

RESEARCH CENTRE

**Inria Centre at Rennes
University**

IN PARTNERSHIP WITH:
CNRS

2024

ACTIVITY REPORT

Project-Team
COMPACT

**COMPRESSiON of mAssively produCED
visual daTa**

IN COLLABORATION WITH: Institut de recherche en informatique et
systèmes aléatoires (IRISA)

DOMAIN

Perception, Cognition and Interaction

THEME

**Vision, perception and multimedia
interpretation**

Inria

Contents

| | |
|--|-----------|
| Project-Team COMPACT | 1 |
| 1 Team members, visitors, external collaborators | 2 |
| 2 Overall objectives | 3 |
| 3 Research program | 5 |
| 4 Application domains | 11 |
| 5 Social and environmental responsibility | 11 |
| 6 Highlights of the year | 12 |
| 7 New software, platforms, open data | 12 |
| 7.1 New software | 12 |
| 7.1.1 color-guidance | 12 |
| 8 New results | 13 |
| 8.1 Axis 1: Compression for specific types of visual data, receivers and media | 13 |
| 8.1.1 Efficient and Fast Light Field Compression via VAE-Based Spatial and Angular Disentanglement | 13 |
| 8.1.2 Visibility-Based Geometry Pruning of Neural Plenoptic Scene Representations | 13 |
| 8.1.3 Coding for Computing and compression rates for universal tasks | 13 |
| 8.1.4 Coding for Machine: learning in the compressed domain | 14 |
| 8.1.5 Storage on DNA: robust transcoding of data | 14 |
| 8.1.6 Partial decoding and RD optimizations for joint multiprofile video coding based on Predictive Residual Coding | 15 |
| 8.1.7 Subsampling design in a multi-profile coding system | 15 |
| 8.2 Axis 2: Sobriety for visual data | 16 |
| 8.2.1 Can image compression rely on CLIP? | 16 |
| 8.2.2 CoCliCo: Extremely low bitrate image compression based on CLIP semantic and tiny color map | 16 |
| 8.2.3 Linearly transformed color guide for low-bitrate diffusion based image compression | 16 |
| 8.2.4 Database compression | 17 |
| 8.3 Axis 3: Acquisition/representation/processing co-design | 17 |
| 8.3.1 Omni-directional Neural Radiance Fields | 17 |
| 8.3.2 GSMorph: Gaussian Splat Morphing with UDF Registration for Dynamic Scene View Synthesis | 18 |
| 8.3.3 Learning-based meta-optics for computational microscopy | 18 |
| 8.4 Axis 4: Learning methods and guarantees | 19 |
| 8.4.1 Deep Equilibrium Maximum A posteriori Score Training | 19 |
| 8.4.2 Message-Passing Guarantees for Graph Coarsening | 19 |
| 8.4.3 Universality of limits of graph neural networks on large random graphs | 19 |
| 8.4.4 Node Regression on Latent Position Random Graphs via Local Averaging | 20 |
| 9 Bilateral contracts and grants with industry | 20 |
| 9.1 Bilateral contracts with industry | 20 |
| 9.1.1 CIFRE contract with TyndallFx on Radiance fields representation for dynamic scene reconstruction | 20 |
| 9.1.2 CIFRE contract with MediaKind on Learned video downscaling for end-to-end Rate-Distortion optimization of video streaming system | 21 |

| | |
|--|-----------|
| 10 Partnerships and cooperations | 21 |
| 10.1 Visits to international teams | 21 |
| 10.2 European initiatives | 21 |
| 10.2.1 H2020 projects | 21 |
| 10.3 National initiatives | 22 |
| 10.3.1 IA Chair: DeepCIM- Deep learning for computational imaging with emerging image modalities | 22 |
| 10.3.2 CominLabs Colearn project: Coding for Learning | 23 |
| 10.3.3 PEPR MolecuArXiv. Targeted project 2: From Digital Data to Synthetic DNA | 23 |
| 10.3.4 PEPR IA. Project SHARP: Theoretical and Algorithmic Principles for frugal ML | 24 |
| 10.3.5 ANR Young researcher grant: MAssive multimedia DAta collection REpurposing (MADARE) | 24 |
| 11 Dissemination | 25 |
| 11.1 Promoting scientific activities | 25 |
| 11.1.1 Scientific events: organisation | 25 |
| 11.1.2 Scientific events: selection | 25 |
| 11.1.3 Journal | 25 |
| 11.1.4 Invited talks | 26 |
| 11.1.5 Leadership within the scientific community | 26 |
| 11.1.6 Scientific expertise | 26 |
| 11.1.7 Research administration | 26 |
| 11.2 Teaching - Supervision - Juries | 26 |
| 11.2.1 Teaching | 26 |
| 11.2.2 Supervision | 27 |
| 11.2.3 Juries | 27 |
| 11.2.4 Internal or external Inria responsibilities | 28 |
| 11.3 Popularization | 28 |
| 11.3.1 Specific official responsibilities in science outreach structures | 28 |
| 12 Scientific production | 28 |
| 12.1 Major publications | 28 |
| 12.2 Publications of the year | 28 |
| 12.3 Cited publications | 30 |

Project-Team COMPACT

Creation of the Project-Team: 2024 July 01

Keywords

Computer sciences and digital sciences

- A5.9. – Signal processing
- A5.9.1. – Sampling, acquisition
- A5.9.2. – Estimation, modeling
- A5.9.3. – Reconstruction, enhancement
- A5.9.4. – Signal processing over graphs
- A5.9.5. – Sparsity-aware processing
- A5.9.6. – Optimization tools
- A8.6. – Information theory
- A8.7. – Graph theory
- A9.2. – Machine learning

Other research topics and application domains

- B3.1. – Sustainable development
- B6.5. – Information systems

1 Team members, visitors, external collaborators

Research Scientists

- Aline Roumy [Team leader, INRIA, Senior Researcher, from Jul 2024]
- Christine Guillemot [INRIA, Senior Researcher, from Jul 2024]
- Nicolas Keriven [CNRS, Researcher, from Jul 2024]
- Thomas Maugey [INRIA, Senior Researcher, from Jul 2024]

Post-Doctoral Fellow

- Remy Leroy [INRIA, Post-Doctoral Fellow, from Jul 2024 until Aug 2024]

PhD Students

- Sara Al Sayyed [INRIA, from Dec 2024]
- Ipek Anil Atalay Appak [CEA, from Jul 2024 until Oct 2024]
- Tom Bachard [UNIV RENNES I, from Jul 2024 until Sep 2024]
- Emmanuel Victor Barobosa Sampaio [INTERDIGITAL, CIFRE, from Dec 2024]
- Stephane Belemkoabga [TYNDAL FX, CIFRE, from Jul 2024]
- Tom Bordin [INRIA, from Jul 2024]
- Davi Rabbouni De Carvalho Freitas [INRIA, from Jul 2024 until Aug 2024]
- Antonin Joly [CNRS, from Jul 2024]
- Esteban Pesnel [MEDIKIND, from Jul 2024]
- Remi Piau [INRIA, from Jul 2024]
- Soheib Takhtardeshir [UNIV MID SWEDEN, from Jul 2024 until Oct 2024]

Technical Staff

- Robin Richard [INRIA, Engineer, from Dec 2024]

Interns and Apprentices

- Sara Al Sayyed [INRIA, Intern, from Jul 2024 until Sep 2024]
- Raphael Leger [CNRS, Intern, from Mar 2024 until Jul 2024]

Administrative Assistant

- Caroline Tanguy [INRIA]

2 Overall objectives

Context

Visual data (images and videos) is omnipresent in various forms (movies, screen content, satellite images, medical images, ...), and provided by different actors ranging from video-on-demand platforms to social networks, and including organizations disseminating Earth observation data. Indeed, video is massively present on the web and accounted for nearly 66% of total internet traffic in 2022 [55]. Therefore, compressing, storing, and transmitting visual data represents a significant societal challenge. Another remarkable fact is that not only does video traffic represent the majority of internet traffic, but it also increases every year. For instance, the number of uploaded hours on Youtube, and shared pictures were multiplied by 10 and 20 respectively in 10 years (Every minute, 48 hours of videos uploaded in 2013 against 500 hours in 2022, and 3.6K shared pictures in 2013 against 66K in 2022 [33]). This acceleration is predicted to continue. Indeed, video traffic on mobile networks accounted for 71% in 2022 and is predicted to reach 80% by 2028 [35]. To address this issue of ever-increasing data volumes, we analyze the usage of videos more finely, and we realize that within video traffic, we can distinguish between **massively generated data** on one hand and **massively viewed data** on the other hand. Massively generated data can either be provided by machines (for instance, in Copernicus, the Earth observation component of the European Union Space Program, 16 TB of observed or prediction data is provided daily [31]), or humans (in 2022, YouTube saw the upload of 500 hours of video content every single minute [33]). Massively viewed data is mostly movies from video-on-demand platforms. These two modes of traffic have different characteristics, and our team proposes to respond specifically to these two contexts. Finally, another consequence of this massive aspect is the energy and ecological impact associated with the processing, storage, and transmission of this data.

General objective

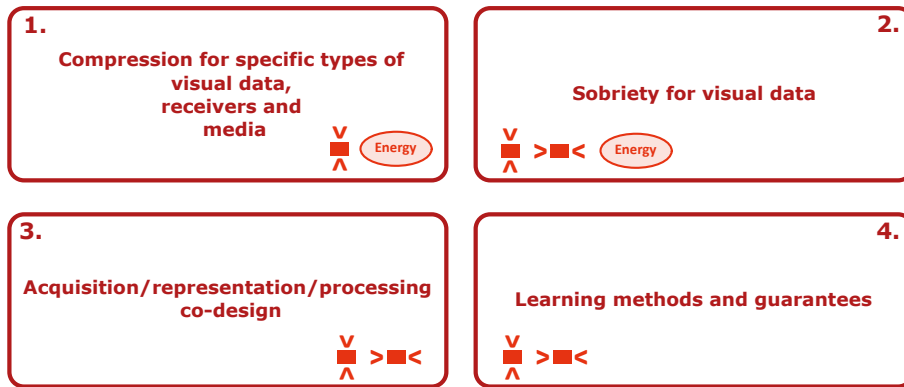
Our main objective is to address the compression problem **in the context of the rapid growth of video usage**, and develop **mathematically grounded algorithms for compressing and processing visual data**. This implies compressing visual data, whose individual volume keeps increasing (new image modalities such as light field, 360, but also higher resolution videos). But it also implies going beyond the classical approach of compressing a single data item to a collection of visual data. To achieve this goal, our team relies on expertise in signal and image processing, statistical machine learning, and information theory. Our originality lies in addressing compression problems in their entirety with contributions that are both practical and theoretical. By doing so, the proposed solutions will address compression challenges comprehensively. More precisely, we begin with a thorough analysis of the compression problem in its practical context taking into account the current context of massively produced data. This will lead to a formulation as an optimization problem and the derivation of information theoretical compression bounds. Subsequently, compression and processing algorithms will be proposed, accompanied by theoretical guarantees regarding content preservation. Finally, validation is performed on real-world data.

Scientific challenges

Compressing this massive data within an ecological transition context leads us to three scientific challenges:

- Reducing the size of **each individual** visual data,
- Reducing the size of a **collection** of visual data,
- Reducing **energy** consumption.

These challenges will be addressed through four main research axes, as shown below:



In the first axis, we will compress data taking into account its usage, i.e., the type of receiver (human versus machine performing inference), as well as its storage mode depending on whether it is hot or cold data. This will both reduce the dimension of the data and provide an energetically efficient solution. In the second axis, the goal is to move towards energy efficiency by proposing algorithms that both reduce the size of individual data and data collections. The third axis also aims to reduce the size of data or data collections, but this time considering the acquisition process and/or a final restoration objective. Finally, many of the proposed methods will be based on machine learning, hence the need to analyze these methods and provide guarantees.

Each axis will be composed of the following sub-axes:

Axis 1. Compression for specific types of visual data, receivers and media,

Axis 1.1. Compression adapted to the data-type,

Axis 1.2. Compression adapted to the user-type: Machine,

Axis 1.3. Compression adapted to the media.

Axis 2. Sobriety for visual data,

Axis 2.1. Ultra-low bitrate visual data compression,

Axis 2.2. Data collection sampling,

Axis 2.3. Low-tech video coders,

Axis 2.4. Sobriety in video usage.

