Silvio Guimaraes (PUC Minas, Brazil) visited the team in July (1 week). His visit was related to the Stic-Amsud project.
- Benjamin Bustos (Univ. Chile, Chile) visited the team in July (1 week). His visit was related to the Stic-Amsud project.

9.5.2. Visits to International Teams

9.5.2.1. Research Stays Abroad

- Yannis Avrithis, National and Kapodistrian University of Athens, 3 visits on April (1 week), June (1 week) and September-October 2019 (3 weeks).

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events: Organisation

10.1.1.1. General Chair, Scientific Chair

- Laurent Amsaleg is general co-chair of ACM Intl. Conf. on Multimedia 2019.
- Guillaume Gravier is technical program co-chair for ACM Intl. Conf. on Multimedia 2019.

10.1.1.2. Member of the Organizing Committees

- Yannis Avrithis is workshops co-chair for ACM Intl. Conf. on Multimedia 2019.
- Simon Malinowski is member of the organizing commitee of the Advanced Analytics and Learning on Temporal Data workshop (Wurzburg, Sept. 2019)
- Vincent Claveau was Finance chair of SIGIR 2019, (Paris, July 2019)

10.1.2. Scientific Events: Selection

10.1.2.1. Chair of Conference Program Committees

- Vincent Claveau was co-chair of the text-mining workshop DeFT in Toulouse, July 2019

10.1.2.2. Member of the Conference Program Committees

- Laurent Amsaleg was a PC member of: Extraction et Gestion de Connaissances, ACM International Conference on Multimedia Retrieval, Multimedia Modeling, Content-Based Multimedia Indexing, IEEE International Conference on Multimedia & Expo, International Conference on Similarity Search and Applications, Multimedia Information Retrieval and Applications.
- Guillaume Gravier was PC member of: European Conf. on Information Retrieval, Intl. Symposium on Multimedia, ACM Intl. Conf. on Multimedia Retrieval, IEEE Workshop on Multimedia Modeling.
- Pascale Sébillot was a PC member of: Annual Meeting of the Association for Computational Linguistics (ACL), International Joint Conference on Artificial Intelligence on Artificial Intelligence (IJCAI), European Conference on Information Retrieval (ECIR), International Conference on MultiMedia Modeling (MMM), Terminologie et intelligence artificielle (TIA) workshop with TALN.
- Yannis Avrithis was PC member of: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), International Conference on Computer Vision (ICCV). He has been selected as Outstanding Reviewer of ICCV (91 selected out of 2506 reviewers in total).
- Vincent Claveau was a PC member of: European Conference on Information Retrieval (ECIR), EMNLP (demo), ACL (demo), TexMine workshop with EGC, DeFT workshop with TALN, TalMed workshop with MEDINFO, CORIA.
- Simon Malinowski is member of the program commitee of the Advanced Analytics and Learning on Temporal Data workshop (Wurzburg, Sept. 2019)
- Ewa Kijak was a PC member of: ACM Intl. Conf. on Multimedia (area chair), Intl. Conf. on Content-Based Multimedia Indexing (CBMI)

- Teddy Furon was a PC member of: IEEE Workshop on Information Forensics and Security (WIFS), Int. Conf. on Similarity Search and Applications (SISAP), IEEE Int. Conf. on Acoustic Speech and Signal Processing (ICASSP), ACM Information Hiding & Multimedia Security (IHMMSEC).

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- Pascale Sébillot is editor of the Journal Traitement Automatique des Langues.
- Pascale Sébillot is member of the editorial board of the Journal Traitement Automatique des Langues.
- Vincent Claveau is Chief Editor of the ISTE journal RIDoWS
- Vincent Claveau is member of the editorial board of the journal Traitement Automatique des Langues (TAL)

10.1.3.2. Reviewer - Reviewing Activities

- Pascale Sébillot reviewed for Traitement Automatique des Langues.
- Vincent Claveau was a reviewer for: Multimedia Tools and Applications, Technique et Sciences Informatiques, Documents numériques, Traitement Automatique des Langues, RIDoWS
- Ewa Kijak was a reviewer for: Multimedia Tools and Applications, Remote Sensing journal
- Laurent Amsaleg was a reviewer for: IEEE Transactions on Information Forensics and Security, ACM Transactions on Information Systems.
- Pascale Sébillot was a reviewer for: Conference Traitement Automatique des Langues Naturelles (TALN).
- Simon Malinowski was a reviewer for the Data Mining and Knowledge Discovery (DAMI) journal, Springer.
- Teddy Furon was a reviewer for: Entropy (MDPI), IEEE Trans. on Dependable and Secure Computing, IEEE Trans. on Information Forensics and Security, IEEE Trans. on Pattern Analysis and Machine Intelligence, IEEE Trans. on Image Processing.

10.1.4. Invited Talks

- Vincent Claveau gave an invited talk at "La vie sociale du fake sur les espaces numériques", Tours, France, nov 2019.
- Ewa Kijak gave a talk on "Tampering detection in images to fight fake news" for the GFAIH (Global Forum on Artificial Intelligence for Humanity) in Paris, during the workshop "AI and information disorder".

