Contrat doctoral Etablissement 2026/2029 ½ Contrat Formulaire de candidature

Titre de la thèse	**Joint Embedding Predictive Architecture for Optimal Control of Multiple IoT Sensors for Swarm 3D Bioacoustics Fauna Survey**

Ecole Doctorale	ED548 Mer & Sciences
Laboratoire	LIS & IM2NP = CDE & DYNI, et CCSI IM2NP lien à aux deux centres **LIS IM2NP CIAN** et le **CMD**.
Discipline	Informatique, traitement du signal et contrôle, Sciences de la mer
Co-financeur 50%	ANR Sylvania et Pelagos : crédits acquis
Directeurs de Thèse & Encadrants	Directeurs = Nicolas Boizot (35%) + Valentin Gies (35%) Encadrants = Sébastien Paris (30%)

Description du sujet de recherche

Contexte, originalité et pertinence par rapport à l'état de l'art :

Le Reinforcement Learning (RL) appliqué à la robotique et aux systèmes autonomes, tels que les drones ou bouée iot se heurte à plusieurs limitations majeures : inefficacité en termes d'échantillons, difficulté à exploiter des observations de grande dimension (notamment visuelles), et complexité du contrôle à horizon long. Dans ce contexte, les approches dites Joint Embedding Predictive Architectures (JEPA) (LeCun 2022, 2025, Bagatella 2026) émergent comme une alternative prometteuse aux paradigmes classiques, en particulier pour l'apprentissage de représentations utiles au contrôle. Contrairement aux approches traditionnelles basées sur la reconstruction (autoencodeurs, world models pixel-level), les JEPA apprennent à prédire des représentations latentes abstraites. Cette propriété les rend particulièrement adaptées aux tâches de contrôle où l'objectif n'est pas de reconstruire fidèlement le monde, mais d'en extraire les variables pertinentes pour l'action. Par exemple, les ROV ou bouées IoT opérant en milieu sous-marin font face à des contraintes nettement plus sévères que les systèmes aériens ou terrestres : absence de GPS (localisation indirecte), forte atténuation visuelle (dépendance aux capteurs acoustiques), dynamique perturbée (courants, turbulence), observabilité partielle (environnement peu structuré). Les hydrophones permettent d'exploiter des signaux acoustiques (balises, communications, sources passives), mais ces signaux sont : bruités, présence de multi-trajets et dépendants du milieu (température, salinité, bathymétrie). Dans ce contexte, les architectures JEPA offrent une approche pertinente pour apprendre une représentation latente pour le contrôle.

Objectifs : Un des apports majeurs des JEPA est leur capacité à apprendre une représentation compacte à partir de données brutes complexes. Dans le cas d'un ROV avec hydrophones, les observations peuvent inclure des signaux temporels multi-capteurs (Poupard 2021), des spectrogrammes ou différences de phase ou de temps d'arrivée (TDoA).

Un encodeur JEPA peut projeter ces observations dans un espace latent z_t capturant : la direction d'arrivée des sources acoustiques, des indices de distance relative et la structure du champ acoustique environnant. Contrairement à une approche analytique (beamforming classique), cette représentation est robuste au bruit, capable d'intégrer des effets complexes (réflexions, réfractions) et adaptée au contrôle. Dans un cadre JEPA, le modèle apprend une dynamique latente du type $z_{t+1} = f(z_t, u_t)$, où z_t encode à la fois l'état du drone et les informations sur l'environnement acoustique, u_t correspond aux commandes (propulseurs, orientation). Cette dynamique capture implicitement à la fois le déplacement du ROV ou capteur IoT, l'évolution relative des sources acoustiques et les perturbations environnementales. Cela permet de contourner la modélisation explicite, souvent

difficile, de la propagation acoustique sous-marine. Une fois ce modèle appris, il devient possible d'effectuer de la planification directement dans l'espace latent. Voici les différentes étapes du principe associé à JEPA : (1) Observation → encodage z_t ; (2) Simulation de trajectoires futures z_{t+k} via le modèle JEPA ; (3) Optimisation d'une séquence d'actions $\{u_t, ..., u_{t+H}\}$ (4) Application en boucle fermée (type Model Predictive Control, MPC).

