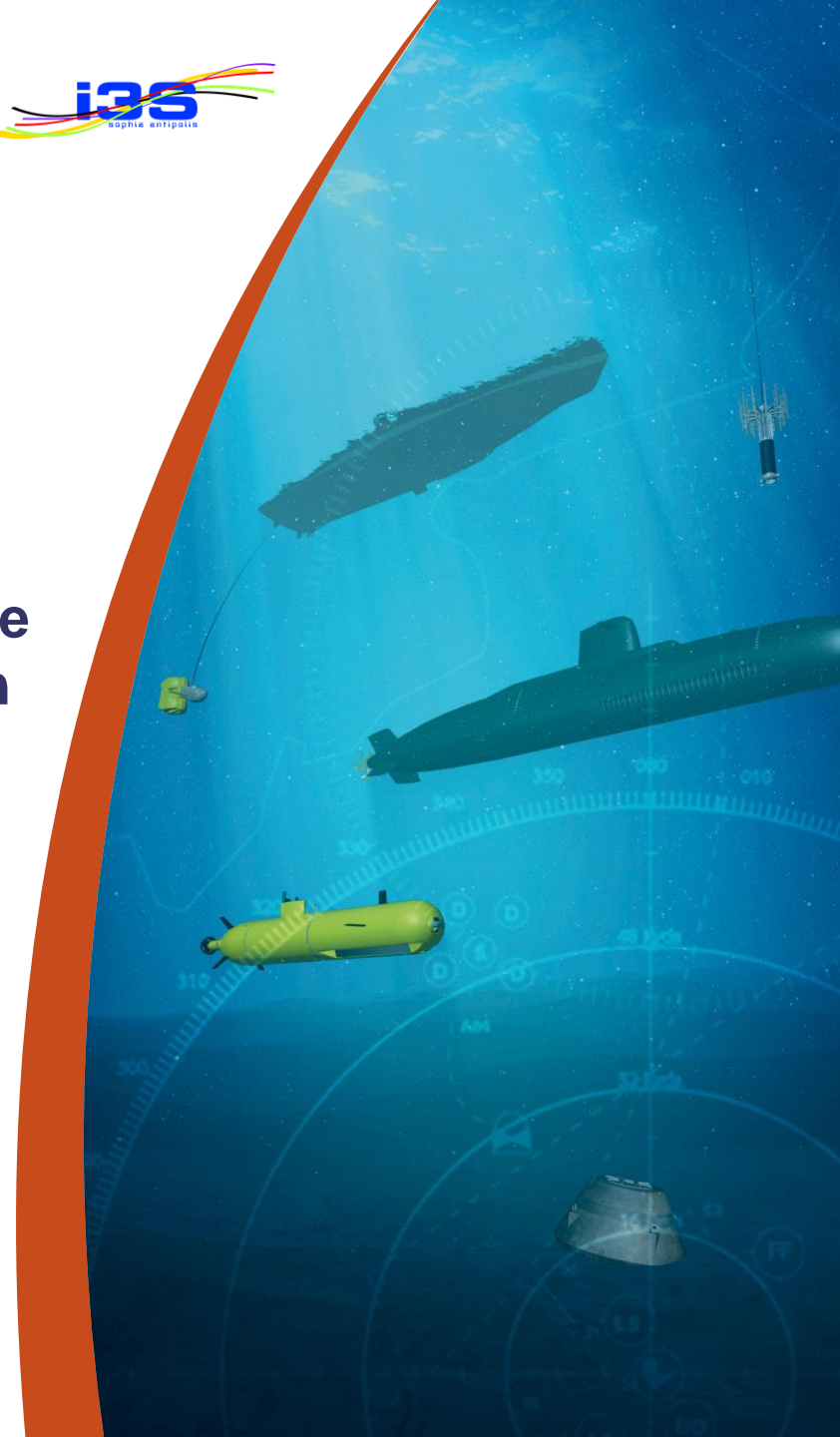


Apprentissage en profondeur fiable et explicable pour la classification acoustique sous-marine

Soutenance de Thomas Mahiout
23 juin 2022

Encadrants :

- **Lionel FILLATRE** (*Université Côte d'azur, I3S, Sophia Antipolis*)
- **Laurent DERUAZ-PEPIN** (*Thales DMS , Sophia Antipolis*)



I. Introduction

II. État de l'art de la classification acoustique

III. Détection & classification de bruits de propulseur

IV. Architectures multi-résolution explicables

V. Conclusions et perspectives

I. Introduction

II. État de l'art de la classification acoustique

III. Détection & classification de bruits de propulseur

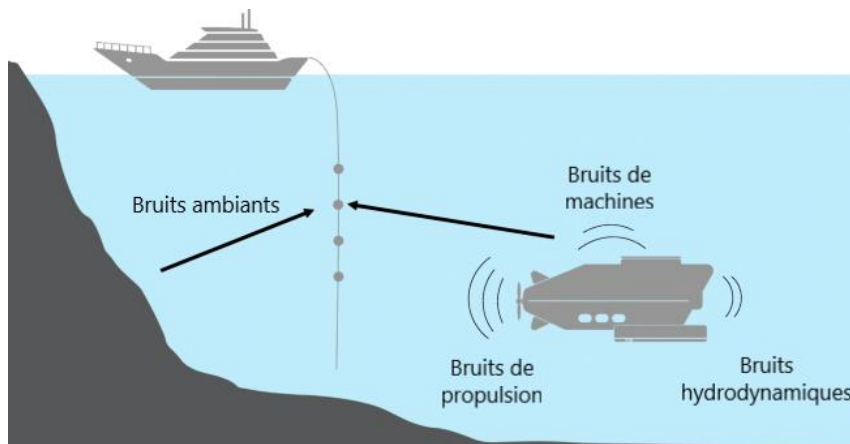
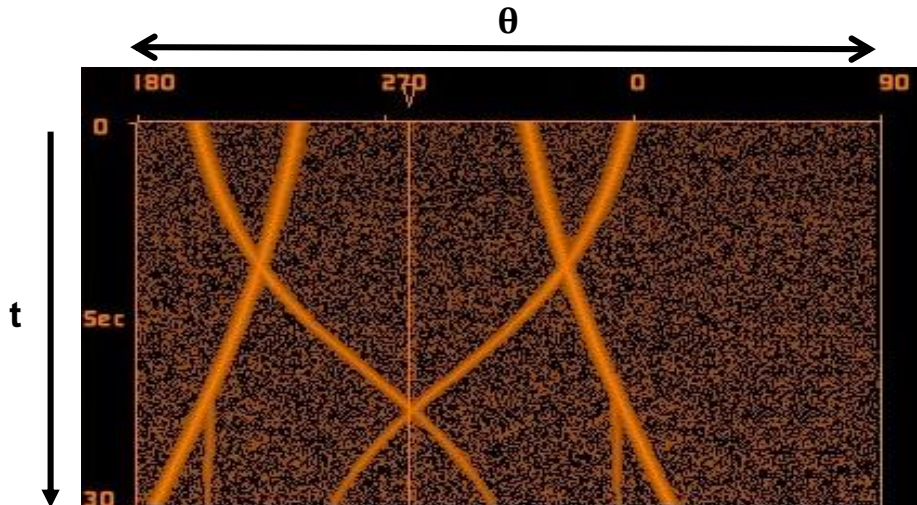
IV. Architectures multi-résolution explicables

V. Conclusions et perspectives

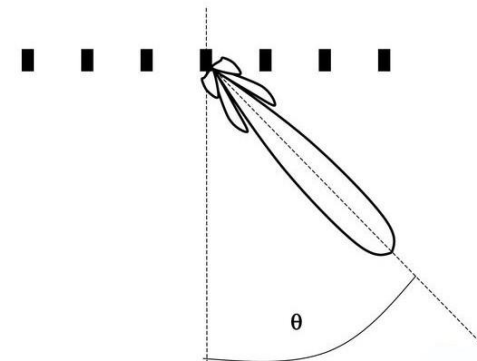
I. Introduction

Contexte opérationnelle de la classification acoustique sous marine

- Après formation de voie sur antenne, suivi de plusieurs cibles potentielles sur waterfall.