Axis 3. Acquisition/representation/processing co-design,

Axis 3.1. Joint optics/processing,

Axis 3.2. Joint representation/processing: Neural Scene Representation.

Axis 4. Learning methods and guarantees.

Axis 4.1. Optimization methods with learned priors,

Axis 4.2. Learning on graphs,

Axis 4.3. Reducing graphs.

Each of these sub-axes addresses one or several of the initial objectives. Indeed, the first scientific challenge is to **reduce the size of each individual visual data**, such as videos or images. This reduction can be achieved either during acquisition (optics/image processing co-design, compressive acquisition, in Axis 3.1) or after acquisition through a processing leading to a compact representations (low-rank implicit representation, Axis 3.2, learned priors, Axis 4.1, or for a given data type, such as light fields, Axis 1.1). Another approach to size reduction is through the utilization of extremely compact storage mediums, such as DNA storage (Axis 1.3). Furthermore, by considering the usage context, significantly higher compression rates can be achieved when the user is interested in the semantic content rather than

the entirety of the visual data (Axis 2.1), or when performing specific data processing tasks, as in the case of video coding for machines (Axis 1.2).

The second challenge focuses on **reducing the size of a collection of visual data**, for instance, by sampling a database. This sampling can be performed by processing individual data items (Axis 2.2) or by using a structured representation of the database in the form of a graph, addressing issues such as graph reduction (graph sampling, graph coarsening in Axis 4.3), and processing data defined on these graphs (Axis 4.2). Reducing the size of a collection of visual data will also be addressed by learning a compact representation of the whole collection (Axis 3.2).

The third challenge, applicable to both previous challenges, involves **reducing energy consumption**. This will be accomplished through DNA storage research, which offers a low-energy cost storage medium, as well as through optimizing solutions with explicit consideration of global energy costs (for instance in the context of streaming) (Axis 1.3). On top of these necessary efforts for improving the efficiency of coding/storage/transmission systems, a global energy consumption will be targeted, involving the study of efficient and acceptable solutions to aim sobriety in video usage (Axis 2.4).

3 Research program

Axis 1: Compression for specific types of visual data, receivers and media

We start from the observation that visual data is massive but in different ways. For instance, data is individually massive because the dimension of each data point increases, and considering the nature of this data is important for efficient compression (Axis 3). Furthermore, visual data is massively present on networks for different reasons. On one hand, there are massively generated data points that, in some cases, are rarely viewed. On the other hand, there are massively viewed data points that represent a smaller volume than the former. Therefore, it is necessary to propose solutions adapted to each use case.

In the case of massively generated data, the volume of this data is such that it cannot all be visualized by humans. Instead, it will be analyzed by machines, which represents new challenges (Axis 3). Additionally, once analyzed by machines, the rarely viewed cold data can be stored on a medium that allows for low-energy-cost storage, such as DNA (Axis 3). As for the massively viewed data, such as in streaming, the challenge is to offer compression algorithms that optimize not for a financial cost but rather for an energy cost (Axis 3).

Axis 1.1: Compression adapted to the data-type

The field of visual data compression knows new challenges triggered by the emergence of novel modalities (light fields, aka plenoptic, 360° videos, and even holographic data). This research axis focuses on compact representation of light fields. Unlike traditional cameras which capture simple 2D images, light field cameras capture very large volumes of high-dimensional data containing information about the light rays as they interact with the physical objects in the scene. A major challenge in the practical use of light field technology is the huge amount of captured data, hence the need for efficient compression solutions. While in the past decade the problem has been addressed using traditional signal processing models, e.g. sparse or low rank models, these models present some limitations in terms of well capturing and representing the characteristics of real data. Real data in general require much more complex models that cannot be fully expressed analytically. By contrast, machine learning (ML) methods are data-driven approaches which, by learning a very large number of parameters, turn out to be more powerful for encoding and expressing complex data properties. This is especially important for plenoptic data which represents the complexity of the visual worlds in terms of reflective, diffusive, semi-transparent and partially-occluded objects at various depths. In this context, this research axis aims at dealing with high dimensional light field data, focusing on problems of dimensionality reduction for compression while enabling rendering of high quality. Another problem that will be investigated corresponds to the case where the light field or plenoptic data is first represented by a deep network model. The problem of data compression then becomes a problem of dimensionality reduction of Deep Network Models, e.g. for Mobile Computational Plenoptics.

Axis 1.2: Compression adapted to the user-type: Machine

The volumes of visual data being generated [33] are such that these data will not only be viewed by humans but also by machines. For instance, in autonomous vehicles, the machine is the perception system that processes videos to detect objects such as pedestrians, vehicles, traffic signs, and barriers. Another example is the case when a tremendous amount of visual data is uploaded (in social media for instance) and analyzed to make recommendations to humans. A notable difference between compression for humans and compression for machines is that in the case of machines the entirety of the image is not necessary but only some elements are needed to perform the analysis. Hence there is a need to develop specific algorithms for compression for machines.

Furthermore, among the use cases of compression for machines, we can distinguish two scenarios. In the case of cameras embedded in autonomous vehicles, it is known, upon acquisition, that these visual data will be destined for machines. However, due to time and/or computational constraints, the analysis cannot be performed at the camera, and the data need to be compressed and sent to a remote machine. Instead, in the second example of data uploaded on a social media, the primary destination of the data was initially a human, but it is later decided, after compression, that these data will be analyzed by a machine. For these two use cases, the challenges are different. In the first case, the challenge is to (i) develop new compression algorithms that take into account the receiver, machine and the task that will be performed. In the second case, the goal is instead to (ii) develop algorithms that process the data directly in the compressed domain when the compression algorithm has been specifically designed for human vision.

To develop new compression algorithms (i), our approach is to first define the achievable compression rates when the receiver is a machine that is not interested in the entirety of the data but aims to perform processing on it. Our approach will differ from the work of the community [34, 37, 38] in that we incorporate a strict guarantee on the quality of the processing output. The long term objective is to design compression algorithms, where the task may not be known in advance or another task may be chosen (for instance, a new category to be detected).

When the objective is to build algorithms that allow for processing compressed data with an existing algorithm primarily designed for humans (ii), our approach is to avoid decompressing the data. By avoiding data decompression, it is possible to work with more compact representations of the data. The community avoids this decompression when compression is learned for a specific task (i), as in [57, 29, 30]. Conversely, our objective is to construct these algorithms when the compression is performed by an existing algorithm intended for human viewers.

Axis 1.3: Compression adapted to the media

Storing on DNA Data volume growth has led to a projected data storage requirement of 175 ZB by 2025 [54]. However, the actual data storage capacity currently falls short of this forecast. Furthermore, a significant portion of this data is rarely accessed and is categorized as "cold" data. One potential solution to address these challenges is DNA storage as it offers several advantages, including high data density, extended retention, and low energy cost [28]. Indeed, in terms of data density, DNA can store about 10^{19} bytes per cm^3 , enabling the storage of all data generated throughout human history within a 30 cm-sided cube [61]. Regarding retention, DNA can endure for centuries, in contrast to contemporary storage mediums that typically last for decades [61]. Additionally, DNA storage is energy-efficient, since it can be stored at reasonable temperatures, if it is kept away from light and humidity.

Nonetheless, making DNA an efficient storage solution involves overcoming numerous challenges. These challenges encompass:

(i) Data Transformation: convert data into a quaternary code (ACGT). (ii) DNA Synthesis: write data, essentially synthesizing DNA. (iii) DNA Sequencing: extract the quaternary code from DNA, i.e., sequencing DNA. (iv) Data Retrieval: transform back the read quaternary code into the original data. Our primary objective is to address the first and fourth challenges by developing compression algorithms that are robust to synthesis and, more significantly, sequencing errors that occur during steps (ii) and (iii). Indeed, efficient DNA storage heavily relies on rapid sequencing methods, which introduce errors. For instance, real time analysis has been achieved at the price of increased error rates with nanopore sequencing, developed by Oxford Nanopore Technologies (ONT). The main difficulty comes from the

type of errors: nanopore introduces not only conventional substitution errors but also unconventional deletion and insertion errors. Deletion differs from erasure errors, where it is known which part is missing (e.g., lost packets on the internet can be identified by packet headers). Such knowledge of the existence and position of the missing part is unavailable for deletions, and this complicates the correction of this type of error. While the research community largely concentrates on constructing error-correcting codes, our approach aims to develop compression algorithms that are resilient to these errors.

Storing and processing on server for streaming In the case of massively viewed visual data, such as in the case of video streaming, a major objective is to significantly reduce the energy consumption of these solutions. Serving requests is energy-intensive due to the various processing steps undergone by the video before transmission. In fact, the same video content is transmitted with variable qualities (in terms of spatial and temporal resolution, as well as compression errors) in order to adapt to the network bandwidth and receiver type (screen size). In practice, for each request, the high-quality stored video is degraded (in resolution and error level) and then re-compressed. At the decoder level, the video is decompressed and potentially super-resolved to reach the screen resolution. Classically, the optimization of the processing chain is performed to reduce latency and the amount of transmitted data. Instead, our focus is to consider energy consumption as a criterion, and to perform a global optimization taking into account not only transmission, but also storage cost and computation to be performed upon request. This work will be carried out in collaboration with streaming specialist companies. The challenge is to build intermediate representations of videos that provide a video stream compatible with the standard and suitable for transmission (network and screen), thereby optimizing the overall energy balance (storage, server processing, transmission, post-processing at the receiver).

Axis 2: Sobriety for visual data

The sixth report of the Intergovernmental Panel on Climate Change (IPCC) [62] states that if we want to keep the global warming under 1.5°C (Paris agreement), one should target, for 2030, a global emission decrease of 50% when compared to those of 2019. This corresponds to a decrease of 7.6% per year [46]. They also state that *this is not the path that is currently taken*. Hence, every part of our society must urgently aim at sobriety. This is in particular the case of the energy consumed by video data creation/streaming/consumption. In this axis, we will explore solutions enabling a significant reduction of the GreenHouse Gas (GHG) emissions due to video usage. Our strategy is to work on two complementary questions: how to significantly decrease the data size (drastic compression in Axis 3 and data collection sampling in Axis 3)? And how to limit the global video creation and usage (Axis 3)?

Axis 2.1: Ultra-low bitrate visual data compression

The goal of this axis is to reduce the storage cost of cold data, by achieving very high compression ratio. Recently, researchers have proven the existence of a trade-off between distortion and perception when compressing data at low bitrate [25]. In other words, targeting low bitrate inevitably leads to move away from the traditional compression's objective, *i.e.*, keeping faithful decoded data, and to target visual plausibility instead. Therefore, the envisaged solution will semantically describe the visual information in a concise representation, thus leading to drastic compression ratios *exactly as a music score is able to describe, for example, a concert in a compact and reusable form*. This enables the compression to withdraw tremendous amount of useless, or at least not essential, information while condensing the important information into a compact semantic description. At the decoder side, a generative process, relying for example on Diffusion Models [53], is in charge of reconstructing the image or video that is close semantically to the input. In a nutshell, the decoded signals target subjective exhaustiveness of the information description, rather than fidelity to the input data, as in the traditional compression algorithms. Naturally, not all the visual content is meant to be regenerated. Users might be willing to retrieve faithfully the content after decompression. Such approaches will therefore be designed according to user's profile taking into account their choice and interaction. This is a complete change of paradigm, which must enable gigantic compression gains. Considering this approach would use heavy deep learning algorithms and may not tackle data that are often decoded, otherwise the energy due to storage cost reduction would be totally negligible when compared with the huge decoding complexity. On the contrary,

this would perfectly fit with cold data. Finally, in order to be coherent with the purpose of sobriety, we will look for solutions that do not require retraining or even fine-tuning of the heavy Diffusion Models.

Axis 2.2: Data collection sampling

As previously stated, the amount of data created every day is huge *and* exploding. This is certainly accelerated by the fact that most of the social network, video platforms or mobile companies offer the possibility to create, stream and store unlimited data size (or with unreachable bounds), leaving the impression that the storage of data is intangible and cost-less in terms of energy consumption. Increasing the awareness of users or companies requires an efficient way to automatically decide what data deserves to be kept or deleted.