Méthodes : Le principe fondamental d'une architecture JEPA repose sur l'apprentissage d'une relation prédictive dans un espace latent (Bardes 2026, Hafner 2020). Plus précisément, deux encodeurs sont utilisés pour projeter respectivement un contexte x_c dans un espace latent h_c et une cible x_t dans un espace latent h_t. Un prédicteur g est ensuite entraîné pour approximer h_t à partir de h_c, en minimisant une distance dans l'espace des représentations. Contrairement aux approches génératives, il n'y a pas de reconstruction explicite de x_t, ce qui permet d'éviter l'apprentissage de détails inutiles (textures, bruit, etc.). Cette approche favorise l'émergence de représentations capturant des invariants structurels et dynamiques du système observé, ce qui est particulièrement pertinent pour les systèmes physiques. Un avantage déterminant des JEPA dans ce contexte est la possibilité d'apprentissage sans supervision. Appliquée à la bioacoustique ou l'acoustique marine en générale, les données exploitables incluent des enregistrements hydrophones lors de missions, des logs de navigation et/ou des données environnementales implicites. Voici deux cas d'usages envisagés dans cette thèse:

A) Localisation passive de source : Un bateau/ROV muni d'hydrophones doit localiser une source acoustique inconnue (cétacé, véhicule). Le JEPA apprend une représentation reliant signatures acoustiques et positions relatives puis le "planner" choisit des actions maximisant l'information (active sensing) afin que le système converge vers la source sans modèle explicite de propagation préalable.

B) Suivi de cible acoustique mobile : Dans le cas d'un suivi (autre drone, animal marin, etc.) : le modèle latent capture la dynamique relative, la planification anticipe les mouvements futurs. Grâce à JEPA, le contrôle reste robuste malgré les pertes de signal temporaires.

Retombées attendues :

A) Optimisation du placement des capteurs : nous proposons de coupler l'approche JEPA avec un objectif de *maximisation de l'information latente et/ou de minimisation de l'incertitude sur les trajectoires individuelles*. Cela conduit au déploiement optimal d'un réseau d'hydrophones (Chouchanne Paris 2012, 2013) qui ne reposerait pas simplement sur la minimisation d'une métrique géométrique (technique de couverture classique). L'optimisation portera donc à la fois sur la discriminabilité latente des sources acoustiques et sur la capacité à séparer plusieurs individus simultanés.

B) Planification d'un ROV pour suivi actif de la mégafaune : Un ROV (Poupard et al 2019, 2020, 2021) équipé d'hydrophones devient un observateur mobile actif qui, contrairement au réseau fixe, peut se déplacer vers l'information mais également il peut améliorer la qualité du signal. Le ROV ne cherche pas seulement à suivre, mais à réduire l'incertitude sur plusieurs individus. Parmi les objectifs réalisables dans l'espace latent, citons : maximiser la séparation des sources, minimiser l'incertitude sur trajectoire, maximiser le SNR des signaux cibles ou maintenir l'observabilité multi-individus.

C) Planification d'un protocole en mer : ici l'objectif est de déterminer la nature des déplacements des cachalots, leur sondes et remontées, en regard du trafic maritime : est-ce que leurs trajectoires se modifient en présence de trafic en surface et si oui, observe t'on des différences de comportement selon le type de trafic ? Si oui, quels sont les indices acoustiques qui permettent aux cachalots de modifier leur trajectoire en fonction du trafic ? Si non, serait-il possible de rajouter une information rayonnante sur les tanker ou ferry pour donner un indice de leur passage à la faune ?
Les cachalots sur les rails maritimes sont seuls, en duo, ou groupe de 1/2 douzaine (ce que l'on constate depuis 20 ans sur zone). Voici un exemple de mode opératoire de base : on vise le déploiement depuis un navire 'amiral' (un catamaran) avec les capacités d'embarquer des bouées (4 ou 6 vue leurs taille de 1m de long, 20 cm de diamètre avec communication Argos / iridium et transmission radio de leur détection). La mesure dynamique devrait être confiée au catamaran qui déploie ce réseau de bouées dérivantes. Une boule de 4 hydros sera plongée depuis le catamaran : c'est elle qui construit un système en dynamique en fonction des bouées dérivantes, sachant que tous les capteurs, bouée et boule depuis le catamaran sont synchronisés à la nanoseconde via le Pulse par Seconde (satellite). La portée des hydrophones est de 6 km environ formant un observatoire très dense permettant : i) la