Antenne acoustique



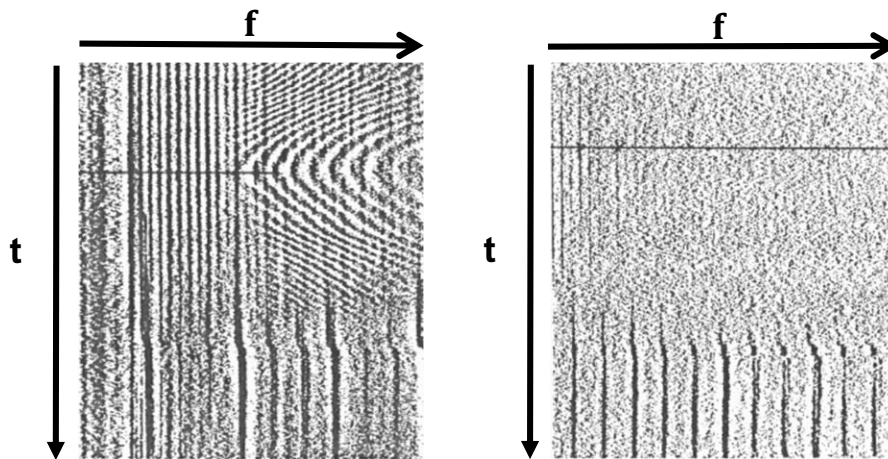
- ASM, environnement complexe où les sources sont noyées dans le bruit ambiant.
- Grande diversité de sources parasites pouvant être détectées.
- Les signaux reçus doivent être classifiés à l'écoute par les marins.

I. Introduction

Contexte de la thèse



- Augmentation du nombre de capteurs sur antennes passives.
- Objectif de différencier automatiquement les détections provenant de bruits de navire et les détections parasites.
- Implique de maîtriser l'environnement acoustique sous-marin et les méthodes actuellement utilisées et comprises par la marine.
- La thèse doit éclaircir deux aspects d'intérêts de l'apprentissage profond : l'explicabilité et la fiabilité.



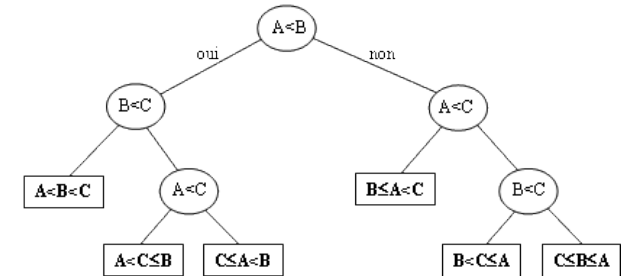
LOFARgram

DEMONgram

I. Introduction

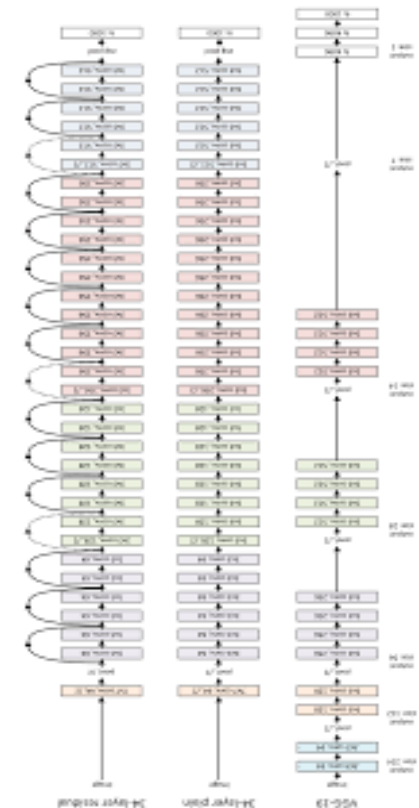
Objectif d'explicabilité

- Recherche d'explicabilité en termes de structure mathématique intrinsèquement explicable plutôt que d'interprétabilité sur boîtes noires.
- Recherche d'un compromis entre explicabilité des boîtes blanches et expressivité des architectures profondes en tirant avantage de leurs profondeurs.



Objectif de fiabilité & robustesse

- Objectif d'établir un lien théorique entre l'architecture d'un réseau profond et la classification/détection optimale de signaux bruités.
- Montrer quelles sont les pertes d'optimalités liées à l'apprentissage et comment les paramètres de l'architecture et de l'apprentissage influencent ces pertes.
- Besoin de maîtriser les performances sur données simulées et de pouvoir déterminer si les entrées sont éloignées de la distribution des données d'apprentissage.



I. Introduction

II. État de l'art de la classification acoustique

III. Détection & classification de bruits de propulseur

IV. Architectures multi-résolution explicables

V. Conclusions et perspectives

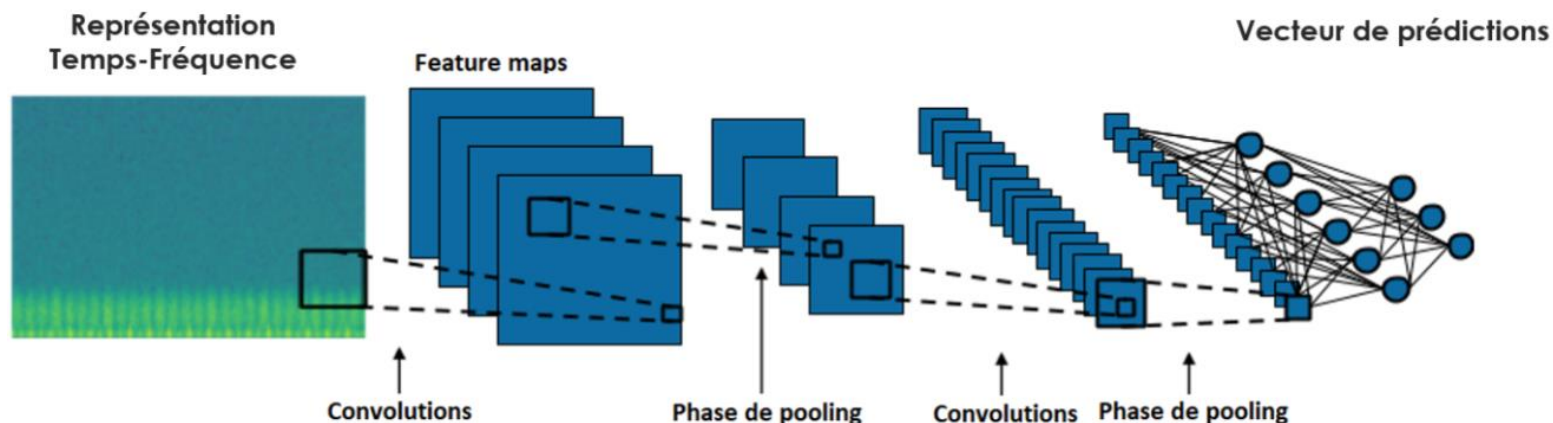
II. État de l'art de la classification acoustique

Classification à partir de caractéristiques extraites

- Utilisation d'une représentation des signaux sous forme de variables (spectre, cepstre, tonalités, chroma, ...) comme entrée d'algorithmes d'apprentissage machine linéaires ou non-linéaires.