In this axis, we will explore data collection sampling, which consists in selecting the images and videos a user would like to keep among a massive data collection, enabling significant data size savings. This requires first modeling the information perceived by a given user when experiencing a data collection (the initial or the sampled one). This model relies on the volume spanned by the sources features in a personalized latent space. In parallel, we will develop methods to learn the structure and statistics that rule a given data collection. Concretely, among all the pictures of an image collection, some coherent patterns (*e.g.*, landscape, portrait), resemblance between images, chronological landscape evolution or any salient content can be learned and described by mathematical tools, for example with graphs or manifolds. Thereafter these structures will be the support of sampling algorithms aiming at the subjective exhaustiveness of the description, *i.e.*, covering the maximum volume of the learned structure. We will thus pose the trade-off between the rate of the samples (not necessarily taken from the input data, but could be a combination of them) and the quality of the obtained description, driven by the user's preferences.

Axis 2.3: Low-tech video coders

All the recent advances in video compression are due to an increase of the complexity: *e.g.*, more tools and more freedom in the choice of parameters [27] or fully deep learning-based algorithms [48]. In such a context, the global energy cost due to video consumption can only explode, which is not compatible with the urgent need of energetic sobriety. Developing low-energetic video compression/decompression algorithms has been explored for a long time [44, 22, 52]. However, most of the time, the achieved low complexity of the compression algorithms comes from the reduction of the capability of the video coder (*e.g.*, less parameters to estimate, removing of some complex functionalities). Such approaches do not put in question the trade-off between complexity and video coding performance, and thus remain limited.

In this axis, we plan to investigate low complexity algorithms that *are not* low-cost versions of a complex algorithm. The proposed methodology is the following. We start from a complex learning-based coder as for example the auto-encoder-like architecture proposed in [47]. Such architectures are able to achieve outstanding performance, with, however a gigantic encoding and decoding complexity. Our goal is to investigate how to *deduce* from this trained network and its millions of parameters, some efficient features for low complexity compression. As an example, we can show that the set of non-linear operations involved in a deep convolutional neural architecture can be modeled as a linear operation *once the input is fixed*, like it is studied in [51, 50]. The strength of the deep architecture resides in its ability to adjust this linear filter to the input. For our purpose, we will, on the contrary, investigate if some common features reside in these linear filters when the input is changed. These common features may constitute, for example, an efficient transform or partitioning operation that does not require anymore millions of parameters. In a nutshell, the intuition will be to take benefit of algorithms trained on a large set of images and to extract from them some common analysis tools.

Axis 2.4: Sobriety in video usage

Rebound effect or the Jevons's paradox [60] refers to the fact that reducing the cost (in terms of energy or resource consumption) of a technology often leads to an increase of the technology usage and thus to a global increase of the cost, in opposition with the initial goal. Video compression is clearly a good example of this rebound effect. Smaller video sizes (and other technology advances) have led to a global increase of the video usage in today's society. As the ultimate goal, for achieving IPCC objectives, is to

reduce the global carbon footprint of video usage, compression nowadays should not only focus on the reduction of each video file individually. The compression problem should be formulated globally. This inevitably raises the following research question: what is the best (most efficient and acceptable) solution for reducing the amount of videos created/stored/consumed? This question naturally includes the study of user's behavior, and thus deals with other research fields in human and social sciences. The goal of the team COMPACT is twofolds: i) to raise a multidisciplinary research effort on that question by connecting different laboratories and ii) to put its expertise in video compression to the service of this crucial question.

Axis 3: Acquisition/representation/processing co-design

In this axis, the goal is to compress either a data or a collection of data, while taking into account either the acquisition process or a final restoration objective.

Axis 3.1: Joint optics/processing

Our goal is the design of an end-to-end optimization framework designed for acquiring high-resolution images across an extensive Depth of Field (DOF) range within a microscopy system. Microscopy is indeed one key potential application of light field imaging. The optics and post-processing algorithm will be modeled as parts of the end-to-end differentiable computational image acquisition system, allowing for simultaneously optimizing both components. Our computational Extended DOF microscopy imaging system will employ a hybrid approach combining an optical setup with a learned wavefront modulating optical element at the Fourier plane based on metasurfaces. The extended depth of field leads to an increased axial resolution which refers to the ability to distinguish features at different depths by refocusing. While we have obtained initial results for 2D microscopy [23], our goal here will be to extend these results to light field microscopy, which has recently retained the attention of the research community [56, 58, 45].

Axis 3.2: Joint representation/processing: Neural Scene Representation

The task of generating high-quality immersive content with a sufficiently high angular and spatial resolution is technologically challenging, due to the complexity of the constrained capture setup and the bottleneck of data storage and of computational cost. Reconstructing the imaged scene (from a few viewpoints), with a sufficient resolution and quality, and in a way that we can observe it from almost continuously varying positions or angles in space is also an important challenge for a wide adoption in consumer applications. To address the two above problems, the concept of NeRF has been introduced as an implicit model that maps 5D vectors (3D coordinates plus 2D viewing directions) to opacity and color values. The model is based on multi-layer perceptrons (MLP) trained by fitting the model to a set of input views. The learned model is an implicit scene representation that can be used to generate any view of the light field using volume rendering techniques. A variety of works have attempted to handle dynamic scenes in radiance field reconstructions but they either constrain the capture process with multi-view or suffer from quality loss when compared to static scene representations. The proposed research, jointly addressing acquisition, representation and scene reconstruction problems [40, 49, 36] will focus on the reconstruction of neural radiance fields from a limited set of input images, especially in the context of unconstrained, monocular captures, on the completion of the NeRF representation when the capture is incomplete due to a limited set of input images or due to motion in the scene, on the representations of dynamic scenes that are both compact (low memory) and limited in computational complexity. The compactness of scenes will be explored considering joint implicit representations for a collection of data points (2D Images or light fields). The implicit representations inspired from the NeRF concept can be seen as neural network based data representations. The generalization of joint implicit representations to unseen data points assumed to reside in the same subspace as the training data points will also be investigated.

Axis 4: Learning methods and guarantees

A difficulty in visual data (image and video) processing is that their distribution is not known. Therefore, learning-based methods have a certain advantage over model-based methods because they can better adapt to this data. We propose to explore two new ideas in the context of these learning-based methods with the goal of obtaining guarantees on the quality of processing. First, in the context of inverse problems, where the dimension of the observed data is lower than that of the data to be restored, we wish to study the construction of learned priors rather than handcrafted ones, with guarantees stemming from a technique called Deep Equilibrium. In a second approach, we aim to exploit the data's structure (such as a graph), build new learning algorithms adapted to this structure, and obtain theoretical guarantees regarding the learning of the graph but also the learning on the constructed graph.

Axis 4.1: Optimization methods with learned priors

Building upon our past work aiming at taking advantage of learned priors in optimization algorithms, i.e. via plug-and-play and unrolled optimization methods, we will further investigate Deep equilibrium (DEQ) [24] models. Unrolled optimization methods, by coupling optimization algorithms with end-to-end trained regularization, recently emerged as powerful solutions to inverse problems. However, training such unrolled neural networks end-to-end can come with a large memory footprint [39], hence their numbers of iterations are in general limited and they do not generally converge. DEQ models can be seen as an extension of unrolled methods with a theoretically infinite amount of iterations. DEQ models leverage fixed-point properties, allowing for simpler back-propagation. We will further study these models to learn image priors and apply them to inverse problems in classical 2D and new imaging modalities (light fields, omni-directional images).

Axis 4.2: Learning on graphs

In the last decades, there has been a multiplication of data that cannot be properly represented by conventional means, but rather by *relationships* between objects, of various natures and with various properties. Such structures are usually represented as *graphs*. This is for instance the case of collections of (visual) data under the form of *relational databases* (Axis 3), formed by drawing “meaningful” relations between individual data points according to some notion of proximity (semantic, geographical, etc.). Moreover, graphs are increasingly used to represent the structure of (potentially pre-trained) neural networks. Processing this structure using graph machine learning and graph signal processing tools gives rise to the recent topic of (*graph*) *meta-networks* [42], which draws connections with all other axes, particularly the definition of low-tech encoders (Axis 3). Finally, graphs are also a popular representation for geometric data exhibiting *invariance* to certain transforms [26] such as 2D or 3D isometries, often encountered in non-conventional visual data (Axis 3).

(Un)structured data such as graphs posit many challenges. Processing and storing them can be computationally burdensome if done naively. The main challenge resides in the fact that the *regularity* of other types of data (fixed-size vectors, regular grids, well-defined boundaries, etc.), at the basis of many methods, cannot be easily defined here. This axis is thus dedicated to advancing the state-of-the-art in processing efficiently graph data, often through the lens of compression. ML techniques have proved extremely efficient in designing adaptive, data-driven methods *for* compression [59], including for database reduction [32]. Conversely, the extraction of information *from* compressed databases, *a fortiori* by ML, is a major requirement of any compression pipeline. Since graphs have become the *de facto* structure to represent modern relational data, *graph ML* (GML) has known a tremendous development in the last few years, with Graph Neural Networks (GNN) at the forefront of it. Acclaimed for their flexibility, these deep architectures however suffer from many issues, with very limited theoretical and empirical comprehension. A major goal will be to deepen this understanding through the use of tools such as statistical models of large random graphs and information theory. New random graph models adapted to modern real-world data will be developed, focusing on databases arising from visual data but also generic databases, whose analysis will help the choice of GNN architecture, and ultimately lead to new architecture improving the state-of-the-art, in terms of performance and/or computational efficiency.

Axis 4.3: Reducing graphs

Data compression approaches on graphs are referred to as *graph reduction* methods. With modern large graphs numbering millions of nodes, these methods have become a staple of many pipelines, including ML methods mentioned above and database reduction (Axis 3). Graph reduction can be broadly sorted into two related families of algorithms: graph sampling, and graph coarsening.

Graph sampling Graph sampling consists in selecting, often randomly, a reduced number of “representative” node from a large graph. The means to do so, and the downstream tasks to achieve with the subsampled graph, can take many different forms. Particularly interesting for us is the role of graph sampling for fast and efficient querying in large databases [41] (Axis 3), and reducing the size of large neural networks (Axis 3). We will focus on theoretically grounded methods using models of random graphs and information theory, taking into account the specificity of the graph data examined through the previous axes. Since graph sampling is also part of several modern architectures of GNNs, we will incorporate our methods in such models, and examine in which measure sampling methods can be adaptive, data-driven, and/or trained in an end-to-end manner, taking inspiration from modern generative models. Validation will be performed along different criteria, focusing on the classical trade-off between compression rate and performance score, with different choices for the latter depending on the application: supervised classification accuracy, clustering coefficient, etc.

Graph coarsening A related, but somewhat more complex and less well-defined, problem to graph sampling is that of *graph coarsening*, that is, producing an entirely new smaller graph from a large given graph. Again, the purposes can be many, and graph coarsening has an important role in many efficient methods to query and store large databases [43]. Traditional graph coarsening methods seek to preserve certain property of the graph, e.g. spectral properties, and build specific loss functions and performance measurements around these notions.

We will examine whether different coarsening criteria could be defined in a task-dependent manner with guarantees, for instance with the purpose of reducing large neural networks with graph meta-networks [42] (Axis 3), or to expressly design well-adapted convolution operators to be incorporated in neural nets acting on non-Euclidean data (Axis 3). On the theoretical side, we will examine if additional regularity under the form of random graphs models can be exploited. An information-theoretical approach could also lead to new methods. Moreover, graph coarsening is at the heart of *pooling* in GNNs, a very promising lead to improving such architectures by making them “hierarchical” like CNNs, which is still largely open despite an extensive literature on the topic. A more theoretically-grounded approach to the problem could lead to significant advances in this domain.

4 Application domains

Our research is inherently motivated by the application of image and video compression and processing (mostly to help compression denoising, extrapolating such as super-resolution, view synthesis; in the case, of communication to machine, the final goal of object detection and tracking will be also considered, but here as to measure the efficiency of the compression). Two major types of visual data will be considered. First, hot data, such as publicly available data commonly streamed. We will also consider cold data, such as the archival of data that is rarely accessed, as in the case of legal repositories.

5 Social and environmental responsibility

Most of the research fields tackled by the COMPACT team, such as image/video compression, data dimensionality reduction, are inherently aligned with the objective of bringing frugality for processing algorithms. In other words, our algorithms are designed to reduce the energy and resources required for data analysis and consumption. However, while crucial, this research goal is not sufficient to achieve an effective reduction of the environmental footprint of the digital world.