10.1.5. Leadership within the Scientific Community

- Laurent Amsaleg is a member of the Steering Committee of SISAP for the 2016-2020 term
- Laurent Amsaleg is a member of the Steering Committee of ACM Multimedia for the 2020-2023 term
- Guillaume Gravier is a member of the scientific board of the pre-GDR Traitement Automatique des Langues.
- Pascale Sébillot is a member of the permanent steering committee of Conf. francophone Traitement Automatique des Langues Naturelles.
- Pascale Sébillot is a member of the board of the pre-GDR Traitement Automatique des Langues, leader of the Intermodality and Multimodality working group.
- Vincent Claveau is deputy head of the GdR CNRS MaDICS https://www.madics.fr
- Vincent Claveau is the finance head of ARIA https://www.asso-aria.org

10.1.6. Scientific Expertise

- Laurent Amsaleg has been involved in the writing of the "Rapport de la mission sur les outils de reconnaissance des contenus protégés et les plateformes de partage en ligne : état de l'art et propositions. Conseil supérieur de la propriété littéraire et artistique, Hadopi".
- Guillaume Gravier was auditioned by the EC for the report "Towards European Media Sovereignty: An industrial media strategy to leverage data, algorithm and artificial intelligence" by Guillaume Klossa, special advisor to EC VP Andrus Ansip.
- Guillaume Gravier was appointed by Technical University of Delft as an evaluator for a professorship position.
- Teddy Furon is scientific advisor for the startup Lamark.

10.1.7. Research Administration

- Guillaume Gravier is deputy director of the Institut de Rercherche en Informatique et Systèmes Aléatoires (IRISA, UMR 6074).
- Guillaume Gravier is a member of the Board of the technology cluster Images & Réseaux.
- Pascale Sébillot was a member of the Conseil National des Universités 27th section (computer science) until Nov 2019.
- Pascale Sébillot is the director of the Computer Science Laboratory, INSA Rennes.
- Pascale Sébillot is the deputy director of the Scientific Advisory Committee of IRISA UMR 6074.
- Pascale Sébillot is a member of the theses advisory committee of the MathSTIC doctoral school.
- Pascale Sébillot is a member of the board of the MathSTIC doctoral school.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Licence: Laurent Amsaleg & Teddy Furon, Indexation multimédia, 10h, L3-cours invité, ENS Rennes, France

Licence: Laurent Amsaleg, Bases de données avancées, 2h, L3-option génie mathématique, INSA Rennes, France

Licence: Guillaume Gravier, Databases, 30h, L2, INSA Rennes, France

Licence: Guillaume Gravier, Probability and statistics, 16h, L3, INSA Rennes, France

Licence: Guillaume Gravier, Natural Language Processing, 12h, L3 & M1, INSA Rennes, France

Licence: Pascale Sébillot, Natural Language Processing, 10h, L3, INSA Rennes, France

Master: Laurent Amsaleg, Bases de données avancées, 25h, M2, INSA Rennes, France

Master: Guillaume Gravier, Data analysis and probabilistic modeling, 30h, M2, University Rennes 1, France

Master: Pascale Sébillot, Natural Language Processing, 6h, M1, INSA Rennes, France

Master: Yannis Avrithis, Deep learning for vision, 20h, M2 SIF, France

Master: Yannis Avrithis, Computer vision, 30h, National and Kapodistrian University of Athens, Greece

Master : Simon Malinowski, Predictive analytics, 48h, M1 MIAGE, Univ. Rennes 1

Master : Simon Malinowski, Machine Learning, 32h, M2 MIAGE, Univ. Rennes 1

Master : Simon Malinowski, Symbolic data mining, 12h, M2 MIAGE, Univ. Rennes 1

Engineering school: Vincent Claveau, Machine Learning, 18h, 3rd year, INSA Rennes, France

Master: Vincent Claveau, Information Retrieval, 10h, M2 MIAGE, Univ. Rennes, France

Master: Ewa Kijak, Image processing, 55h, M1, ESIR, France

Master: Ewa Kijak, Supervised machine learning, 21h, M2R, University Rennes 1, France

Master: Ewa Kijak, Supervised machine learning, 15h, M2, University Rennes 1, France

Master: Ewa Kijak, Image classification, 45h, M1, ESIR, France

Master: Ewa Kijak, Image indexing, 17h, M2, University Rennes 1, France

Master: Ewa Kijak, Computer vision, 22h, M2, ESIR, France

Master: Teddy Furon, Rare Events, 20h, M2, Insa Rennes, France

10.2.2. Supervision

PhD in progress: Mathieu Laroze, Active learning on adaptive representations for object detection in high-resolution imaging, June 2016, Romain Dambreville, Chloe Friguet, Ewa Kijak & Sébastien Lefevre (with OBELIX, IRISA team)

PhD in progress: Hanwei Zhang, Deep Learning in Adversarial Contexts, October 2017, Laurent Amsaleg, Yannis Avrithis, Teddy Furon & Ewa Kijak

PhD in progress: Marzieh Gheisari Khorasgani, Secure identification in the Internet of Things, January 2018, Laurent Amsaleg & Teddy Furon