Classification à partir de représentations 2D

- Utilisation d'algorithmes de CNN 2D exploitant des représentations issues du traitement du signal :
 - Représentations Temps-Fréquence : Spectrogram (LOFARgram) / Scalogram
 - Représentations Temps-Cepstre : MFCCs
 - Spectre cyclique/Stationnarité d'ordre I : DEMONgram, Scattering Transform



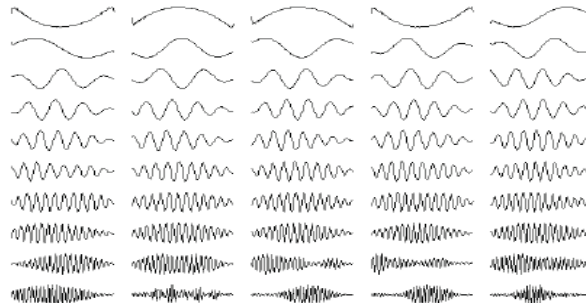
- Choix des représentations et du prétraitement lié aux données à classifier.
- Bonnes performances sur de petits jeux d'entraînement.

II. État de l'art de la classification acoustique

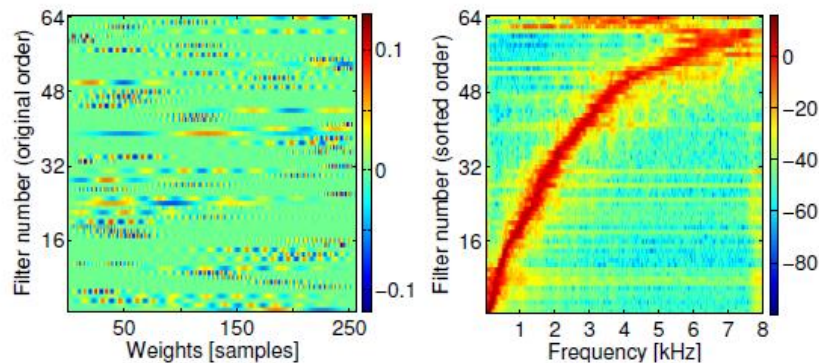
Classification à partir de signaux acoustiques bruts

- Première couche de convolution agissant comme une banque de filtres.

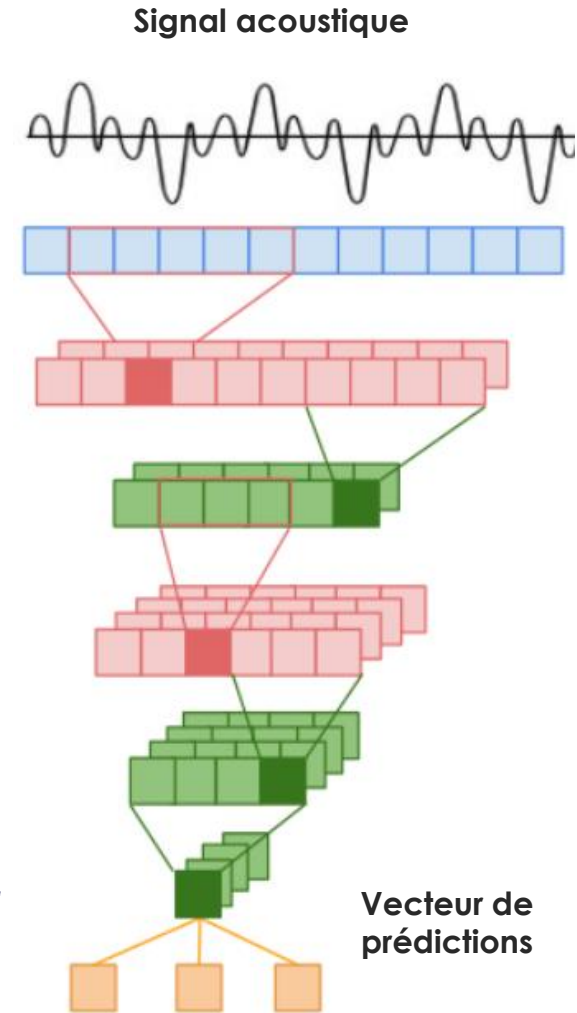
*Sailor,
"Unsupervised filterbank
learning using CRBM
for environmental sound
classification"*



- Les couches suivantes peuvent approximer des structures déjà connues dans le traitement du signal, qu'on peut visualiser lors de la combinaison des features maps.
- Structure mathématique complexe avec plus d'une couche.



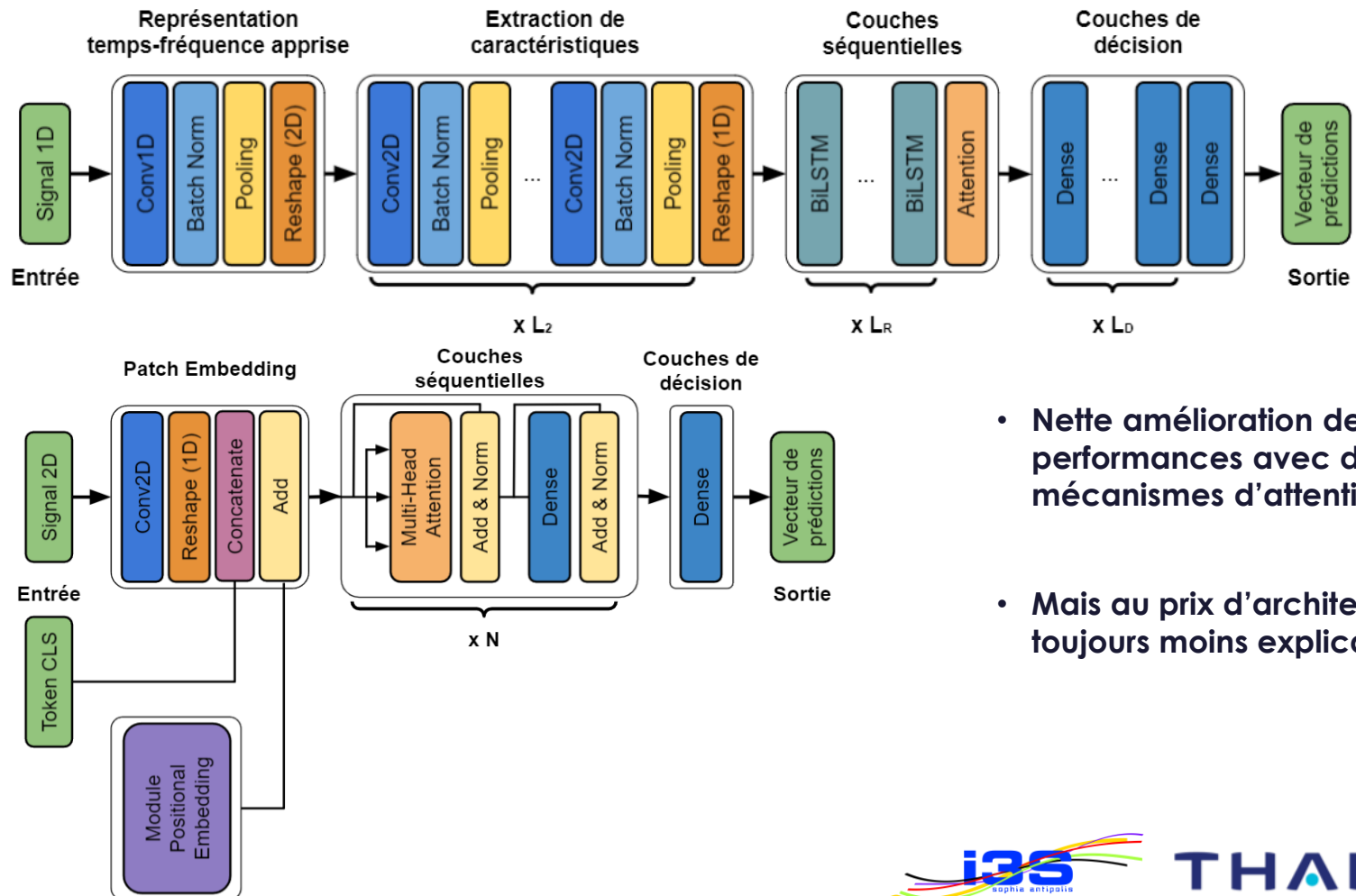
*Golik et al,
"CNN for acoustic
modeling of raw
time signal"*