Indeed, the well-known rebound effect makes that such reductions at the algorithm level implies an increase at a broader level (e.g., more videos being created, more learning models being deployed, etc.). The COMPACT team is well aware of this challenge, and is therefore making a strong effort to build collaborations with Social and Human Science researchers. This interdisciplinary approach aims to explore to what extent some limits in the technology usage may be set.

6 Highlights of the year

- The team COMPACT has been involved in the organisation of the french conference CORESA (Nov 2024) taking place at INSA Rennes
- N. Keriven has received an ERC Starting Grant for his project *MALAGA: Reinventing the Theory of Machine Learning on Large Graphs*, that will start in 2025.

Project MALAGA starts with the observation that, in sharp contrast with traditional ML, the field of Graph ML has somewhat jumped from early methods to deep learning, without the decades-long development of well-established notions to compare, analyze and improve algorithms. As a result, 1) Graph Neural Networks (GNNs), all based on the so-called message-passing paradigm, have significant limitations both practical and theoretical, and it is not clear how to address them, and 2) GNNs do not take into account the specificities of graphs coming from domains as different as biology or the social sciences. Overall, these are the symptoms of a major issue: Graph ML is hitting a glass ceiling due to its severe lack of a grand, foundational theory. The ambition of project MALAGA is to develop such a theory. Solving the crucial limitations of the current theory is highly challenging: current mathematical tools cannot analyze the learning capabilities of GML methods in a unified way, existing statistical graph models do not faithfully represent the many characteristics of modern graph data. MALAGA will develop a radically new understanding of GML problems, and of the strengths and limitations of a large panel of algorithms, with the goal to significantly boost the performance, reliability and adaptivity of GNNs.

7 New software, platforms, open data

7.1 New software

7.1.1 color-guidance

Keyword: Image compression

Scientific Description: This study addresses the challenge of controlling the global color aspect of images generated by a diffusion model without training or fine-tuning. We rewrite the guidance equations to ensure that the outputs are closer to a known color map, without compromising the quality of the generation. Our method results in new guidance equations. In the context of color guidance, we show that the scaling of the guidance should not decrease but rather increase throughout the diffusion process. In a second contribution, our guidance is applied in a compression framework, where we combine both semantic and general color information of the image to decode at low cost. We show that our method is effective in improving the fidelity and realism of compressed images at extremely low bit rates (0.001 bpp), performing better on these criteria when compared to other classical or more semantically oriented approaches.

Functional Description: Official implementation of the article: "Linearly transformed color guide for low-bitrate diffusion based image compression" Paper(<https://arxiv.org/pdf/2404.06865>)

Publication: [hal-04882103](https://hal.archives-ouvertes.fr/hal-04882103)

Contact: Tom Bordin

8 New results

8.1 Axis 1: Compression for specific types of visual data, receivers and media

8.1.1 Efficient and Fast Light Field Compression via VAE-Based Spatial and Angular Disentanglement

Participants: Soheib Takhtardeshir, Christine Guillemot.

Light field (LF) imaging captures both spatial and angular information, which is essential for applications such as depth estimation, view synthesis, and post-capture refocusing. However, the efficient processing of this data, particularly in terms of compression and execution speed, presents challenges. We have developed a Variational Autoencoder (VAE)-based framework to disentangle the spatial and angular features of light field images, focusing on fast and efficient compression. Our method uses two separate sub-encoders, one for spatial and one for angular features, to allow for independent processing in the latent space. Evaluations on standard light field datasets have shown that our approach reduces execution time significantly, with a slight trade-off in Rate-Distortion (RD) performance, making it suitable for real-time applications. This framework also enhances other light field processing tasks, such as view synthesis, scene reconstruction, and super-resolution.

8.1.2 Visibility-Based Geometry Pruning of Neural Plenoptic Scene Representations

Participants: Davi R. Freitas, Christine Guillemot.

The need for more realistic 3D scene representations has fomented the development of models for a wide range of applications. In this context, solutions that attempt to model the light's behavior through the plenoptic function have provided considerable advancements using neural-based approaches, often presenting a trade-off between rendering time and model sizes. In collaboration with Ioan Tabus from Tampere Univ., we have firsts proposed a framework for a comparative assessment of implicit and explicit plenoptic scene representation [12]. We have then proposed a pruning framework to reduce the sizes of these models by computing the visibility over the training data so that it can be applied to different representations. In particular, we have implemented our solution primarily for the 3D Gaussian Splatting, while also showing an example of the Neural Radiance Fields (NeRF)-style of rendering using PlenOctrees. We have shown that our pruning solution produces smaller models in terms of the number of elements – be it voxels, points, or Gaussians – with minimal losses in terms of rendering novel views. We further assessed our solution by combining it with state-of-the-art (SOTA) compression solutions for both rendering schemes. Results over the NeRF-Synthetic dataset have shown comparable quality to the SOTA for PlenOctrees, achieving marginal gains for lower bitrates. For 3DGS, the combination of our pruning method and compression solutions achieves a compression ratio of up to 37.5 times over the uncompressed 3DGS models, with only a 0.5 dB decrease in rendering quality. When compared against other SOTA compression methods, our solution produces models 1.4 times smaller, with less than a 0.1 dB loss over novel views.

8.1.3 Coding for Computing and compression rates for universal tasks

Participants: Aline Roumy.

Coding for Computing refers to the problem of compressing data when the goal is not merely to visualize the data but also to perform inference tasks on it. The challenge arises when the inference task is not known prior to encoding. We model this uncertainty by allowing the task to be specified only at the

decoder. Specifically, let X represent the data to be compressed, Y represent the selection of the task to be performed, and $f(X,Y)$ denote the computation executed on X .

In this framework, we address the problem of organizing data during compression to account for possible tasks. This is modeled by introducing a function of the task selection, denoted $g(Y)$, which is available at the encoder. We consider the zero-error information-theoretic problem. Without any side information at the encoder, the zero-error compression rate is generally unknown, as it depends on coloring infinite products of graphs. We first propose a condition under which an analytic expression for the compression rate can be derived with side information available at the encoder. We then investigate the optimization of this side information at the encoder. Finally, we design two greedy algorithms that provide achievable sets of points in the side information optimization problem, based on partition refinement and coarsening. One of these algorithms runs in polynomial time [7].

8.1.4 Coding for Machine: learning in the compressed domain

Participants: Rémi Piau, Thomas Maugey, Aline Roumy.

Due to the ever-growing resolution of visual data, images are often downsampled before being processed. This typically requires fully decoding the data before performing downsampling in the pixel domain. To avoid the computational overhead of full decoding, we propose an alternative approach: learning directly in the compressed domain by subsampling the encoded bitstream. This method eliminates the need to decode the entire dataset, and enables the processing of more compact data representations. However, this task is challenging due to the entropy coding used in most image and video compression schemes, which loses synchronization due to the variable-length nature of the encoded data. Despite these difficulties, we have developed a technique for partial decoding of JPEG-encoded data. Our results demonstrate that partial decoding is feasible and delivers reasonable performance for image classification tasks.

8.1.5 Storage on DNA: robust transcoding of data

Participants: Sara Al Sayyed, Aline Roumy, Thomas Maugey.

The storage of information on DNA requires complex biotechnological processes, which introduce significant noise during both reading and writing. Processes such as synthesis, sequencing, storage, or manipulation of DNA can cause errors that compromise the integrity of the stored data.

Two approaches are possible to address this challenge. The first is to protect the data by adding redundancy and correcting errors in post-processing. The second is to constrain the representation of the data to prevent the generation of such errors. It is this second approach that we have explored. Specifically, we considered the constraint of avoiding homopolymers—sequences where the same polymer is repeated consecutively beyond a given length. This threshold is determined by the type of synthesis and sequencing method used. We proposed two transcoding methods:

- The first completely avoids homopolymers, albeit at the cost of a slight increase in the length of the encoded sequence.
- The second ensures a fixed encoded sequence length, tolerating the creation of a limited number of homopolymers.

Both transcoding methods not only outperform state-of-the-art techniques but also demonstrate that there is no trade-off between compression efficiency and homopolymer avoidance. Indeed, the overheads associated with each method—additional length for the first and the quantity of homopolymers for the second—are negligible.

8.1.6 Partial decoding and RD optimizations for joint multiprofile video coding based on Predictive Residual Coding

Participants: Reda Kaafarani, Aline Roumy, Thomas Maugey.

Video streaming services heavily rely on HTTP-based Adaptive bitrate (ABR) streaming technologies to serve video content to varying end-user device capabilities and network conditions. To adapt to the end-client request, an ABR system encodes, and usually stores, the same video content in different resolution and bitrate pairs, which define a set of encoding profiles. Considering the signal redundancy between representations, we seek to optimize the trade-off between storage bit-cost, transcoding complexity, and transmission efficiency of the different representations in a multi-profile delivery system for standard ABR streaming. The two common approaches for multi-profile video delivery are Simulcast (SC) and Full Transcoding (FT) which set two extremes in terms of criteria optimization. SC offers the lowest transcoding complexity and best transmission efficiency but requires the largest amount of storage, while FT has the lowest storage cost but the highest transcoding cost and transmission bitrate overhead. As a third, in-between approach, the Guided Transcoding using Deflation and Inflation (GTDI) method based on predictive residual coding was proposed. GTDI preserves the best transmission efficiency of SC and reduces its storage but still requires some transcoding complexity as full decoding of the layers is required to regenerate the requested stream. Instead, we propose the idea of partial decoding for standard stream re-generation to explore new tradeoffs. Indeed, the proposed approach significantly lowers the transcoding complexity of the GTDI approach, but sacrifices storage or transmission cost. In addition, we propose novel optimization methods, which improves all methods. For instance, for the GTDI method, this further improves its storage savings by up to -10.3% with a slight 1.7% transmission bitrate overhead in a realistic video delivery scenario.

8.1.7 Subsampling design in a multi-profile coding system

Participants: Esteban Pesnel, Aline Roumy, Thomas Maugey.

The objective of this study is to optimize the subsampling filter in a multi-profile encoding context, without touching the oversampling filter, considered fixed. Indeed, some state-of-the-art methods tend to also optimize the oversampling filter, using machine learning methods (super-resolution, etc.). These methods inevitably lead to a significant increase in algorithmic complexity on the client side, while posing a problem from a standardization point of view. Thus, in the context of the thesis, we will not seek to optimize the decoding and oversampling bricks on the client side. Moreover, learning an optimal undersampling filter is complex, because it must ideally take into account the impact on the rate distortion of the encoder, used to encode undersampled images. However, current encoders are non-differentiable, and thus pose a problem during the error gradient backpropagation step. Some advances in the state of the art propose differentiable neural networks to model the output of an encoder, thus allowing to optimize the learning of the optimal subsampling filter, without being used for inference. We also plan to study and improve these existing contributions, as they do not interfere with the constraints imposed in this study. Following this state-of-the-art work, we have set about the in-depth study of conventional (i.e., non-learned) resampling solutions, via the theoretical aspect (interpolation equations, frequency filtering), and the practical aspect (implementation solutions), by not considering the encoder during this step. A set of relevant conventional resampling filters to study has been defined: conventional linear filters (bicubic, bilinear, Lanczos), non-linear filters (bilateral, median), as well as a filter based on the pseudo-inverse, optimal solution minimizing the L2 norm, of a bicubic upsampling kernel. Two types of experiments have been performed so far. A first one based on the end-to-end distortion of the system, using different downsamplers generating a low-resolution image, followed by a fixed bicubic upsampler. The measured distortion is that between the output of the downsampler and the input image, without encoder (i.e., the downsampled image is directly upsampled and not encoded as in a classical ABR system). In a second step, we studied the performance of these same solutions, when

the downsampled image is encoded and then reconstructed before being upsampled, by encoding the downsampled images with a set of defined codecs JPEG; VVC. The upsampling filter is always set to a bicubic. The distortion is measured between the input image (i.e., before downsampling), and the reconstructed image, at the output of the upscaler. All of these experiments have made it possible to establish a hierarchy among conventional filters, for different ranges of bitrates, compression standards, and resampling factor. It also makes it possible to define the convex hull of the best conventional filter, a reference to beat for the rest of our contributions.