localisation de source acoustique en maximisant l'information directionnelle (réduction d'incertitude) ; ii) le suivi de balise (minimisant une distance latente à une signature acoustique cible) ; iii) l'évitement de zones bruitées(planification de trajectoires améliorant par exemple le SNR) ; iv) navigation coopérative c'est-à-dire la fusion de signaux entre plusieurs ROV. L'intérêt fondamental est que ces objectifs peuvent être définis directement dans l'espace latent, sans nécessiter une reconstruction explicite du champ acoustique.

Mots clés : Apprentissage auto-supervisé acoustique, Représentation latente pour contrôle sous-marin, Observabilité, Modélisation latente des dynamiques ROV, Navigation basée sur hydrophones

Références :

-LeCun, Y. (2022). A Path Towards Autonomous Machine Intelligence.

-Assran, M. et al. (2023). Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture.

-Bardes, A., Ponce, J., & LeCun, Y. (2022). VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning.

-Hafner, D. et al. (2020). Dream to Control: Learning Behaviors by Latent Imagination.

-Schrittwieser, J. et al. (2020). Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model.

- Bagatella M., Matteo Pirotta, Ahmed Touati, Alessandro Lazaric, Andrea Tirinzoni (2026). TD-JEPA: Latent-predictive Representations for Zero-Shot Reinforcement Learning.

-Arberet, S., Gribonval, R., & Bimbot, F. (2009). A robust method to count and locate audio sources in a multichannel underdetermined mixture. IEEE Transactions on Signal Processing, 58(1), 121-133.

-Albert, C.H.,Yoccoz,N.G., Edwards, T.C.,Graham, C.H.,Zimmermann, N.E. Thuiller, W. (2010) Sampling in ecology and evolution – bridging the gap between theory and practice. Ecography, 33, 1028–1037.

-Chouchane, S Paris, F Le Gland, M Ouladsine (2012) Splitting method for spatio-temporal sensors deployment in underwater systems, European Conference on Evolutionary Computation in Combinatorial

-Chouchane, S Paris, F Le Gland, C Musso, DT Pham (2013) On the probability distribution of a moving target. Asymptotic and non-asymptotic results, 14th International Conference on Information Fusion, 1-8

-Barchasz, V., Gies, V., Marzetti, S., & Glotin, H. (2020). A novel low-power high speed accurate and precise DAQ with embedded artificial intelligence for long term biodiversity survey. In Proc. Acustica Symp.

-Ferrari, M., Glotin, H., Marxer, R., & Asch, M. (2020) Open access dataset of marine mammal transient studies and end-to-end CNN classification, IEEE Int. Conf. on Joint Neural Network

-Fourniol, M., Gies, V., Barchasz, V., Kussener, E., Barthelemy, H., Vauché, R., & Glotin, H. (2018) Low-power wake-up system based on frequency analysis for environmental IoT IEEE Int. Conf on Mechatronic & Emb. Sys. & App. 1-6.

-Gies, V., Marzetti, Barchasz,, Barthélemy, H. Glotin (2020) Ultra-low Power Embedded Unsupervised Learning Smart Sensor for Industrial Fault Classification. IEEE Internet of Things and Intelligence System (IoTaIS) (pp. 181-187)

-Gies V. et al., (2020) Ultra low-power always-on wake-up by pulse pattern adaptive recognition for long term biodiversity monitoring, in IEEE Int. C. on Internet of Things and Intelligence System (IoTAIS)