II. État de l'art de la classification acoustique

Solutions plus récentes

- Existence de nombreuses variantes au CNN 1D pour l'apprentissage d'une représentation 2D : CRBM, CNN 1D et 2D, CRNN, VAE CNN

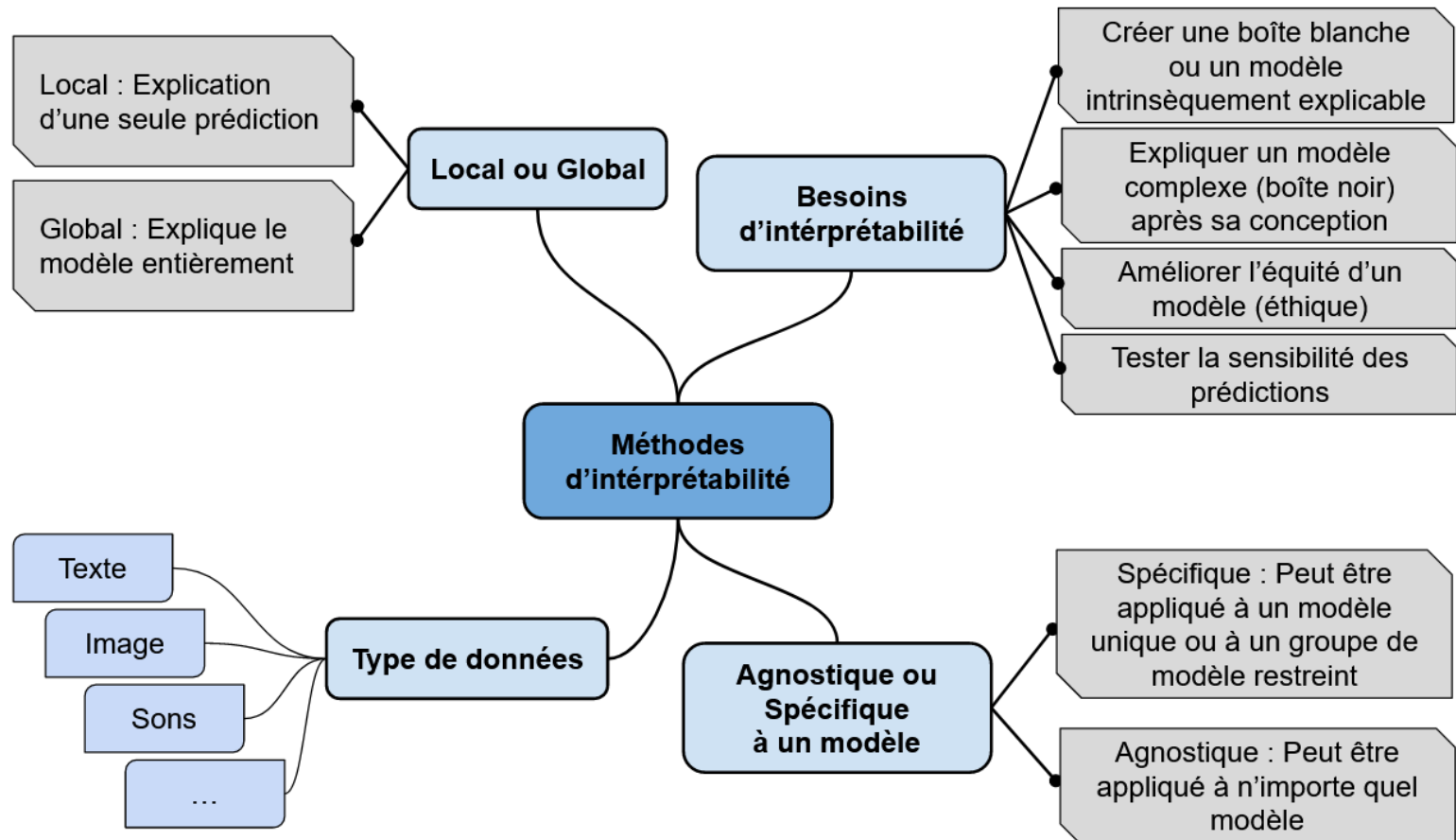


- Nette amélioration des performances avec des mécanismes d'attention.
- Mais au prix d'architectures toujours moins explicables....

II. État de l'art de la classification acoustique

XAI et formes d'explicabilité

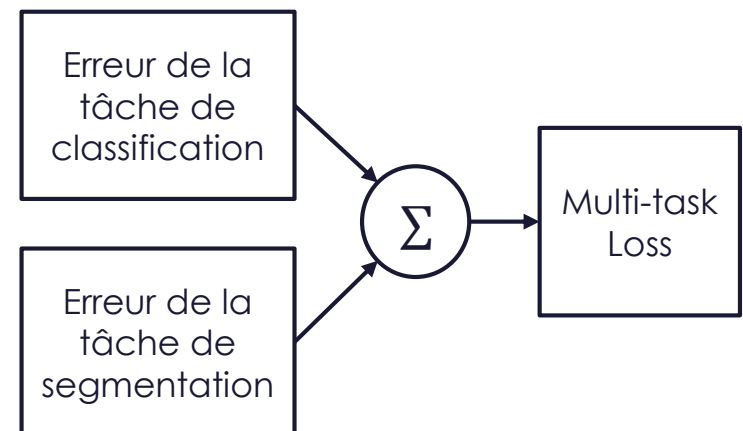
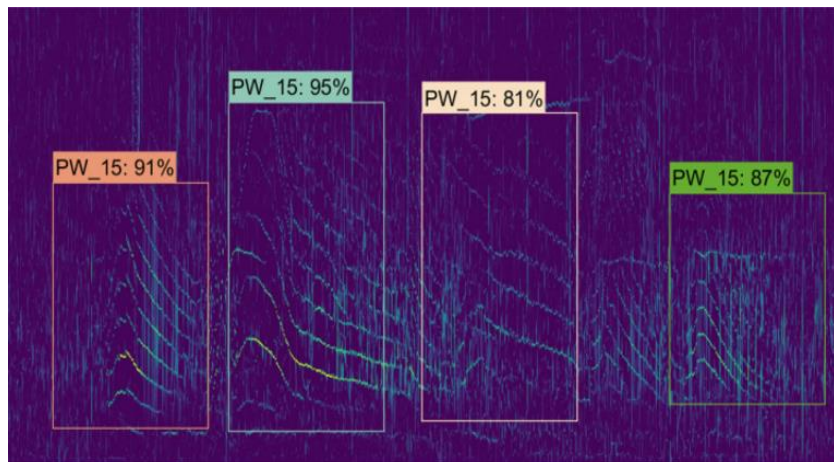
- XAI : capacité d'expliquer ou de présenter en termes compréhensible, un algorithme à un humain.