8.2 Axis 2: Sobriety for visual data

8.2.1 Can image compression rely on CLIP?

Participants: Tom Bachard, Thomas Maugey.

Coding algorithms are usually designed to faithfully reconstruct images, which limits the expected gains in compression. A new approach based on generative models for new compression algorithms that can reach drastically lower compression rates. Instead of pixel fidelity, these algorithms aim at faithfully generating images that have the same high-level interpretation as their inputs. In that context, the challenge becomes to set a good representation for the semantics of an image. While text or segmentation maps have been investigated and have shown their limitations, in this paper, we ask the following question: do powerful foundation models such as CLIP provide a semantic description suited for compression? By suited for compression, we mean that this description is robust to traditional compression tools and, in particular, quantization. In [5], we showed that CLIP fulfills semantic robustness properties. This makes it an interesting support for generative compression. To make that intuition concrete, we proposed a proof-of-concept for a generative codec based on CLIP. Results demonstrated that our CLIP-based coder beats state-of-the-art compression pipelines at extremely low bitrates (0.0012 BPP), both in terms of image quality (65.3 for MUSIQ) and semantic preservation (0.86 for the Clip score).

8.2.2 CoCliCo: Extremely low bitrate image compression based on CLIP semantic and tiny color map

Participants: Tom Bachard, Tom Bordin, Thomas Maugey.

After demonstrating that CLIP foundation model was suited for compression (see previous section), we build a complete encoder-decoder solution based on CLIP to achieve ultra-low bitrate image compression: CoCliCo (for COMpression based on CLIP and COLOR) in [9]. In order to complete the CLIP description, we additionally transmit a small color map to depict the general color shape of the image to recover at the decoder. At the decoder, this small color map is noised and used to initialize the diffusion model. To summarize, we encode the inputs into a CLIP latent vector and a tiny color map, and we use a conditional diffusion model for reconstruction. When compared to the most recent traditional and generative coders, our approach reaches drastic compression gains while keeping most of the high-level information and a good level of realism.

8.2.3 Linearly transformed color guide for low-bitrate diffusion based image compression

Participants: Tom Bordin, Thomas Maugey.

The drawback of the CoCliCo method proposed in the previous section is that the color map is used as an initialization of the diffusion model, which does not offer a reconstruction that is always with the desired color organization. In [6], we addressed the challenge of controlling the global color aspect of images generated by a diffusion model without training or fine-tuning. We rewrote the guidance

equations to ensure that the outputs are closer to a known color map, without compromising the quality of the generation. Our method results in new guidance equations. In the context of color guidance, we showed that the scaling of the guidance should not decrease but rather increase throughout the diffusion process. Then, our guidance is applied in a compression framework, where we combined both semantic and general color information of the image to decode at very low cost. We showed that our method is effective in improving the fidelity and realism of compressed images at extremely low bit rates (10^{-2} bps), performing better on these criteria when compared to other classical or more semantically oriented approaches. The implementation of our method is available on gitlab. An example of the obtained results is shown in Figure 1.

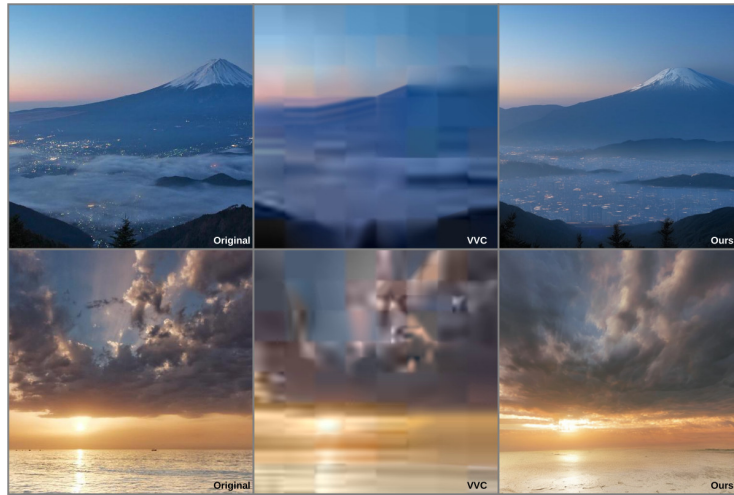


Figure 1: Images decoded using 25×25 resolution color maps and quantized CLIP latent for encoding (0.0041 bps).

8.2.4 Database compression

Participants: Claude Petit, Aline Roumy, Thomas Maugey.

In this work, we move beyond the traditional compression of single data items and address the problem of compressing entire datasets by extracting representative data items. We formulate this as the extraction of submatrices with maximum volume. While existing methods perform well for submatrices whose size is below the rank of the dataset matrix, they struggle to optimize the volume when the submatrix size exceeds the rank. This year, we proposed several novel greedy and one-shot algorithms for extracting maximum-volume rectangular submatrices of sizes greater than the dataset matrix rank. These algorithms are based on a new interpretation of volume maximization through the spectrum of the covariance matrix. Simulations demonstrate that the proposed algorithms outperform both uniform sampling and determinantal point processes (DPPs), producing results that are remarkably close to the exact maximum volume. One of our algorithms achieves quadratic complexity in terms of operations, which, to the best of our knowledge, is the lowest known complexity for maximum-volume extraction algorithms [19].

8.3 Axis 3: Acquisition/representation/processing co-design

8.3.1 Omni-directional Neural Radiance Fields

Participants: Kai Gu, Christine Guillemot, Thomas Maugey.

Radiance Field (RF) representations and their latest variant, 3D- Gaussian Splatting (3D-GS), have revolutionized the field of 3D vision. Novel View Synthesis (NVS) from RF typically requires dense inputs, and for 3D-GS in particular, a high-quality point cloud from a multi-view stereo model is usually necessary. Sparse input RFs are commonly regularized by various priors, such as smoothness, depth, and appearance. Meanwhile, 3D scene segmentation has also achieved significant results with the aid of RFs, and combining the field with different semantic and physical attributes has become a trend. To further tackle NVS and 3D segmentation problems under sparse-input conditions, we have, in collaboration with Sebastian Knorr from Jena Univ., introduced RegSegField, a novel pipeline to utilize 2D segmentations to aid the reconstruction of 3D objects. This method introduces a novel mask-visibility loss by matching 2D segments across different views, thus defining the 3D regions for different objects. To further optimize the correspondence of 2D segments, we have introduced a hierarchical feature field supervised by a contrastive learning method, allowing iterative updates of matched mask areas. To resolve the inconsistent segmentation across different views and refine the mask matching with the help of RF geometry, we also employed a multi-level hierarchy loss. With the help of the hierarchy loss, our method facilitates scene segmentation at discrete granularity levels, whereas other methods require sampling at different scales or determining similarity thresholds. Our experiments show that our regularization approach outperforms various depth-guided NeRF methods and even enables sparse reconstruction of 3D-GS with random initialization [13].

8.3.2 GSMorph: Gaussian Splat Morphing with UDF Registration for Dynamic Scene View Synthesis

Participants: Saphane Belemkoabga, Christine Guillemot, Thomas Maugey.

Monocular novel view synthesis of dynamic scenes presents significant challenges, especially when the views are beyond the original capture trajectory. The reconstruction task is inherently ill-posed due to the limited information available from a monocular setting. To overcome this challenge, we have proposed a method that models both the static geometry and the dynamics of the scenes while preserving geometry consistency during deformations. Using RGB-D data from portable devices, we have introduced a robust depth processing technique to enhance incomplete sensor data. Keyframes are processed using a 3D Gaussian Splatting (3DGS), enhanced with a novel loss function based on an Unsigned Distance Function (UDF) to achieve precise geometry. We also introduce an algorithm for estimating deformation fields for each keyframe, processed in parallel, followed by a fusion step to unify the different geometries and deformation fields. Based on a proposed dataset and an original evaluation methodology involving multiple levels of difficulty, we have shown the effectiveness, both in terms of rendering quality and processing times, of our approach across various scenarios, in comparison with recent methods.

8.3.3 Learning-based meta-optics for computational microscopy

Participants: Ipek Anil Atalay Appak, Christine Guillemot.

Multi-spectral fluorescence microscopy is crucial for analyzing sub-cellular interactions in biological samples but faces challenges like chromatic aberration and limited depth of field (DOF), which degrade image quality and complicate capturing full-depth images in a single exposure. Traditional methods to extend DOF often struggle with light scattering and have limited effectiveness. We have introduced a Novel differentiable learning framework, to design meta-optics specifically for managing chromatic aberration and defocus? The proposed approach, coupled with an advanced image reconstruction

algorithm, enables an extended depth of field, and improves imaging performance across broad spectral ranges.

8.4 Axis 4: Learning methods and guarantees

8.4.1 Deep Equilibrium Maximum A posteriori Score Training

Participants: Christine Guillemot, Samuel Willingham.

Inverse problems are an important part of image processing and deal with the reconstruction of degraded observations. This relates to problems like completion, deblurring and super-resolution. Several approaches allow for high-quality reconstructions. Among those, maximum a posteriori approaches leverage iterative procedures to solve a minimization problem, where the data-distribution is represented by a regularization term. Deep neural networks have proven to be an excellent tool for the representation of the prior data-distribution to be leveraged to find a reconstructed image that is most likely to have caused the observed degraded image, as investigated in our earlier work [15]. Recently, diffusion-based approaches have been suggested to sample images from the posterior distribution. These approaches use a pre-trained score function that is trained on images perturbed by Gaussian noise, which may not represent the distribution of images encountered at inference - sampling from the posterior. In collaboration with Marten Sjostrom from MidSweden Univ., we have proposed a strategy to train a neural network based score function using deep equilibrium models in the context of maximum a posteriori image reconstruction algorithms. The investigation has demonstrated that such a score function can improve reconstruction performance when sampling from a posterior distribution while maintaining the same quality when sampling from the prior.

8.4.2 Message-Passing Guarantees for Graph Coarsening

Participants: Antonin Joly, Nicolas Keriven.

Graph coarsening aims to reduce the size of a large graph while preserving some of its key properties, and has been used in many applications to reduce computational load and memory footprint. For instance, in graph machine learning, training Graph Neural Networks (GNNs) on coarsened graphs leads to drastic savings in time and memory. However, GNNs rely on the Message-Passing (MP) paradigm, and classical spectral preservation guarantees for graph coarsening do not directly lead to theoretical guarantees when performing naive message-passing on the coarsened graph. In [14], we propose a new message-passing operation specific to coarsened graphs, which exhibit theoretical guarantees on the preservation of the propagated signal. Interestingly, and in a sharp departure from previous proposals, this operation on coarsened graphs is oriented, even when the original graph is undirected. We conduct node classification tasks on synthetic and real data and observe improved results compared to performing naive message-passing on the coarsened graph.

8.4.3 Universality of limits of graph neural networks on large random graphs

Participants: Nicolas Keriven.

We propose a notion of universality for graph neural networks (GNNs) in the large random graphs limit, tailored for node-level tasks. When graphs are drawn from a latent space model, GNNs on growing graph sequences are known to converge to limit objects called "continuous GNNs", or cGNNs. A cGNN inputs and outputs functions defined on a latent space, and as such, is a non-linear operator between functional spaces. Therefore, we propose to evaluate the expressivity of a cGNN through its ability to approximate

arbitrary non-linear operators between functional spaces. This is reminiscent of Operator Learning, a branch of machine learning dedicated to learning non-linear operators between functional spaces. In this field, several architectures known as Neural Operators (NOs) are indeed proven to be universal. The justification for the universality of these architectures relies on a constructive method based on an encoder-decoder strategy into finite dimensional spaces, which enables invoking the universality of usual MLP. In [10], we adapt this method to cGNNs. This is however far from straightforward: cGNNs have crucial limitations, as they do not have access to the latent space, but observe it only indirectly through the graph built on it. Our efforts will be directed toward circumventing this difficulty, which we will succeed, at this stage, only with strong hypotheses.

8.4.4 Node Regression on Latent Position Random Graphs via Local Averaging

Participants: Nicolas Keriven.