-Gies V., Fourniol M., Barchasz V., Kussener E., Barthelemy H., Vauché R.,Glotin H., (2021) Ultra low-power 2.5w cmos implementation of a mixed analog-digital wake-up system based on frequency analysis, IEEE NEWCAS pp. 1–6

-Gies V., Marzetti S., P. Best, V. Barchasz, S. Paris (2021) A 30 w embedded real-time cetacean smart detector, Electronics, 10.7

-Gies V., Fourniol M., Barchasz V., Kussener E., Barthelemy H., Vauché R., Glotin H., (2018) Analog ultra low-power acoustic wake-up system based on frequency detection, Int. Conf. on Internet of Things and Intelligence System (IOTAIS), pp. 109–115

-Glotin, H, Mischenko, Giraudet (2015) Joint constraints imposed on multiband time transitivity & doppler-effect differences, for separating, characterizing & locating sound sources, Patent WO2015177172A1 https://patents.google.com/patent/WO2015177172A1/en

-Poupard, M., Ferrari, M., Schluter, J., Marxer, R., Giraudet, P., Barchasz, V., ... & Glotin, H. (2019) Real-time passive acoustic 3d tracking of deep diving cetaceans by small non-uniform mobile surface antenna. IEEE CAcoustics, Speech Signal Processing (pp. 8251-8255)

-Poupard, M., Symonds, H., Spong, P., Glotin, H (2021) Evidence of Intra-group Orca Call RateModulation using a Small-aperture FourHydrophone Array, in Frontiers Journal.

-Poupard, M (2020) Contributions en Méthodes bioacoustiques multiéchelles: spécifiques, populationnelles, individuelles et comportementales, Doctoral dissertation, UnivToulon.

-Stevenson, BC, van Dam‐Bates, P, Young, CKY, Measey, J. (2021) A spatial capture–recapture model to estimate call rate and population density from passive acoustic surveys. Methods Ecol Evol. 12: 432– 442.

-Takeda, R.,Komatani, K. (2016). Sound source localization based on deep neural networks with directional activate function exploiting phase information. IEEE int. conf. on acoustics, speech signal proc. (ICASSP) (pp. 405-409)

-Xiao, X et al (2015) A learning-based approach to direction of arrival estimation in noisy and reverberant environments. IEEE int. C. Acoustics, Speech Signal Proc (ICASSP) (pp. 2814-2818)

Encadrement et conditions matérielles pour le doctorant

Directeurs = Nicolas Boizot(35%) + Valentin Gies (35%)
Encadrants = Sébastien Paris (30%)

Source du cofinancement RH : (table fournie par la DAF Me Anais Cipriani le 26 avril 2026):

Source de financement	Montant	Durée	Début du financement	Fin du financement
RS21-SYLVANA	29 625,00 €	9 mois	01/10/2026	30/06/2027
RC24-PELAGOX	30 225,00 €	9 mois	01/07/2027	31/03/2028
ED548	59 850,00 €	18 mois	01/04/2028	30/09/2029
Total	**119 700,00 €**	**36 mois**		

a) Participation de SYLVANIA ANR pour les tâches :

WP2: Software

2.1: Geometry Optimization for passive acoustics

2.2: Representation Learning

2.3: Unsup/sup Learning

2.4: Localisation +Classification

b) Participation de Pelagos, projet avec Le Parc National de Port Cros / Pelagos pour l'étude des comportements de cétacés

Source financement MATERIEL :

Cette thèse bénéficiera directement de tout le matériel du projet REGION CLETRIAC = 100 Keuros de système d'écoute sous-marine distribuée = 10 bouées d'écoute avec IA et synchronisation Satellite PPS, qui sont construites en 2026-2027. D'autre part les sorties en Mer seront assurées en collaboration avec le programme WhaleWay La voix des cachalots de Longitudes181 et CIAN qui y collaborent depuis 5 ans et pour les 5 années à venir.

Compétences attendues et personnes à contacter :

Diplomé.e Master Informatique IA, Robotique, Mécatronique, ou Traitement du signal.

Personnes à contacter :

Nicolas Boizot et Sébastien Paris, Valentin Gies, prénom.nom @ univ-tln.fr