II. État de l'art de la classification acoustique

Formes d'interprétabilité

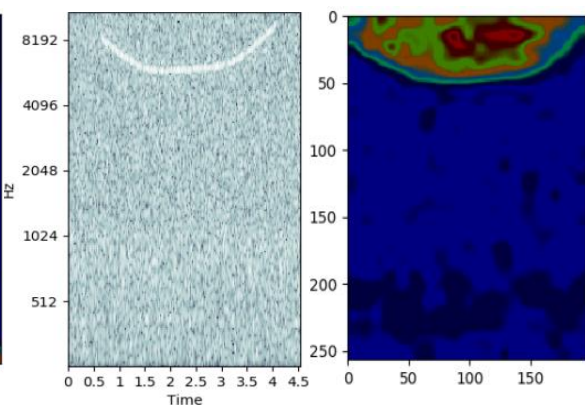
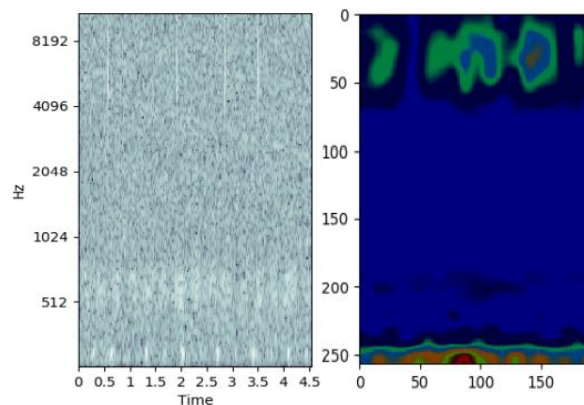
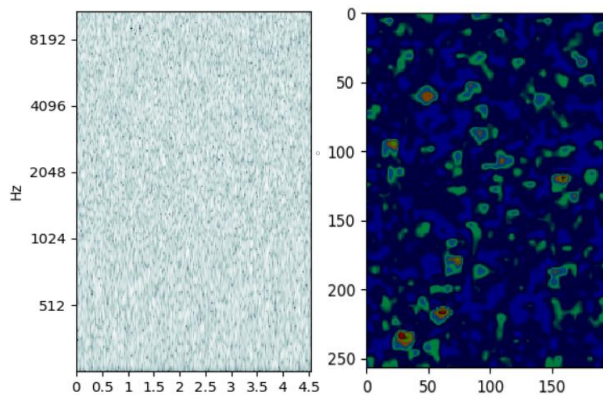
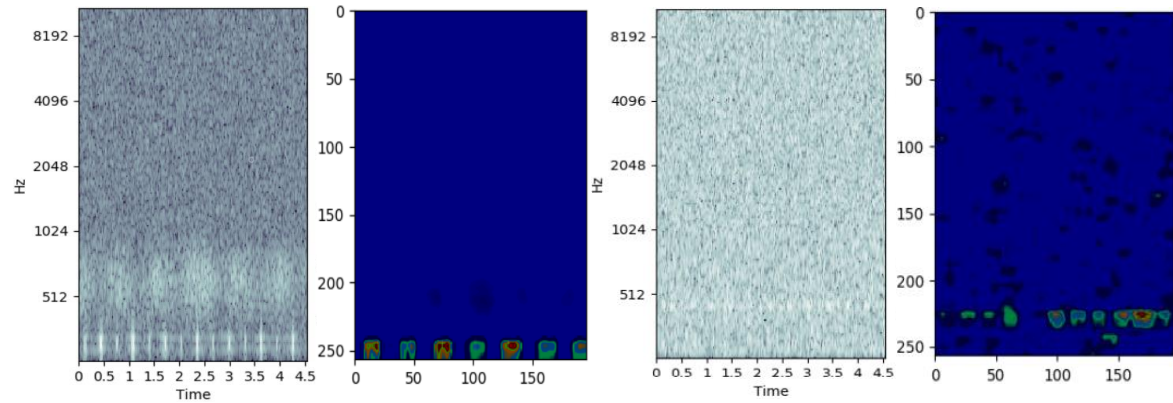
- Méthodes de description : utilisation d'un algo boîte blanche devant décrire localement le comportement d'un algorithme profond (SHAP, LIME).
- Méthodes de visualisation : mettre en évidence les caractéristiques d'une entrée qui entraînent une classification.
- Méthodes restituant des sorties intrinsèquement explicables, où un label et des explications doivent être fournis lors de chaque classification (Multi-task learning).
 - Tentatives réalisées pour prédire le nombre de pales, la vitesse de rotation des pales, mais également une segmentation temps-fréquence.



II. État de l'art de la classification acoustique

Méthodes de visualisation

- Méthodes basées sur la rétro-propagation du gradient d'une prédiction via GradCam.
- Utilisation possible sur modèles CNN 2d où l'information est bien localisée et sur CNN 1d et 2d.
- Ne permet pas d'interpréter l'obtention de meilleures confiances en intégrant sur de longues périodes.



III. Détection & classification de bruits de propulseur

I. Introduction

II. État de l'art de la classification acoustique

III. Détection & classification de bruits de propulseur

IV. Architectures multi-résolution explicables

V. Conclusions et perspectives

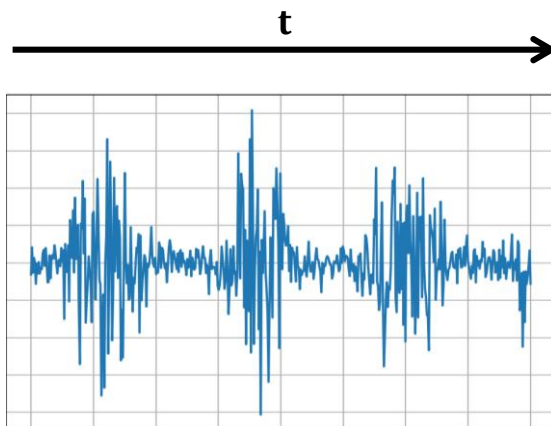
III. Détection & classification de bruits de propulseur

Bruit de cavitation de l'hélice

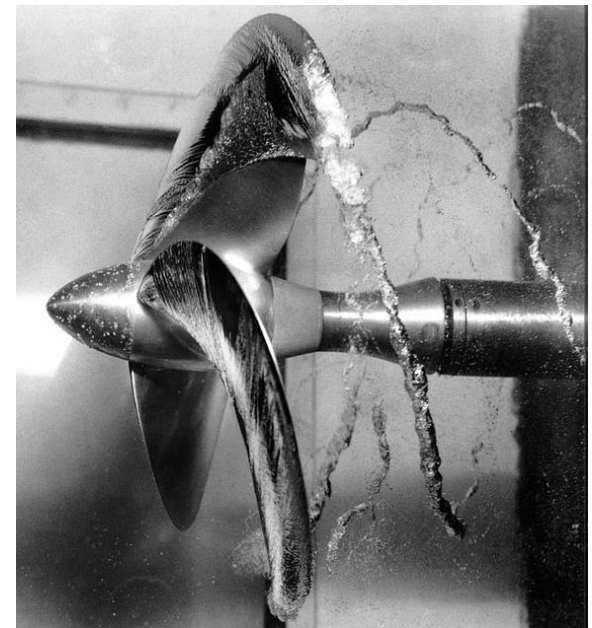
- Signal reçu $x(t) = s(t) + n_a(t)$ contenant un bruit d'hélice et un bruit ambiant blanc et Gaussien.

$$s(t) = \sum_{i=1}^{n_p} m_{i,\omega}(t) \cdot n_{i,c}(t) + n_{0,c}(t)$$

- $n_{i,c}(t)$, bruit de cavitation de la pale d'indexe i .
 - $m_{i,\omega}(t)$, modulation en amplitude de la pale d'indexe i .
 - $n_{0,c}(t)$, contribution du bruit de cavitation constant.
- Bruit cyclostationnaire comme signal cible.