PhD in progress: Antoine Perquin, Universal speech synthesis through embeddings of massive heterogeneous data, October 2017, Laurent Amsaleg, Gwénolé Lecorvé & Damien Lolive (with Expression, IRISA team)

PhD in progress: Benoit Bonnet, Adversarial images, November 2019, Teddy Furon & Patrick Bas

PhD in progress: Tong Xue, Visualization and collaborative analysis of document collections and extracted knowledge for data journalism, October 2018, Laurent Amsaleg & Anastasia Bezerianos

PhD in progress: Mikail Demirdelen, User-adapted multi-document multimedia synthesis, started October 2016, Guillaume Gravier and Pascale Sébillot

PhD in progress: Cheikh Brahim El Vaigh, Incremental content to data linking leveraging ontological knowledge in data journalism, started October 2017, Guillaume Gravier, Pascale Sébillot and François Goasdoué (with CEDAR, Inria team)

PhD in progress: Cyrielle Mallart, Incremental dynamic construction of knowledge graphs from text mining, started December 2018, Guillaume Gravier, Michel Le Nouy (Ouest-France), Pascale Sébillot

PhD in progress: François Torregrossa, Heterogeneous data embedding for professional search, started November 2018, Robin Allessiardo (So Local), Vincent Claveau, Guillaume Gravier

PhD in progress: Oriane Siméoni, Invariance and supervision in visual learning, started October 2016, Yannis Avrithis & Guillaume Gravier

PhD in progress: Yann Lifchitz, Few shot learning for object recognition in aerial images. Started March 2018, Yannis Avrithis & Sylvaine Picard (Safran Tech).

PhD in progress: Zhaohui Yang, learning visual models with minimal human supervision, started January 2018, Miaojing Shi & Yannis Avrithis & Chao Xu (Peking University, Beijing)

PhD in progress: Raquel Almeida, Learning hierarchichal models for multimedia data, started January 2019, Ewa Kijak & Simon Malinowski & Laurent Amsaleg

PhD defended: Cédric Maigrot, Détection de fausses informations dans les réseaux sociaux, defended April 2019, Laurent Amsaleg, Vincent Claveau & Ewa Kijak. See [1].

PhD defended: Ricardo Carlini-Sperandio, Unsupervised motif mining in multimedia time series, defended December 2019, Laurent Amsaleg.

10.2.3. Juries

- Guillaume Gravier was involved in the following juries:
  - PhD Eloi Zabrocki, Paris-Sorbonne Université, Nov. 2019, reviewer

  - HDR Benoît Favre, Aix-Marseille Université, Nov. 2019, reviewer
- Pascale Sébillot was involved in the following juries:
  - HDR Natalia Grabar, Université Paris-Saclay, May 2019, member
- Vincent Claveau was reviewer for the mid-term PhD auditions of: Ygor Gallina (LS2N), Tsanta Randriatsitohaina (LIMSI), Oussama Ahmia (IRISA), Hyun Jung Kang (MoDyCo)
- Vincent Claveau was a member of the PhD jury of C. Artaud (Univ. La Rochelle)
- Teddy Furon was reviewer for the mid-term PhD auditions of: Alexandre Sablayrolles (Inria Grenoble - Facebook FAIR), Solène Bernard (CRIStAL Lille)
- Teddy Furon was a member of the PhD jury of Clément Feutry (CentraleSupelec Paris).

10.3. Popularization

10.3.1. Articles and contents

- Vincent Claveau was interviewed for his activities about fake news detection in: hors-série 199 de Sciences et avenir Oct/Nov 2019; Sciences Ouest n375; Télérama; Science et vie; Science et avenir; 01net N 894.
- Vincent Claveau was interviewed for his activities about text mining the Grand Débat contributions: AEG infos; Ouest France.
- Ewa Kijak and Vincent Claveau were interviewed for the TV show Télématin Science, nov 2019.
- Ewa Kijak was interviewed about fake news and deepfake detection in: AFP, Le Télégramme, Data Analytics Post, Sciences Ouest n375, Le Temps, FranceTV Info (TV show "Vrai ou Fake"), tv5 monde
- Ewa Kijak was invited on the radio show "La méthode scientifique", France culture, on "Deepfake : faut-il le voir pour le croire ?"
- Teddy Furon was interviewed by Data Analytics Post about the connection of deepfake and watermarking.

10.3.2. Education

LINKMEDIA has been actively involved into the "L codent, L créent" initiative with Cyrielle Mallart and Oriane Siméoni.