Node regression consists in predicting the value of a graph label at a node, given observations at the other nodes. To gain some insight into the performance of various estimators for this task, in [21] we perform a theoretical study in a context where the graph is random. Specifically, we assume that the graph is generated by a Latent Position Model, where each node of the graph has a latent position, and the probability that two nodes are connected depend on the distance between the latent positions of the two nodes. In this context, we begin by studying the simplest possible estimator for graph regression, which consists in averaging the value of the label at all neighboring nodes. We show that in Latent Position Models this estimator tends to a Nadaraya Watson estimator in the latent space, and that its rate of convergence is in fact the same. One issue with this standard estimator is that it averages over a region consisting of all neighbors of a node, and that depending on the graph model this may be too much or too little. An alternative consists in first estimating the true distances between the latent positions, then injecting these estimated distances into a classical Nadaraya Watson estimator. This enables averaging in regions either smaller or larger than the typical graph neighborhood. We show that this method can achieve standard nonparametric rates in certain instances even when the graph neighborhood is too large or too small.

9 Bilateral contracts and grants with industry

9.1 Bilateral contracts with industry

9.1.1 CIFRE contract with TyndallFx on Radiance fields representation for dynamic scene reconstruction

Participants: Stephane Belemkoabga, Christine Guillemot, Thomas Maugey.

- Title : Radiance fields representation for dynamic scene reconstruction
- Partners : TyndallFx (R. Mallart), Inria-Rennes.
- Funding : TyndallFx, ANRT.
- Period : Oct-2023-Sept. 2026.

The goal of this project is to design novel methods for modeling and compact representation of radiance fields for scene reconstruction and view synthesis. The problems that are addressed are those of fast and efficient estimation of the camera pose parameters and of the 3D model of the scene based on Gaussian splatting, and as as the one of tracking and modeling the deformation of the model due to the global camera motion and to the motion of the different objects in the scene.

9.1.2 CIFRE contract with MediaKind on Learned video downscaling for end-to-end Rate-Distortion optimization of video streaming system

Participants: Esteban Pesnel, Aline Roumy, Thomas Maugey.

- Title : Learned video downscaling for end-to-end Rate-Distortion optimization of video streaming system
- Partners : MediaKind, Inria-Rennes.
- Funding: MediaKind, ANRT.
- Period : November 2023-October 2026.

This CIFRE contract aims to optimize a streaming solution by addressing constraints related to distribution, standards, and deployment. The focus is on developing downscaling techniques that enhance the end-to-end streaming process, considering bitrate-distortion optimization. While the upscaling filter on client devices is fixed due to standardization, encoding and downscaling on the server side remain flexible, offering an opportunity for improvement within the streaming pipeline.

10 Partnerships and cooperations

10.1 Visits to international teams

Research stays abroad

Participants: Tom Bordin.

- Visited institution: La Sapienza University
- Country: Roma, Italy
- Dates: June-July 2024 (2 months)
- Context of the visit: research collaboration with Prof Sergio Barbarossa, on the topic of semantic compression.
- Mobility program/type of mobility: outgoing international mobility for doctoral researchers from the doctoral school MATISSE.

10.2 European initiatives

10.2.1 H2020 projects

PLENOPTIMA

Participants: Davi Freitas, Kai Gu, Anil Ipek Atalay Appak, Christine Guillemot, Thomas Maugey, Soheib Takhtardeshir, Samuel Willingham.

Title: Plenoptima: Plenoptic Imaging

Duration: From January 1, 2021 to December 31, 2025

Partners:

- INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE (INRIA), France
- MITTUNIVERSITETET (MIUN), Sweden
- TECHNISCHE UNIVERSITÄT BERLIN (TUB), Germany
- TAMPEREEN KORKEAKOULUSAATIO SR (TAMPERE UNIVERSITY), Finland
- "INSTITUTE OF OPTICAL MATERIALS AND TECHNOLOGIES "ACADEMICIAN JORDAN MALINOWSKI" - BULGARIAN ACADEMY OF SCIENCES" (IOMT), Bulgaria

Inria contact: Christine Guillemot

Coordinator: Tampere University (Finland, Atanas Gotchev)

Summary: Plenoptic Imaging aims at studying the phenomena of light field formation, propagation, sensing and perception along with the computational methods for extracting, processing and rendering the visual information.

The PLENOPTIMA ultimate project goal is to establish new cross-sectorial, international, multi-university sustainable doctoral degree programmes in the area of plenoptic imaging and to train the first fifteen future researchers and creative professionals within these programmes for the benefit of a variety of application sectors. PLENOPTIMA develops a cross-disciplinary approach to imaging, which includes the physics of light, new optical materials and sensing principles, signal processing methods, new computing architectures, and vision science modelling. With this aim, PLENOPTIMA joints five of strong research groups in nanophotonics, imaging and machine learning in Europe with twelve innovative companies, research institutes and a pre-competitive business ecosystem developing and marketing plenoptic imaging devices and services.

PLENOPTIMA advances the plenoptic imaging theory to set the foundations for developing future imaging systems that handle visual information in fundamentally new ways, augmenting the human perceptual, creative, and cognitive capabilities. More specifically, it develops 1) Full computational plenoptic imaging acquisition systems; 2) Pioneering models and methods for plenoptic data processing, with a focus on dimensionality reduction, compression, and inverse problems; 3) Efficient rendering and interactive visualization on immersive displays reproducing all physiological visual depth cues and enabling realistic interaction.

All ESRs are registered in Joint/Double degree doctoral programmes at academic institutions in Bulgaria, Finland, France, Germany and Sweden. The programmes will be made sustainable through a set of measures in accordance with the Salzburg II Recommendations of the European University Association.

10.3 National initiatives

10.3.1 IA Chair: DeepCIM- Deep learning for computational imaging with emerging image modalities

Participants: Christine Guillemot, Remi Leroy.

- Funding: ANR (Agence Nationale de la Recherche).
- Period: Sept. 2020 - Aug. 2024.

The project aimed at leveraging recent advances in three fields: image processing, computer vision and machine (deep) learning. It focused on the design of models and algorithms for data dimensionality reduction and inverse problems with emerging image modalities. The first research challenge concerned the design of learning methods for data representation and dimensionality reduction. These methods have been encompassing the learning of sparse and low rank models, of signal priors or representations in latent spaces of reduced dimensions. This also included the learning of efficient and, if possible, lightweight architectures for data recovery from the representations of reduced dimension. Modeling

joint distributions of pixels constituting a natural image is also a fundamental requirement for a variety of processing tasks. This is one of the major challenges in generative image modeling, field conquered in recent years by deep learning. Based on the above models, our goal was also to develop algorithms for solving a number of inverse problems with novel imaging modalities. Solving inverse problems to retrieve a good representation of the scene from the captured data requires prior knowledge on the structure of the image space. Deep learning based techniques designed to learn signal priors, to be used as regularization models.

10.3.2 CominLabs Colearn project: Coding for Learning

Participants: Rémi Piau, Thomas Maugey, Aline Roumy.

- Partners: Inria-Rennes (Compact team); LabSTICC, IMT Atlantique, (team Code and SI3); IETR, INSA Rennes (Syscom team).
- Funding: Labex CominLabs.
- Period: Sept. 2021 - Dec. 2025.

The amount of data available online is growing so fast that it is essential to rely on advanced Machine Learning techniques so as to automatically analyze, sort, and organize the content uploaded by e.g. sensors or users. The conventional data transmission framework assumes that the data should be completely reconstructed, even with some distortions, by the server. Instead, this project aims to develop a novel communication framework in which the server may also apply a learning task over the coded data. The project will therefore develop an Information Theoretic analysis so as to understand the fundamental limits of such systems, and develop novel coding techniques allowing for both learning and data reconstruction from the coded data.

10.3.3 PEPR MolecuArXiv. Targeted project 2: From Digital Data to Synthetic DNA

Participants: Sara Al Sayeed, Aline Roumy, Thomas Maugey.

- Partners: I3S, LabSTIC, IMT-Atlantique, Irisa/Inria (GenScale and Compact team), IPMC, Eurecom.
- Funding: France 2030.
- Period: Sept. 2022 - Dec. 2029.

The PEPR MolecuArXiv aims to develop future data storage devices on molecular media, including DNA and artificial polymers. This involves not only parallelizing synthesis devices but also discovering new molecules and information technologies to accelerate the synthesis of storage media, their encoding and decoding, and exploring various molecular supports.

Within the targeted project "From Digital Data to Synthetic DNA," the objective is to make physical and logical storage efficient through custom-designed codes tailored to the physicochemical constraints of DNA writing and reading. This effort is conducted in collaboration with partners from other targeted projects, such as "Next-Generation DNA Synthesis" and "Synthetic Digital Polymers."

Several key challenges are addressed, including robustness to noise. Processes like synthesis, sequencing, storage, or manipulation of DNA can introduce errors that threaten the integrity of the stored data. These errors are non-classical compared to those encountered in wired and wireless communication channels and require specific handling. This issue is approached from the perspectives of both compression and error-correcting codes.

Another critical challenge is data access. A significant advantage of storing information on DNA, apart from its durability, is its extremely high density, enabling vast amounts of data to be stored compactly.

Due to this high density, it is essential to facilitate rapid access to the required data items. New data representations are studied to enable fast random access to the data relying merely on biological and chemical processes.

10.3.4 PEPR IA. Project SHARP: Theoretical and Algorithmic Principles for frugal ML

Participants: Antonin Joly, Nicolas Keriven.

- Partners: LIP, LAMSADE, LIGM, ISIR, CEA (team List), Inria (team Ockham, Compact, Genesis), CNRS.
- Funding: France 2030.
- Period: 2024 - 2029.

The amount of resources needed to train and deploy state of the art deep neural models is enormous, from CPUs, GPUs and storage to bandwidth and dataset acquisition (with possible annotations). More importantly, it steadily increases at an unsustainable pace, leading to a growing and alarming environmental footprint, and results in significant societal tensions. This raises serious concerns, not only in flagship domains such as computer vision (CV) and natural language processing (NLP) where all players become dependent on the main worldwide players (with the risk of an AI oligopoly), but also in many domains where data is scarce and resource-hungry models cannot be successfully trained with current know-how.

The major challenge of the SHARP project is to achieve a leap forward in frugality by designing, analyzing and deploying intrinsically efficient models (neural or not) able to achieve the versatility and performance of the best models while requiring only a vanishing fraction of the resources currently needed.

To achieve its vision, SHARP will design a principled theoretical and algorithmic framework to leverage prior knowledge and the modern avatars of the notion of sparsity of predictors and/or algorithms, and establish a new paradigm of computational representation learning bypassing the current technical and computing bottlenecks. Two showcase demonstrations of the impact of SHARP will be the frugal training of compact transformers with negligible performance loss, and the development of effective representation learning models on small unlabeled datasets, for a selected downstream application. To achieve such frugality in AI, SHARP will rely on three pillars: frugal architectures, frugal learning principles, and learning with small and scarce datasets.

10.3.5 ANR Young researcher grant: MASSive multimedia DATA collection REPurposing (MADARE)

Participants: Tom Bordin, Robin Richard, Thomas Maugey.

- Funding: ANR (Agence Nationale de la Recherche)
- Period: pr. 2022 - Oct. 2025.

Compression algorithms are nowadays overwhelmed by the tsunami of visual data created everyday. Despite a growing efficiency, they are always constrained to minimize the compression error, computed in the pixel domain. The Data Repurposing framework, proposed in the MADARE project, will tear down this barrier, by allowing the compression algorithm to “reinvent” part of the data at the decoding phase, and thus saving a lot of bit-rate by not coding it. Concretely, a data collection is only encoded to a compact description that is used to guarantee that the regenerated content is semantically coherent with the initial one. In practice, it opens several research directions: how to organise the latent space (in which the coded descriptions lie) such that the information is efficiently and intelligibly represented? How to regenerate a

synthesized content from this compact description (based for example on guided diffusion algorithms)? Finally, how to extend this idea to video? By revisiting the compression problem, the MADARE project aims gigantic compression ratios enabling, among other benefits, to reduce the impact of exploding data creation on the cloud servers' energy consumption.

11 Dissemination

11.1 Promoting scientific activities

11.1.1 Scientific events: organisation

General chair, scientific chair

- T. Maugey was general co-chair for the national conference CORESA in Rennes, November
- T. Maugey was the general chair of two pluridisciplinary workshops on “digital sobriety” co-organized with Marsouin (in June and December)

Member of the organizing committees

- T. Maugey was international liaison chair for the international conference IEEE PCS in Taiwan, June
- N. Keriven was co-organizer of the Learning on Graphs conference Paris meetup, Nov. 2024.