Cavitation d'une hélice



J.G.Lourens "Passive sonar ml estimator for ship propeller speed"

III. Détection & classification de bruits de propulseur

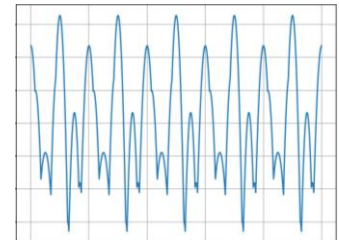
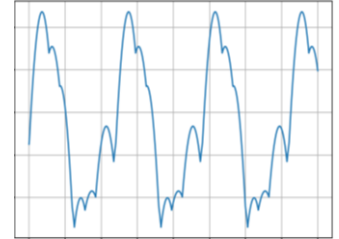
■ Problème de détection de bruit de propulseur

- Détection d'un signal reçu avec une signature inconnu (problème composite) :

$$H_0: \{x(k) \sim \mathcal{N}(0, \sigma_a^2), k \in \llbracket K \rrbracket\},$$

$$H_1: \{x(k) \sim \mathcal{N}(0, \sigma_\theta^2(k) + \sigma_a^2), \theta \in \{\theta_1, \dots, \theta_M\}, k \in \llbracket K \rrbracket\}$$

$$\sigma_\theta^2(k) = \sum_{i=1}^p m_{i,\omega}^2(k) \cdot \sigma_i^2 + \sigma_0^2$$



■ Détecteur/Classifieurs optimal

- Test optimal obtenu à partir du rapport de vraisemblance bayésien $\delta^*(x) = \mathbb{1}_{\Lambda(x) \geq \tau}$:

$$\Lambda^*(x) = c_0 + \sum_{m=1}^M \left(c_m \exp \left(\sum_{k=1}^K b_{m,k} x^2(k) \right) \right)$$

III. Détection & classification de bruits de propulseur

Classe d'architecture d'approximation bayésienne

- δ^* minimise le risque réel pour le problème de détection.

$$\mathcal{R}^* = \min_f \mathcal{R}(f) = \mathcal{R}(\delta^*)$$

- Objectif d'obtenir une convergence structurelle vers le détecteur optimal δ^* et pas juste en terme de risque $\mathcal{R}(g_\theta) - \mathcal{R}^*$.
- Proposition d'une structure d'architectures profondes $f_\theta(x)$, dont les couches approchent les opérations élémentaires du classifieur de Bayes :

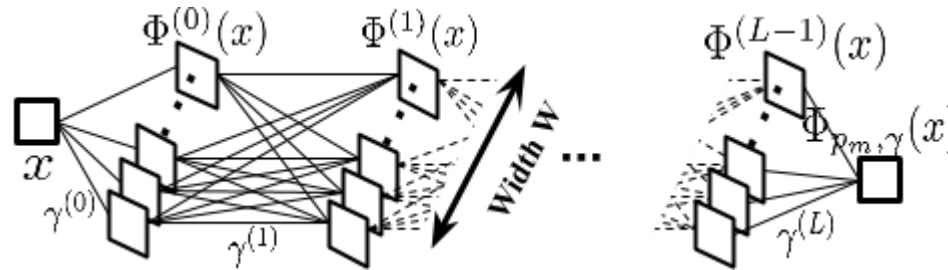
$$\Lambda^*(x) = c_0 + \sum_{m=1}^M \left(c_m \exp \left(\sum_{k=1}^K b_{m,k} x^2(k) \right) \right)$$

$$f_\theta(x) = \lambda_0 + \sum_{m=1}^M \lambda_m \Phi_{e^x, \gamma_2} \left(\sum_{k=1}^K \varphi_{m,k} \Phi_{x^2, \gamma_1}(x(k)) \right)$$

III. Détection & classification de bruits de propulseur

Approximation de fonction par réseau de neurones

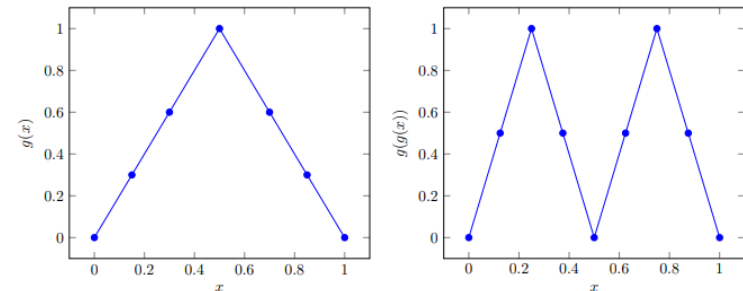
- Réseau de neurones complètement connectés $\Phi_{g,\gamma} : \mathbb{R} \rightarrow \mathbb{R}$ de largeur $W=9$ avec fonction d'activation ReLU :



$$\begin{cases} \Phi^{(0)}(t) = \Gamma_r(\gamma^{(0)} \cdot t + B^{(0)}), t \in \mathbb{R}, \\ \Phi^{(l)}(t) = \Gamma_r(\gamma^{(l)} \Phi^{(l-1)}(t) + B^{(l)}), 1 \leq l < L-1 \\ \Phi_{g,\gamma}(t) = \gamma^{(L)} \Phi^{(L-1)}(t) + B^{(L)}, \end{cases}$$

- Possibilité d'approximer une fonction g avec un nombre de couche L qui croît de façon polylogarithmique par rapport à l'erreur d'approximation désirée:

$$\|\Phi_{g,\gamma} - g\|_{L^\infty(I)} \leq \epsilon$$



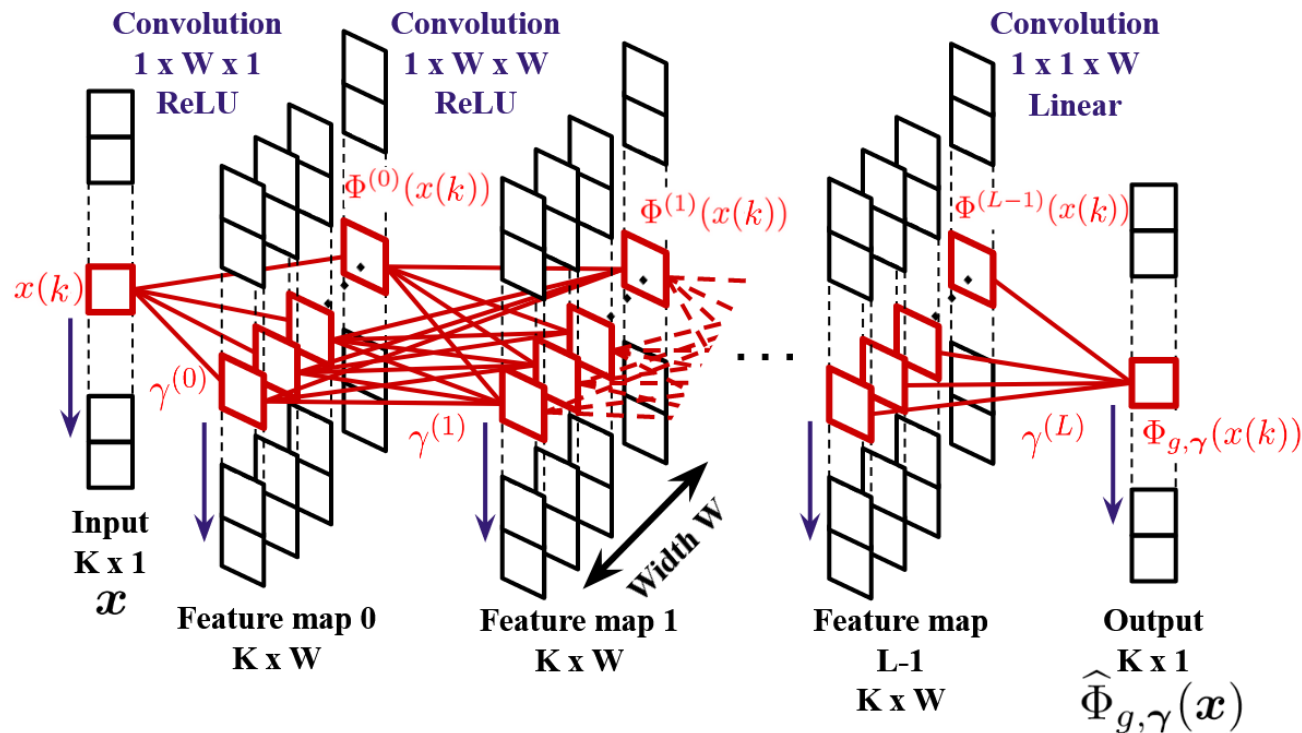
- Théorème montrant que pour $f_\theta(x)$, on a :