10.3.3. Interventions

- Guillaume Gravier gave an invited talk entitled "Strengths, weaknesses and limits of AI illustrated" at the West Data Festival (professional forum, 100 attendees)
- Ewa Kijak gave an invited talk entitled "Fighting Fake news and deep fakes" at the CEPIC congress (Coordination of European Picture Agencies Stock, Press and Heritage)

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] C. MAIGROT. *Detection of misleading information on social networks*, Université de Rennes 1 [UR1], April 2019, https://tel.archives-ouvertes.fr/tel-02404234

Articles in International Peer-Reviewed Journals

[2] L. AMSALEG, M. E. HOULE, E. SCHUBERT. *Introduction to Special Issue of the 9th International Conference on Similarity Search and Applications (SISAP 2016)*, in "Information Systems", February 2019, vol. 80, 107 p. [*DOI :* 10.1016/J.IS.2018.11.006], https://hal.inria.fr/hal-02106914

[3] M. CUI, L. LI, M. SHI. *A Selective Biogeography-Based Optimizer Considering Resource Allocation for Large-Scale Global Optimization*, in "Computational Intelligence and Neuroscience", July 2019, pp. 1-18 [*DOI :* 10.1155/2019/1240162], https://hal.inria.fr/hal-02383076

[4] C. DALLOUX, N. GRABAR, V. CLAVEAU. *Detecting negation: machine learning and a corpus for French*, in "Revue des Sciences et Technologies de l'Information - Série TSI : Technique et Science Informatiques", December 2019, pp. 1-21, https://hal.archives-ouvertes.fr/hal-02402913

[5] S. LI, X. SONG, H. LU, L. ZENG, M. SHI, F. LIU. *Friend recommendation for cross marketing in online brand community based on intelligent attention allocation link prediction algorithm*, in "Expert Systems with Applications", July 2019, vol. 139, pp. 1-11 [*DOI :* 10.1016/J.ESWA.2019.112839], https://hal.inria.fr/hal-02383107

[6] N. PAPANELOPOULOS, Y. AVRITHIS, S. KOLLIAS. *Revisiting the medial axis for planar shape decomposition*, in "Computer Vision and Image Understanding", February 2019, vol. 179, pp. 66-78 [*DOI :* 10.1016/J.CVIU.2018.10.007], https://hal.inria.fr/hal-01930939

[7] O. SIMÉONI, A. ISCEN, G. TOLIAS, Y. AVRITHIS, O. CHUM. *Graph-based Particular Object Discovery*, in "Machine Vision and Applications", March 2019, vol. 30, n^{o} 2, pp. 243-254 [*DOI :* 10.1007/S00138-019-01005-Z], https://hal.inria.fr/hal-02370238

International Conferences with Proceedings

[8] A. ANTONINI, F. BENATTI, E. KING, F. VIGNALE, G. GRAVIER. *Modelling changes in diaries, correspondence and authors' libraries to support research on reading: the READ-IT approach*, in "ODOCH 2019 - First International Workshop on Open Data and Ontologies for Cultural Heritage", Rome, Italy, June 2019, pp. 1-12, https://hal.archives-ouvertes.fr/hal-02130008

[9] C. B. EL VAIGH, F. GOASDOUÉ, G. GRAVIER, P. SÉBILLOT. *Using Knowledge Base Semantics in Context-Aware Entity Linking*, in "DocEng 2019 - 19th ACM Symposium on Document Engineering", Berlin, Germany, ACM, September 2019, pp. 1-10 [*DOI :* 10.1007/978-3-030-27520-4_8], https://hal.inria.fr/hal-02171981

[10] M. GHEISARI, T. FURON, L. AMSALEG. *Privacy Preserving Group Membership Verification and Identification*, in "CVPR 2019 - IEEE Conference on Computer Vision and Pattern Recognition", Long Beach, United States, IEEE, June 2019, pp. 1-9, https://arxiv.org/abs/1904.10327 , https://hal.archives-ouvertes.fr/hal-02107442

[11] S. GÍSLASON, B. Þ. JÓNSSON, L. AMSALEG. *Integration of Exploration and Search: A Case Study of the M^3 Model*, in "MMM 2019 - 25th International Conference on MultiMedia Modeling", Thessaloniki, Greece, LNCS, Springer, December 2019, vol. 11295, pp. 156-168 [*DOI :* 10.1007/978-3-030-05710-7_13], https://hal.inria.fr/hal-02106893

[12] A. ISCEN, G. TOLIAS, Y. AVRITHIS, O. CHUM. *Label Propagation for Deep Semi-supervised Learning*, in "CVPR 2019 - IEEE Computer Vision and Pattern Recognition Conference", Long Beach, United States, IEEE, June 2019, pp. 1-10, https://hal.inria.fr/hal-02370297

[13] C. LEVERGER, S. MALINOWSKI, T. GUYET, V. LEMAIRE, A. BONDU, A. TERMIER. *Toward a Framework for Seasonal Time Series Forecasting Using Clustering*, in "IDEAL 2019", Manchester, United Kingdom, October 2019, pp. 328-340 [*DOI :* 10.1007/978-3-030-33607-3_36], https://hal.inria.fr/hal-02371221

[14] Y. LIFCHITZ, Y. AVRITHIS, S. PICARD, A. BURSUC. *Dense Classification and Implanting for Few-Shot Learning*, in "CVPR 2019 - IEEE Computer Vision and Pattern Recognition Conference", Long Beach, United States, IEEE, June 2019, pp. 1-10, https://hal.inria.fr/hal-02371279

[15] Y. LIU, M. SHI, Q. ZHAO, X. WANG. *Point in, Box out: Beyond Counting Persons in Crowds*, in "CVPR 2019 - IEEE Computer Vision and Pattern Recognition Conference", Long Beach, United States, IEEE, June 2019, pp. 1-10, https://hal.inria.fr/hal-02383057