11.1.2 Scientific events: selection

Member of the conference program committees

- A. Roumy was a member of the technical program committee of the (Conference on Computer Vision and Pattern Recognition) CVPR 2024 workshop on New Trends in Image Restoration and Enhancement (NTIRE).
- A. Roumy was a member of the technical program committee of the (European Conference on Computer Vision) ECCV 2024 workshop on Advances in Image Manipulation (AIM).

Reviewer

- A. Roumy was a meta-reviewer for the 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) conference.

11.1.3 Journal

Member of the editorial boards

- T. Maugey is associate editor of the EURASIP Journal on Advances in Signal Processing.
- T. Maugey is associate editor of the IEEE Signal Processing Letter.
- A. Roumy is Senior Associate Editor of the IEEE Transactions on Image Processing.
- N. Keriven is associate editor of the IEEE Transactions on Signal and Information Processing over Networks journal.

11.1.4 Invited talks

- T. Maugey gave a talk at Padova university (Italy) on “semantic image compression” (July)
- T. Maugey gave a talk at the EIT Digital Summer School, University of Rennes (July)
- A. Roumy gave a tutorial on “Information theory for image and video compression: fundamental results and recent challenges”, Coresa conference, Nov. 2024.
- N. Keriven gave a talk at Asilomar conference (US) on “Seeking universal approximation for continuous limits of GNNs on large random graphs”, Oct. 2024.

11.1.5 Leadership within the scientific community

- T. Maugey is member of the EURASIP Technical Area Committee on Visual Information Processing
- A. Roumy is a member of the IEEE Image, Video, and Multidimensional Signal Processing Technical Committee (IVMSP TC).
- A. Roumy is a member of the Executive board of the National Research group in Image and Signal Processing (GRETSI).

11.1.6 Scientific expertise

- C. Guillemot is a member of the ERC-Cell at Inria.
- A. Roumy served as a member of Board of Examiners (Comité de sélection) for 2 Professor positions (Professeur des Universités) at Rennes University, May 2024.
- A. Roumy has been a member of the Inria Delegation committee of the Inria center at Rennes University.
- A. Roumy was a reviewer for the evaluation committee for the appointment of a maître de conférences (associate professor), Telecom Paris, Sept. 2024.

11.1.7 Research administration

- C. Guillemot is a member of the research commission of the University of Rennes, as Inria representative
- C. Guillemot has been a member of the jury for recruiting a Professor at the Univ. of Nantes, May 2024.
- C. Guillemot has been a member of the jury of the concours CRCN/ISFP of Rennes, April 2024.
- A. Roumy is a member of the research commission and of the academic board of the University of Rennes 2, as Inria representative
- A. Roumy has been a member of the Inria Delegation committee of the Inria center at Rennes University

11.2 Teaching - Supervision - Juries

11.2.1 Teaching

- T. Maugey has given a course on Graph Image Processing, 10 hours, M2 SiVOS, Univ. of Rennes, France.
- T. Maugey has given a course on Representation, editing and perception of digital images, 21 hours, M2 SIF, Univ. of Rennes, France.

- T. Maugey has given a course on Ecological Transition and digital world, 6 hours, L3 SIF, ENS Rennes, France.
- A. Roumy has given an Engineering degree course on the foundations of Image compression, 36 hours, Univ. of Rennes, ESIR, France.
- A. Roumy has given an Engineering degree course on Image and Video compression, 10 hours, Univ. of Rennes, ESIR, France.
- N. Keriven has given a course on Machine Learning, 24 hours, ENSTA Paris, France.
- N. Keriven has given a course on Signal Processing, 12 hours, ENS Rennes, France.

11.2.2 Supervision

- C. Guillemot and T. Maugey are co-supervising the PhD thesis of S. Belemkoabga in the context of the Cifre contract with TyndallFX.
- C. Guillemot is co-supervising the double degree PhD thesis of S. Willingham (together with M. Sojstrom from MidSeden University), of S. Takhtardeshir (together with M. Sojstrom and R. Olsson from MidSeden University), of Davi Freitas (together with I. Tabus from Tampere Univ.), of K. Gu (together with T. Maugey from Inria and S. Knorr from Technical Univ. Berlin), and of A. Atalay Appak (together with H. Caglayan and E. Sahim from Tampere University). These PhDs are double degrees in the framework of the European Marie Curie project Plenoptima.
- T. Maugey is supervising the PhD thesis of Tom Bachard (ended in December 2024) and Tom Bordin (University of Rennes) and Emmanuel Victor Barobosa Sampaio (Cifre Contract with InterDigital).
- T. Maugey and A. Roumy are co-supervising the PhD thesis of Rémi Piau in the context of the Cominlabs project CoLearn
- T. Maugey and A. Roumy are co-supervising the PhD thesis of Esteban Pesnel in the context of the Cifre contract with MediaKind
- T. Maugey and A. Roumy are co-supervising the PhD thesis of Sara Al Sayyed in the context of the PEPR MolecularXiv
- N. Keriven and A. Roumy are co-supervising the PhD thesis of Antonin Joly in the context of the PEPR SHARP
- A. Roumy was supervising the PhD thesis of Claude Petit
- N. Keriven was supervising the PhD thesis of M. Cordonnier with S. Vaiter (CNRS, LJAD, Nice) and N. Tremblay (CNRS, GIPSA-lab, Grenoble)
- N. Keriven is supervising the PhD thesis of M. Gjorgjevski with S. Barthelmé (CNRS, GIPSA-lab, Grenoble) and Y. de Castro (Pr Centrale Lyon).
- N. Keriven was supervising the internship of R. Léger

11.2.3 Juries

- C. Guillemot was member of the PhD committees of:
 - N. Madali, Insa-Rennes, Sept. 2024, as chair.
 - M. Dufraisse, Univ. Paris-Saclay, Oct. 2024, as rapporteur.
 - H. McGriff, Univ. Bourgogne-Franche Comté, Dec. 2024, as chair.
- T. Maugey was member of the PhD committees of:
 - Jingwen Zhu (Sept 2024), University of Nantes as president

- Melan VIJAYARATNAM (Oct. 2024), Télécom Paris as reviewer
- Claude PETIT (Dec. 2024), Inria, France
- Joshua MARAVAL (Dec. 2024), IETR/INSA, Rennes, France
- Natacha LUKA (Dec. 2024), ENPC as reviewer
- A. Roumy was member of the PhD committees of
 - Andreas Pastor, Polytech' Nantes, Nantes University, Aug. 2024, as a chair.
 - Xavier Pic, Univ. Côte d'Azur, Sept. 2024, as a reviewer.
- and member of the Habilitation (HDR) committee of
 - Enzo Tartaglione, Institut Polytechnique de Paris, Telecom Paris, Sept. 2024.
- N. Keriven was member, as rapporteur, of the PhD jury of M. Ballu, Cambridge University, July 2024.

11.2.4 Internal or external Inria responsibilities

- A. Roumy is a member of the Gender Equality committee of Inria-Rennes and Irisa, responsible for the working group on career interruptions and support.
- A. Roumy is a member of the mentorship program as a mentor.

11.3 Popularization

11.3.1 Specific official responsibilities in science outreach structures

- T. Maugey is Scientific mediation officer in the scientific mediation team of Inria centre at Rennes University.

12 Scientific production

12.1 Major publications

- [1] T. Bordin and T. Maugey. 'Linearly transformed color guide for low-bitrate diffusion based image compression'. In: *IEEE Transactions on Image Processing* (18th Dec. 2024), p. 15. URL: <https://inria.hal.science/hal-04882103>. In press.
- [2] N. Charpenay, M. Le Treust and A. Roumy. 'Side Information Design in Zero-Error Coding for Computing'. In: *Entropy* 26.4 (16th Apr. 2024), pp. 1–18. DOI: [10.3390/e26040338](https://doi.org/10.3390/e26040338). URL: <https://hal.science/hal-04679240>.
- [3] A. Joly and N. Keriven. 'Graph Coarsening with Message-Passing Guarantees'. In: *Advances in Neural Information Processing Systems (NeurIPS)*. Advances in Neural Information Processing Systems (NeurIPS). Vancouver, Canada, 2024. URL: <https://hal.science/hal-04617519>.
- [4] M. Le Pendu and C. Guillemot. 'Preconditioned Plug-and-Play ADMM with Locally Adjustable Denoiser for Image Restoration Mikael'. In: *SIAM Journal on Imaging Sciences* (Nov. 2022), pp. 1–30. URL: <https://hal.science/hal-03857826>.

12.2 Publications of the year

International journals

- [5] T. Bachard and T. Maugey. 'Can image compression rely on CLIP?' In: *IEEE Access* (2024), pp. 1–17. DOI: [10.1109/ACCESS.2024.3408651](https://doi.org/10.1109/ACCESS.2024.3408651). URL: <https://hal.science/hal-04600515> (cit. on p. 16).

- [6] T. Bordin and T. Maugey. ‘Linearly transformed color guide for low-bitrate diffusion based image compression’. In: *IEEE Transactions on Image Processing* (18th Dec. 2024), p. 15. URL: <https://inria.hal.science/hal-04882103>. In press (cit. on p. 16).
- [7] N. Charpenay, M. Le Treust and A. Roumy. ‘Side Information Design in Zero-Error Coding for Computing’. In: *Entropy* 26.4 (16th Apr. 2024), pp. 1–18. DOI: [10.3390/e26040338](https://doi.org/10.3390/e26040338). URL: <https://hal.science/hal-04679240> (cit. on p. 14).
- [8] J. Shi, Y. Xu and C. G. Fellow. ‘Learning Kernel-Modulated Neural Representation for Efficient Light Field Compression’. In: *IEEE Transactions on Image Processing* 33 (1st July 2024), pp. 4060–4074. URL: <https://hal.science/hal-04627349>.

International peer-reviewed conferences

- [9] T. Bachard, T. Bordin and T. Maugey. ‘CoCliCo: Extremely low bitrate image compression based on CLIP semantic and tiny color map’. In: *IEEE PCS. PCS 2024 - Picture Coding Symposium*. Taichung, Taiwan, 2024, pp. 1–5. URL: <https://inria.hal.science/hal-04478601> (cit. on p. 16).
- [10] M. Cordonnier, N. Keriven, N. Tremblay and S. Vaiter. ‘Seeking universal approximation for continuous limits of graph neural networks on large random graphs’. In: *Asilomar 2024 - Asilomar Conference on Signals, Systems, and Computers*. Pacific Grove, United States: IEEE, 2024, pp. 1–5. URL: <https://hal.science/hal-04831761> (cit. on p. 20).
- [11] M. Dufraisse, R. Leroy, P. Trouvé-Peloux, F. Champagnat and J.-B. Volatier. ‘Multitask deep co-design for extended depth of field and depth from defocus’. In: *2024 - Unconventional Optical Imaging IV*. Strasbourg, France: SPIE, 2024, p. 72. DOI: [10.1117/12.3017194](https://doi.org/10.1117/12.3017194). URL: <https://hal.science/hal-04718782>.
- [12] D. R. Freitas, R. L. de Queiroz, I. Tabus and C. Guillemot. ‘A Comparative Assessment Of Implicit and Explicit Plenoptic Scene Representations’. In: *MMSP 2024 - 26th IEEE International Workshop on Multimedia Signal Processing*. West Lafayette (IN), United States, 2024, pp. 1–6. URL: <https://inria.hal.science/hal-04709909> (cit. on p. 13).
- [13] K. Gu, T. Maugey, S. Knorr and C. Guillemot. ‘RegSegField: Mask-Regularization and Hierarchical Segmentation for Novel View Synthesis from Sparse Inputs’. In: *CVMP 2024 - 21st ACM SIGGRAPH European Conference on Visual Media Production*. London, United Kingdom, 2024, pp. 1–10. DOI: [10.1145/3697294.3697299](https://doi.org/10.1145/3697294.3697299). URL: <https://inria.hal.science/hal-04725951> (cit. on p. 18).
- [14] A. Joly and N. Keriven. ‘Graph Coarsening with Message-Passing Guarantees’. In: *Advances in Neural Information Processing Systems (NeurIPS)*. Advances in Neural Information Processing Systems (NeurIPS). Vancouver, Canada, 2024. URL: <https://hal.science/hal-04617519> (cit. on p. 19).
- [15] S. Willingham, M. Sjöström and C. Guillemot. ‘Training Methods for Regularizing Gradients on Multi-Task Image Restoration Problems’. In: *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. VISAPP 2024 - 19th International Conference on Computer Vision Theory and Applications*. Vol. 3. Rome, Italy, 2024, pp. 145–153. DOI: [10.5220/0012368100003660](https://doi.org/10.5220/0012368100003660). URL: <https://hal.science/hal-04763140> (cit. on p. 19).