$$\sup_{x \in \mathcal{X}} |f_\theta(x) - \Lambda^*(x)| \leq C' \epsilon$$

III. Détection & classification de bruits de propulseur

Approximation de fonction par CNN

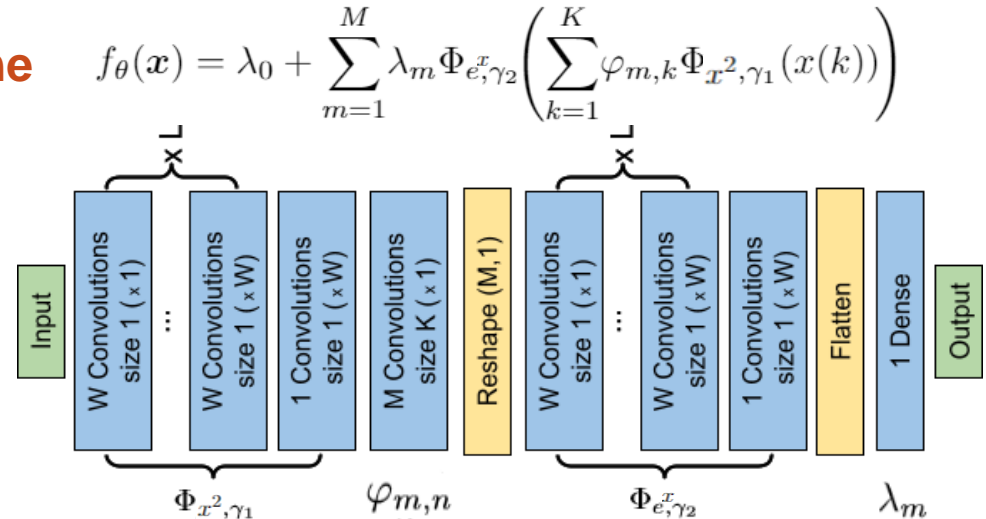
- Adaptation du réseau de neurones $\Phi_{g,\gamma} : \mathbb{R} \rightarrow \mathbb{R}$ en tant que réseau convolutif avec des filtres de tailles 1×1 .
- Calcul efficaces et explicable des fonctions vectorielles.



III. Détection & classification de bruits de propulseur

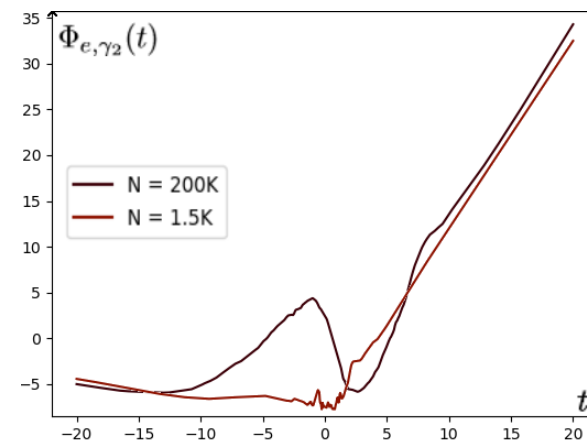
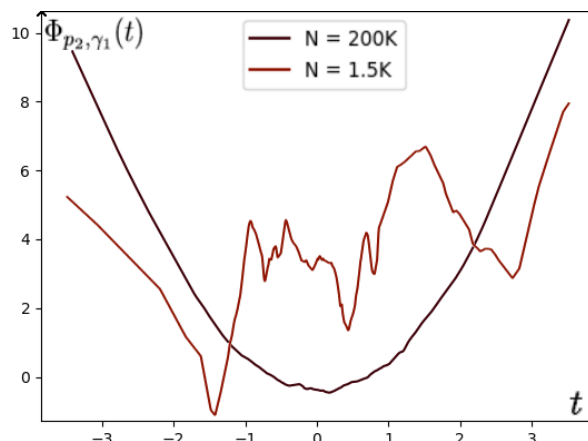
Architecture profonde bayésienne

- Implémentation de l'architecture $f_{\theta}(x)$ en tant que CNN.



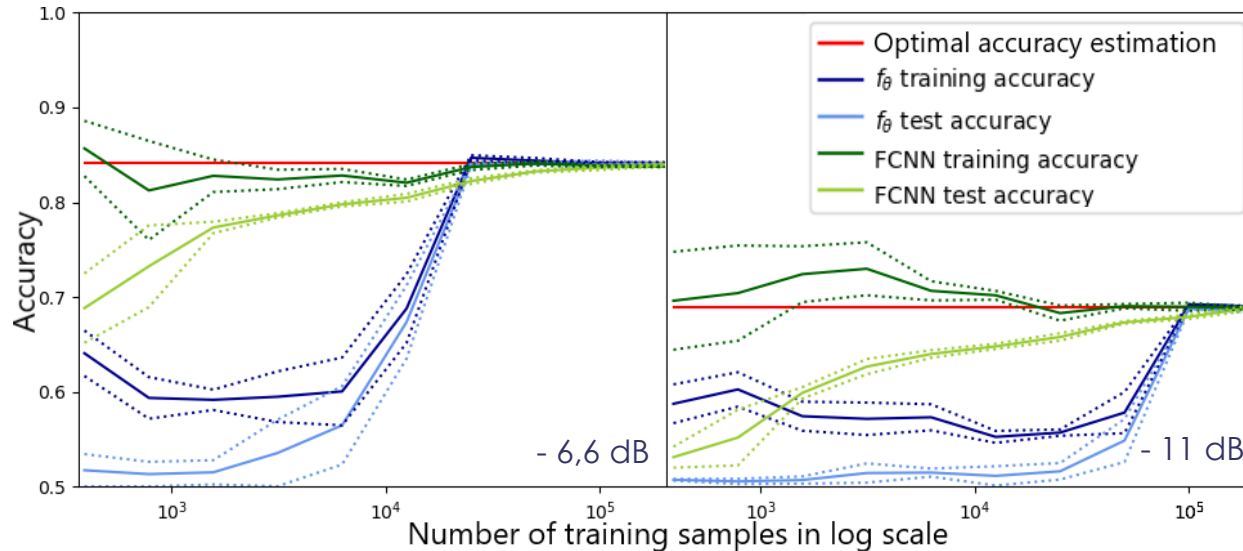
Visualisation des fonctions apprises

- Possibilité d'extraire les poids des fonctions apprises pour une visualisation.