[16] H. RAGNARSDÓTTIR, Þ. ÞORLEIKSDÓTTIR, O. S. KHAN, B. Þ. JÓNSSON, G. Þ. GUÐMUNDSSON, J. ZAHÁLKA, S. RUDINAC, L. AMSALEG, M. WORRING. *Exquisitor: Breaking the Interaction Barrier for Exploration of 100 Million Images*, in "MM 2019 - 27th ACM International Conference on Multimedia", Nice, France, ACM Press, October 2019, pp. 1029-1031 [*DOI :* 10.1145/3343031.3350580], https://hal.inria.fr/hal-02378272

[17] F. C. G. REIS, R. ALMEIDA, E. KIJAK, S. MALINOWSKI, S. J. F. GUIMARÃES, Z. DO PATROCINIO. *Combining convolutional side-outputs for road image segmentation*, in "IJCNN 2019 - International Joint Conference on Neural Networks", Budapest, Hungary, IEEE, July 2019, pp. 1-8 [*DOI :* 10.1109/IJCNN.2019.8851843], https://hal.inria.fr/hal-02370834

[18] M. SHI, Z. YANG, C. XU, Q. CHEN. *Revisiting Perspective Information for Efficient Crowd Counting*, in "CVPR 2019 - IEEE Computer Vision and Pattern Recognition Conference", Long Beach, United States, IEEE, June 2019, pp. 1-10, https://hal.inria.fr/hal-01831109

[19] O. SIMÉONI, Y. AVRITHIS, O. CHUM. *Local Features and Visual Words Emerge in Activations*, in "CVPR 2019 - IEEE Computer Vision and Pattern Recognition Conference", Long Beach, United States, IEEE, June 2019, pp. 1-10, https://hal.archives-ouvertes.fr/hal-02374156

[20] R. SOUZA, R. ALMEIDA, R. MIRANDA, Z. KLEBER GONÇALVES DO PATROCINIO, S. MALINOWSKI, S. J. F. GUIMARÃES. *BRIEF-based mid-level representations for time series classification*, in "CIARP 2019 - 24th Iberoamerican Congress on Pattern Recognition", La Havane, Cuba, October 2019, pp. 449-457 [*DOI :* 10.1007/978-3-030-33904-3_42], https://hal.inria.fr/hal-02371260

Conferences without Proceedings

[21] C. DALLOUX, V. CLAVEAU, N. GRABAR. *Speculation and negation detection in french biomedical corpora*, in "RANLP 2019 - Recent Advances in Natural Language Processing", Varna, Bulgaria, September 2019, pp. 1-10, https://hal.archives-ouvertes.fr/hal-02284444

[22] T. FURON. *Watermarking error exponents in the presence of noise: The case of the dual hypercone detector*, in "IH&MMSEC'19 - 7th ACM Workshop on Information Hiding and Multimedia Security", Paris, France,

ACM, July 2019, pp. 173-181 [*DOI :* 10.1145/3335203.3335731], https://hal.archives-ouvertes.fr/hal-02122206

[23] M. GHEISARI, T. FURON, L. AMSALEG. *Group Membership Verification with Privacy: Sparse or Dense?*, in "WIFS 2019 - IEEE International Workshop on Information Forensics and Security", Delft, Netherlands, IEEE, December 2019, pp. 1-6, https://hal.archives-ouvertes.fr/hal-02307926

[24] M. GHEISARI, T. FURON, L. AMSALEG, B. RAZEGHI, S. VOLOSHYNOVSKIY. *Aggregation and embedding for group membership verification*, in "ICASSP 2019 - IEEE International Conference on Acoustics, Speech and Signal Processing", Brighton, United Kingdom, IEEE, May 2019, pp. 2592-2596, https://arxiv.org/abs/1812.03943 - accepted at ICASSP 2019 [*DOI :* 10.1109/ICASSP.2019.8682422], https://hal.archives-ouvertes.fr/hal-02091036

[25] N. GRABAR, C. GROUIN, T. HAMON, V. CLAVEAU. *Annotated corpus with clinical cases in French.*, in "TALN 2019 - 26e Conference on Traitement Automatique des Langues Naturelles", Toulouse, France, July 2019, pp. 1-14, https://hal.archives-ouvertes.fr/hal-02391878

[26] N. GRABAR, C. GROUIN, T. HAMON, V. CLAVEAU. *Information Retrieval and Information Extraction from Clinical Cases. Presentation of the DEFT 2019 Challenge*, in "DEFT 2019 - Défi fouille de texte", Toulouse, France, July 2019, pp. 1-10, https://hal.archives-ouvertes.fr/hal-02280852

[27] C. GROUIN, N. GRABAR, V. CLAVEAU, T. HAMON. *Clinical Case Reports for NLP*, in "BioNLP 2019 - 18th ACL Workshop on Biomedical Natural Language Processing", Florence, Italy, ACL, August 2019, pp. 273–282 [*DOI :* 10.18653/v1/W19-5029], https://hal.archives-ouvertes.fr/hal-02371243