Doctoral dissertations and habilitation theses

- [16] T. Bachard. ‘Semantic compression of large image collections’. Université de rennes, 29th Nov. 2024. URL: <https://inria.hal.science/tel-04909266>.
- [17] K. Gu. ‘Spherical light field representation and reconstruction from omnidirectional imagery’. université de rennes; Technische Universität Berlin, 25th Nov. 2024. URL: <https://hal.science/tel-04895833>.
- [18] R. Kaafarani. ‘Optimization of a multi-profile encoding system’. Université de Rennes, 20th June 2024. URL: <https://theses.hal.science/tel-04803100>.

- [19] C. Petit. ‘Data Sampling: Database Compression, Graph Reduction, and Compressed Sensing; Three Aspects of Dimensionality Reduction.’ Université de Rennes, 26th Nov. 2024. URL: <https://inria.hal.science/tel-04826139> (cit. on p. 17).

Reports & preprints

- [20] E. Bannier, S. Castellan, S. Derrien, F. Galassi, L. Garnier, L. Hoyet, A. l’Azou, N. Lahaye, M. J.-M. Macé, O. Martineau, A. Masson, T. Maugey, B. Ninassi, E. Rohou, M. Simonin and F. Taïani. *Reducing GHG emissions from business travel: A collaborative approach at IRISA/Inria*. Groupe de travail « missions » IRISA / Centre Inria de l’Université de Rennes, Mar. 2024, pp. 1–16. URL: <https://univ-rennes.hal.science/hal-04506138>.
- [21] M. Gjorgjevski, N. Keriven, S. Barthelme and Y. D. Castro. *Node Regression on Latent Position Random Graphs via Local Averaging*. 29th Oct. 2024. URL: <https://hal.science/hal-04795829> (cit. on p. 20).

12.3 Cited publications

- [22] J. J. Ahmad, H. A. Khan and S. A. Khayam. ‘Energy efficient video compression for wireless sensor networks’. In: *2009 43rd Annual Conference on Information Sciences and Systems*. IEEE. 2009, pp. 629–634 (cit. on p. 8).
- [23] A. Atalay Appak, E. Sahin, C. Guillemot and H. Caglayan. ‘Learning flat optics for extended depth of field microscopy imaging’. In: *Nanophotonics* (2023) (cit. on p. 9).
- [24] S. Bai, J. Z. Kolter and V. Koltun. ‘dDeep Equilibrium Models’. In: *NEURIPS*. 2019 (cit. on p. 10).
- [25] Y. Blau and T. Michaeli. ‘The perception-distortion tradeoff’. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 6228–6237 (cit. on p. 7).
- [26] M. M. Bronstein, J. Bruna, T. Cohen and P. Veličković. ‘Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges’. In: (2021) (cit. on p. 10).
- [27] B. Bross, Y.-K. Wang, Y. Ye, S. Liu, J. Chen, G. J. Sullivan and J.-R. Ohm. ‘Overview of the versatile video coding (VVC) standard and its applications’. In: *IEEE Transactions on Circuits and Systems for Video Technology* 31.10 (2021), pp. 3736–3764 (cit. on p. 8).
- [28] L. Ceze, J. Nivala and K. Strauss. ‘Molecular digital data storage using DNA’. In: *Nature Reviews Genetics* 20.8 (2019), pp. 456–466 (cit. on p. 6).
- [29] L. D. Chamain, F. Racapé, J. Bégaint, A. Pushparaja and S. Feltman. ‘End-to-End optimized image compression for machines, a study’. In: *2021 Data Compression Conference (DCC)*. ISSN: 2375-0359. Mar. 2021, pp. 163–172. DOI: [10.1109/DCC50243.2021.00024](https://doi.org/10.1109/DCC50243.2021.00024) (cit. on p. 6).
- [30] H. Choi and I. V. Bajić. ‘Scalable Image Coding for Humans and Machines’. In: *IEEE Transactions on Image Processing* 31 (2022). Conference Name: IEEE Transactions on Image Processing, pp. 2739–2754. DOI: [10.1109/TIP.2022.3160602](https://doi.org/10.1109/TIP.2022.3160602) (cit. on p. 6).
- [31] Copernicus. *Access to data*. en. <https://www.copernicus.eu/en/access-data>. Aug. 2023. URL: <https://www.copernicus.eu/en/access-data> (cit. on p. 3).
- [32] G. Cormode, M. Garofalakis, P. J. Haas and C. Jermaine. ‘Synopsis for Massive Data: Samples, Histograms, Wavelets, Sketches’. In: *Foundations and Trends in Databases* 4 (2011), pp. 1–294 (cit. on p. 10).
- [33] Domo. *Data Never Sleeps 10.0*. en. <https://www.domo.com/data-never-sleeps>. June 2022 (cit. on pp. 3, 6).
- [34] L. Duan, J. Liu, W. Yang, T. Huang and W. Gao. ‘Video coding for machines: A paradigm of collaborative compression and intelligent analytics’. In: *IEEE Transactions on Image Processing* 29 (2020), pp. 8680–8695 (cit. on p. 6).
- [35] Ericsson. *Mobility Report*. en. <https://www.ericsson.com/en/reports-and-papers/mobility-report>. June 2023. URL: <https://www.ericsson.com/en/reports-and-papers/mobility-report> (cit. on p. 3).

- [36] Y. Feng and A. Varshney. ‘Signet: Efficient neural representation for light fields’. In: *IEEE/CVF International Conference on Computer Vision (ICCV)*. 2021 (cit. on p. 9).
- [37] W. Gao, S. Liu, X. Xu, M. Rafie, Y. Zhang and I. Curcio. ‘Recent standard development activities on video coding for machines’. In: *arXiv preprint arXiv:2105.12653* (2021) (cit. on p. 6).
- [38] A. Kipnis, S. Rini and A. J. Goldsmith. ‘The Rate-Distortion Risk in Estimation From Compressed Data’. In: *IEEE Transactions on Information Theory* 67.5 (May 2021). Conference Name: IEEE Transactions on Information Theory, pp. 2910–2924. DOI: [10.1109/TIT.2021.3068981](https://doi.org/10.1109/TIT.2021.3068981) (cit. on p. 6).
- [39] B. Le Bon, M. Le Pendu and C. Guillemot. ‘Stochastic Unrolled Proximal Point Algorithm for linear image inverse problems’. In: *EUSIPCO 2023 - 31st European Signal Processing Conference*. Helsinki, Finland, 2023 (cit. on p. 10).
- [40] G. Le Guludec and C. Guillemot. ‘Joint Neural Representation For Multiple Light Fields’. In: *ICASSP 2023 - IEEE International Conference on Acoustics, Speech and Signal Processing*. Ed. by IEEE. Rhodes, Greece: IEEE, June 2023, pp. 1–5. URL: <https://hal.science/hal-04054325> (cit. on p. 9).
- [41] J. Leskovec and C. Faloutsos. ‘Sampling from large graphs’. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2006* (2006), pp. 631–636 (cit. on p. 11).
- [42] D. Lim, H. Maron, M. T. Law, J. Lorraine and J. Lucas. ‘Graph Metanetworks for Processing Diverse Neural Architectures’. In: *International Conference on Learning Representations (ICLR)*. 2024 (cit. on pp. 10, 11).
- [43] A. Loukas. ‘Graph reduction with spectral and cut guarantees’. In: *Journal of Machine Learning Research* 20 (2019), pp. 1–42 (cit. on p. 11).
- [44] ‘Low-complexity video compression for wireless sensor networks’. In: *2003 International Conference on Multimedia and Expo. ICME’03. Proceedings (Cat. No. 03TH8698)*. Vol. 3. IEEE. 2003, pp. III–585 (cit. on p. 8).
- [45] Z. Lu, Y. Liu and M. e. a. Jin. ‘Virtual-scanning light-field microscopy for robust snapshot high-resolution volumetric imaging’. In: *Nat Methods* (2023) (cit. on p. 9).
- [46] V. Masson-Delmotte, P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P. R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock et al. ‘Global warming of 1.5 C’. In: *An IPCC Special Report on the impacts of global warming of 1.5* (2018), pp. 43–50 (cit. on p. 7).
- [47] F. Mentzer, E. Agustsson, J. Ballé, D. Minnen, N. Johnston and G. Toderici. ‘Neural video compression using gans for detail synthesis and propagation’. In: *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXVI*. Springer. 2022, pp. 562–578 (cit. on p. 8).
- [48] F. Mentzer, G. D. Toderici, M. Tschannen and E. Agustsson. ‘High-fidelity generative image compression’. In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 11913–11924 (cit. on p. 8).
- [49] B. e. a. Mildenhall. ‘Nerf: Representing scenes as neural radiance fields for view synthesis’. In: *ECCV*. 2020 (cit. on p. 9).
- [50] S. Mohan. ‘Robust and Interpretable Denoising Via Deep Learning’. PhD thesis. New York University, 2022 (cit. on p. 8).
- [51] S. Mohan, Z. Kadkhodaie, E. P. Simoncelli and C. Fernandez-Granda. ‘Robust and interpretable blind image denoising via bias-free convolutional neural networks’. In: *arXiv preprint arXiv:1906.05478* (2019) (cit. on p. 8).
- [52] Z. Pan, H. Qin, X. Yi, Y. Zheng and A. Khan. ‘Low complexity versatile video coding for traffic surveillance system’. In: *International Journal of Sensor Networks* 30.2 (2019), pp. 116–125 (cit. on p. 8).
- [53] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu and M. Chen. ‘Hierarchical text-conditional image generation with clip latents’. In: *arXiv preprint arXiv:2204.06125* (2022) (cit. on p. 7).

- [54] D. R.-J. G.-J. Rydning, J. Reinsel and J. Gantz. ‘The digitization of the world from edge to core’. In: *Framingham: International Data Corporation* 16 (2018), pp. 1–28 (cit. on p. 6).
- [55] Sandvine. *Global Internet Phenomena Report*. <https://www.sandvine.com/global-internet-phenomena-report-2023>. 2023. URL: <https://www.sandvine.com/global-internet-phenomena-report-2023> (cit. on p. 3).
- [56] R. Sims, S. A. Rehman and M. O. L. et al. ‘Single molecule light field microscopy’. In: *Optica* (2020) (cit. on p. 9).
- [57] R. Torfason, F. Mentzer, E. Agustsson, M. Tschannen, R. Timofte and L. Van Gool. ‘Towards Image Understanding from Deep Compression without Decoding’. In: *Int. Conf. on Learning Representations (ICLR)*. 2018. URL: <http://arxiv.org/abs/1803.06131> (visited on 10/02/2020) (cit. on p. 6).
- [58] J. P. Vizcaíno, F. Saltarin, Y. Belyaev, R. Lyck, T. Lasser and P. Favaro. ‘Learning to Reconstruct Confocal Microscopy Stacks From Single Light Field Images’. In: *IEEE Transactions on Computational Imaging* 7 (2021), pp. 775–788. DOI: [10.1109/TCI.2021.3097611](https://doi.org/10.1109/TCI.2021.3097611) (cit. on p. 9).
- [59] Y. Yang, S. Mandt and L. Theis. ‘An Introduction to Neural Data Compression’. In: *Foundations and Trends in Computer Graphics and Vision* 15.2 (2023), pp. 113–200 (cit. on p. 10).
- [60] R. York and J. A. McGee. ‘Understanding the Jevons paradox’. In: *Environmental Sociology* 2.1 (2016), pp. 77–87 (cit. on p. 8).
- [61] V. Zhirnov, R. M. Zadegan, G. S. Sandhu, G. M. Church and W. L. Hughes. ‘Nucleic acid memory’. In: *Nature materials* 15.4 (2016), pp. 366–370 (cit. on p. 6).
- [62] Z. Zhongming, L. Linong, Y. Xiaona, Z. Wangqiang, L. Wei et al. ‘AR6 synthesis report: Climate change 2022’. In: (2022) (cit. on p. 7).