[28] A. PERQUIN, G. LECORVÉ, D. LOLIVE, L. AMSALEG. *Évaluation objective de plongements pour la synthèse de parole guidée par réseaux de neurones*, in "Traitement automatique du langage naturel", Toulouse, France, July 2019, https://hal.archives-ouvertes.fr/hal-02419483

[29] Y. WANG, R. EMONET, E. FROMONT, S. MALINOWSKI, E. MENAGER, L. MOSSER, R. TAVENARD. *Classification de séries temporelles basée sur des "shapelets" interprétables par réseaux de neurones antagonistes*, in "CAp 2019 - Conférence sur l'Apprentissage automatique", Toulouse, France, July 2019, pp. 1-2, https://hal.archives-ouvertes.fr/hal-02268004

Scientific Books (or Scientific Book chapters)

[30] *Proceedings of the workshop DeFt - Défi Fouille de Textes*, July 2019, https://hal.archives-ouvertes.fr/hal-02391768

[31] L. AMSALEG, O. CHELLY, M. E. HOULE, K.-I. KAWARABAYASHI, M. RADOVANOVIĆ, W. TREERATANAJARU. *Intrinsic Dimensionality Estimation within Tight Localities*, in "Proceedings of the 2019 SIAM International Conference on Data Mining", Society for Industrial and Applied Mathematics, May 2019, pp. 181-189 [*DOI :* 10.1137/1.9781611975673.21], https://hal.inria.fr/hal-02125331

[32] C. FABRE, E. MORIN, S. ROSSET, P. SÉBILLOT. *Varia - Préface - 60-1*, ATALA, January 2020, vol. 60, n^{o} 1, pp. 7-11, https://hal.archives-ouvertes.fr/hal-02433763

[33] E. MORIN, S. ROSSET, P. SÉBILLOT. *Varia - Préface - 59-1*, ATALA, May 2019, vol. 59, n^{o} 1, pp. 7-11, https://hal.archives-ouvertes.fr/hal-01789046

Books or Proceedings Editing

[34] L. AMSALEG, B. HUET, M. LARSON, G. GRAVIER, H. HUNG, C.-W. NGO, W. T. OOI (editors). *Proceedings of the 27th ACM International Conference on Multimedia*, ACM Press, Nice, France, October 2019 [*DOI :* 10.1145/3343031], https://hal.inria.fr/hal-02378803

Research Reports

[35] B. CARAMIAUX, F. LOTTE, J. GEURTS, G. AMATO, M. BEHRMANN, F. BIMBOT, F. FALCHI, A. GARCIA, J. GIBERT, G. GRAVIER, H. HOLKEN, H. KOENITZ, S. LEFEBVRE, A. LIUTKUS, A. PERKIS, R. REDONDO, E. TURRIN, T. VIEVILLE, E. VINCENT. *AI in the media and creative industries*, New European Media (NEM), April 2019, pp. 1-35, https://arxiv.org/abs/1905.04175 , https://hal.inria.fr/hal-02125504

Other Publications

[36] A. ANTONINI, M. CARMEN SUÁREZ-FIGUEROA, A. ADAMOU, F. BENATTI, F. VIGNALE, G. GRAVIER, L. LUPI, B. OUVRY-VIAL. *Understanding the phenomenology of reading through modelling*, 2019, pp. 1-22, Work in progress from the READ-IT program (www.readitproject.eu), Project Leader Brigitte Ouvry-Vial, funded by the EU- Joint Programming Initiative for Cultural Heritage, https://hal.archives-ouvertes.fr/hal-02305957

[37] A. ANTONINI, F. VIGNALE, G. GRAVIER, B. OUVRY-VIAL. *The Model of Reading : Modelling principles, Definitions, Schema, Alignments*, 2019, READ-IT Model of Reading -V2, https://hal-univ-lemans.archives-ouvertes.fr/hal-02301611

[38] A. ISCEN, G. TOLIAS, Y. AVRITHIS, O. CHUM. *Label Propagation for Deep Semi-supervised Learning*, November 2019, https://arxiv.org/abs/1904.04717 - Accepted to CVPR 2019, https://hal.inria.fr/hal-02370207

[39] A. ISCEN, G. TOLIAS, Y. AVRITHIS, O. CHUM, C. SCHMID. *Graph Convolutional Networks for Learning with Few Clean and many Noisy Labels*, November 2019, https://arxiv.org/abs/1910.00324 - working paper or preprint [*DOI :* 10.00324], https://hal.inria.fr/hal-02370212

[40] Y. LIFCHITZ, Y. AVRITHIS, S. PICARD, A. BURSUC. *Dense Classification and Implanting for Few-Shot Learning*, November 2019, https://arxiv.org/abs/1903.05050 - CVPR 2019, https://hal.inria.fr/hal-02370192

[41] O. SIMÉONI, Y. AVRITHIS, O. CHUM. *Local Features and Visual Words Emerge in Activations*, November 2019, https://arxiv.org/abs/1905.06358 - working paper or preprint, https://hal.inria.fr/hal-02370209

[42] O. SIMÉONI, M. BUDNIK, Y. AVRITHIS, G. GRAVIER. *Rethinking deep active learning: Using unlabeled data at model training*, November 2019, https://arxiv.org/abs/1911.08177 - working paper or preprint, https://hal.inria.fr/hal-02372102

[43] Z. YANG, M. SHI, Y. AVRITHIS, C. XU, V. FERRARI. *Training Object Detectors from Few Weakly-Labeled and Many Unlabeled Images*, December 2019, https://arxiv.org/abs/1912.00384 - working paper or preprint, https://hal.inria.fr/hal-02393688

[44] H. ZHANG, Y. AVRITHIS, T. FURON, L. AMSALEG. *Smooth Adversarial Examples*, November 2019, https://arxiv.org/abs/1903.11862 - working paper or preprint, https://hal.inria.fr/hal-02370202

[45] H. ZHANG, Y. AVRITHIS, T. FURON, L. AMSALEG. *Walking on the Edge: Fast, Low-Distortion Adversarial Examples*, December 2019, https://arxiv.org/abs/1912.02153 - 13 pages, 9 figures, https://hal.inria.fr/hal-02404216

References in notes

[46] L. AMSALEG, J. E. BAILEY, D. BARBE, S. ERFANI, M. E. HOULE, V. NGUYEN, M. RADOVANOVIC. *The Vulnerability of Learning to Adversarial Perturbation Increases with Intrinsic Dimensionality*, in "WIFS", 2017

[47] L. AMSALEG, O. CHELLY, T. FURON, S. GIRARD, M. E. HOULE, K.-I. KAWARABAYASHI, M. NETT. *Estimating Local Intrinsic Dimensionality*, in "KDD", 2015

[48] L. AMSALEG, G. Þ. GUÐMUNDSSON, B. Þ. JÓNSSON, M. J. FRANKLIN. *Prototyping a Web-Scale Multimedia Retrieval Service Using Spark*, in "ACM TOMCCAP", 2018, vol. 14, n^{o} 3s

[49] L. AMSALEG, B. Þ. JÓNSSON, H. LEJSEK. *Scalability of the NV-tree: Three Experiments*, in "SISAP", 2018

[50] R. BALU, T. FURON, L. AMSALEG. *Sketching techniques for very large matrix factorization*, in "ECIR", 2016

[51] S. BERRANI, H. BOUKADIDA, P. GROS. *Constraint Satisfaction Programming for Video Summarization*, in "ISM", 2013

[52] B. BIGGIO, F. ROLI. *Wild Patterns: Ten Years After the Rise of Adversarial Machine Learning*, in "Pattern Recognition", 2018

[53] P. BOSILJ. *Image indexing and retrieval using component trees*, Université de Bretagne Sud, 2016

[54] X. BOST. *A storytelling machine? : Automatic video summarization: the case of TV series.*, University of Avignon, France, 2016

[55] M. BUDNIK, M. DEMIRDELEN, G. GRAVIER. *A Study on Multimodal Video Hyperlinking with Visual Aggregation*, in "ICME", 2018

[56] N. CARLINI, D. A. WAGNER. *Audio Adversarial Examples: Targeted Attacks on Speech-to-Text*, in "CoRR", 2018, vol. abs/1801.01944

[57] R. CARLINI SPERANDIO, S. MALINOWSKI, L. AMSALEG, R. TAVENARD. *Time Series Retrieval using DTW-Preserving Shapelets*, in "SISAP", 2018

[58] V. CLAVEAU, L. E. S. OLIVEIRA, G. BOUZILLÉ, M. CUGGIA, C. M. CABRAL MORO, N. GRABAR. *Numerical eligibility criteria in clinical protocols: annotation, automatic detection and interpretation*, in "AIME", 2017

[59] A. DELVINIOTI, H. JÉGOU, L. AMSALEG, M. E. HOULE. *Image Retrieval with Reciprocal and shared Nearest Neighbors*, in "VISAPP", 2014

[60] H. FARID. *Photo Forensics*, The MIT Press, 2016

[61] M. GAMBHIR, V. GUPTA. *Recent automatic text summarization techniques: a survey*, in "Artif. Intell. Rev.", 2017, vol. 47, n^{O} 1

[62] I. GOODFELLOW, Y. BENGIO, A. COURVILLE. *Deep Learning*, MIT Press, 2016

[63] G. GRAVIER, M. RAGOT, L. AMSALEG, R. BOIS, G. JADI, E. JAMET, L. MONCEAUX, P. SÉBILLOT. *Shaping-Up Multimedia Analytics: Needs and Expectations of Media Professionals*, in "MMM, Special Session Perspectives on Multimedia Analytics", 2016

[64] A. ISCEN, L. AMSALEG, T. FURON. *Scaling Group Testing Similarity Search*, in "ICMR", 2016

[65] A. ISCEN, G. TOLIAS, Y. AVRITHIS, O. CHUM. *Mining on Manifolds: Metric Learning without Labels*, in "CVPR", 2018

[66] B. Þ. JÓNSSON, G. TÓMASSON, H. SIGURÞÓRSSON, Á. ERÍKSDÓTTIR, L. AMSALEG, M. K. LARUSDOTTIR. *A Multi-Dimensional Data Model for Personal Photo Browsing*, in "MMM", 2015

[67] B. Þ. JÓNSSON, M. WORRING, J. ZAHÁLKA, S. RUDINAC, L. AMSALEG. *Ten Research Questions for Scalable Multimedia Analytics*, in "MMM, Special Session Perspectives on Multimedia Analytics", 2016

[68] H. KIM, P. GARRIDO, A. TEWARI, W. XU, J. THIES, N. NIESSNER, P. PÉREZ, C. RICHARDT, M. ZOLLHÖFER, C. THEOBALT. *Deep Video Portraits*, in "ACM TOG", 2018

[69] M. LAROZE, R. DAMBREVILLE, C. FRIGUET, E. KIJAK, S. LEFÈVRE. *Active Learning to Assist Annotation of Aerial Images in Environmental Surveys*, in "CBMI", 2018

[70] S. LEROUX, P. MOLCHANOV, P. SIMOENS, B. DHOEDT, T. BREUEL, J. KAUTZ. *IamNN: Iterative and Adaptive Mobile Neural Network for Efficient Image Classification*, in "CoRR", 2018, vol. abs/1804.10123

[71] A. LODS, S. MALINOWSKI, R. TAVENARD, L. AMSALEG. *Learning DTW-Preserving Shapelets*, in "IDA", 2017

[72] C. MAIGROT, E. KIJAK, V. CLAVEAU. *Context-Aware Forgery Localization in Social-Media Images: A Feature-Based Approach Evaluation*, in "ICIP", 2018

[73] D. SHAHAF, C. GUESTRIN. *Connecting the dots between news articles*, in "KDD", 2010

[74] M. SHI, H. CAESAR, V. FERRARI. *Weakly Supervised Object Localization Using Things and Stuff Transfer*, in "ICCV", 2017

[75] R. SICRE, Y. AVRITHIS, E. KIJAK, F. JURIE. *Unsupervised part learning for visual recognition*, in "CVPR", 2017

[76] R. SICRE, H. JÉGOU. *Memory Vectors for Particular Object Retrieval with Multiple Queries*, in "ICMR", 2015

[77] O. SIMÉONI, A. ISCEN, G. TOLIAS, Y. AVRITHIS, O. CHUM. *Unsupervised Object Discovery for Instance Recognition*, in "WACV", 2018

[78] O. SIMÉONI, A. ISCEN, G. TOLIAS, Y. AVRITHIS, O. CHUM. *Unsupervised Object Discovery for Instance Recognition*, in "WACV", 2018

[79] H. O. SONG, Y. XIANG, S. JEGELKA, S. SAVARESE. *Deep Metric Learning via Lifted Structured Feature Embedding*, in "CVPR", 2016

[80] C. TSAI, M. L. ALEXANDER, N. OKWARA, J. R. KENDER. *Highly Efficient Multimedia Event Recounting from User Semantic Preferences*, in "ICMR", 2014

[81] C. B. E. VAIGH, F. GOASDOUÉ, G. GRAVIER, P. SÉBILLOT. *Using Knowledge Base Semantics in Context-Aware Entity Linking*, in "DocEng", 2019

[82] O. VINYALS, A. TOSHEV, S. BENGIO, D. ERHAN. *Show and Tell: Lessons Learned from the 2015 MSCOCO Image Captioning Challenge*, in "TPAMI", 2017, vol. 39, n^{o} 4

[83] V. VUKOTIĆ, C. RAYMOND, G. GRAVIER. *Bidirectional Joint Representation Learning with Symmetrical Deep Neural Networks for Multimodal and Crossmodal Applications*, in "ICMR", 2016

[84] V. VUKOTIĆ, C. RAYMOND, G. GRAVIER. *Generative Adversarial Networks for Multimodal Representation Learning in Video Hyperlinking*, in "ICMR", 2017

[85] V. VUKOTIĆ. *Deep Neural Architectures for Automatic Representation Learning from Multimedia Multimodal Data*, INSA de Rennes, 2017

[86] J. WESTON, S. CHOPRA, A. BORDES. *Memory Networks*, in "CoRR", 2014, vol. abs/1410.3916

[87] H. YU, J. WANG, Z. HUANG, Y. YANG, W. XU. *Video Paragraph Captioning Using Hierarchical Recurrent Neural Networks*, in "CVPR", 2016

[88] J. ZAHÁLKA, M. WORRING. *Towards interactive, intelligent, and integrated multimedia analytics*, in "VAST", 2014

[89] L. ZHANG, M. SHI, Q. CHEN. *Crowd Counting via Scale-Adaptive Convolutional Neural Network*, in "WACV", 2018

[90] X. ZHANG, X. ZHOU, M. LIN, J. SUN. *ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices*, in "CoRR", 2017, vol. abs/1707.01083

[91] A. DA SILVA PINTO, D. MOREIRA, A. BHARATI, J. BROGAN, K. W. BOWYER, P. J. FLYNN, W. J. SCHEIRER, A. ROCHA. *Provenance filtering for multimedia phylogeny*, in "ICIP", 2017