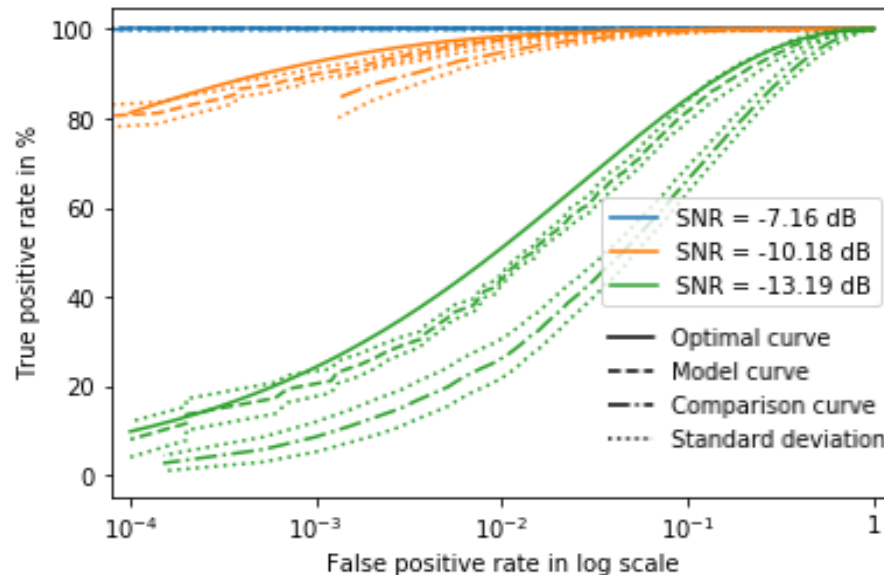


III. Détection & classification de bruits de propulseur

Influence du SNR sur la généralisation



- Possibilité de converger vers les performances et même la structure mathématique du détecteur optimal.
- Nombre d'exemples nécessaire lié au SNR dans les données d'apprentissage.
- Cas avec une seule signature.
- Possibilité de tracer des courbes COR.

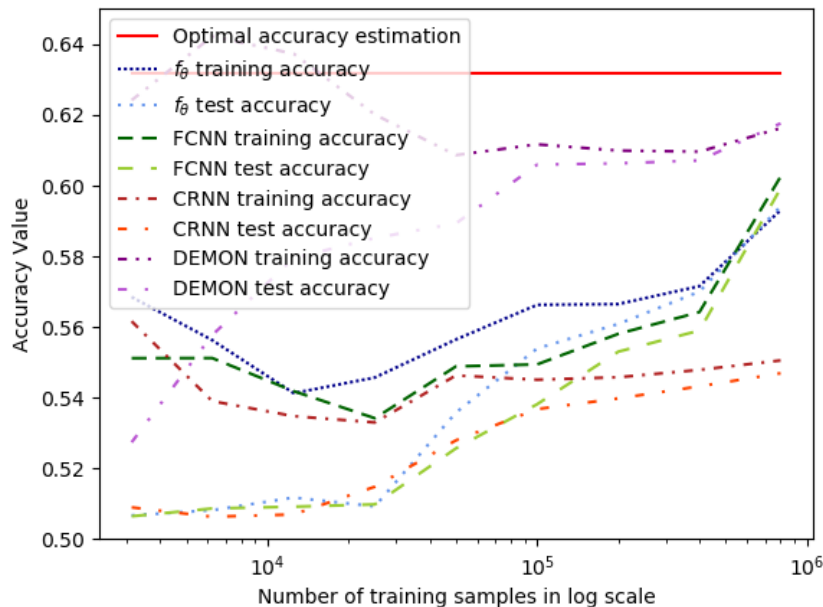


III. Détection & classification de bruits de propulseur

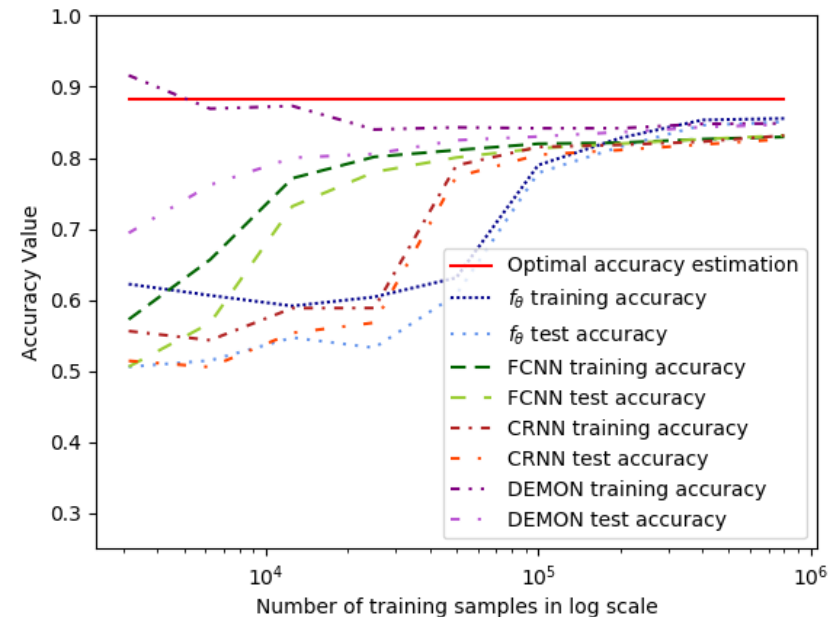
Résultats de généralisation

- Comparaison des performances de $f_{\theta}(x)$ (3 k) au détecteur optimal, à un modèle FCNN 1D (1 m), à un CNN 1D sur DEMON (90 k) et un CRNN 1D (380 k).

K = 64 échantillons



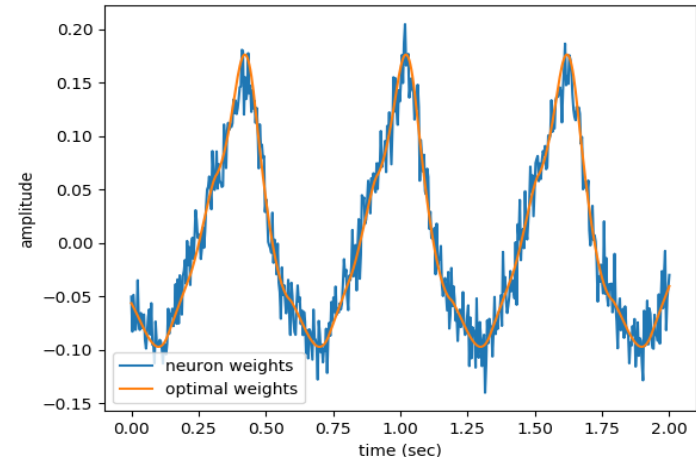
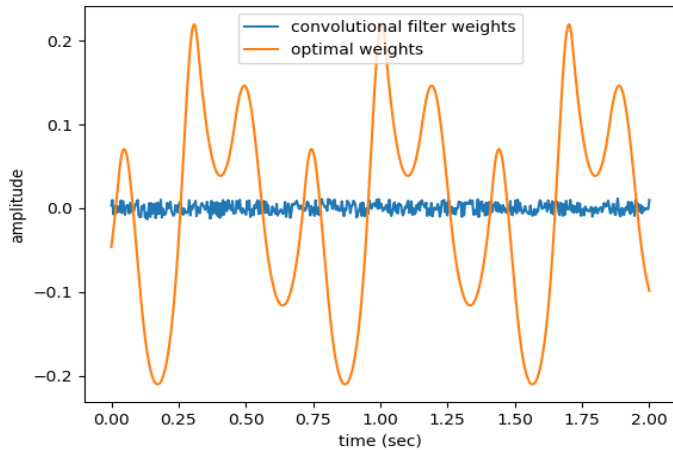
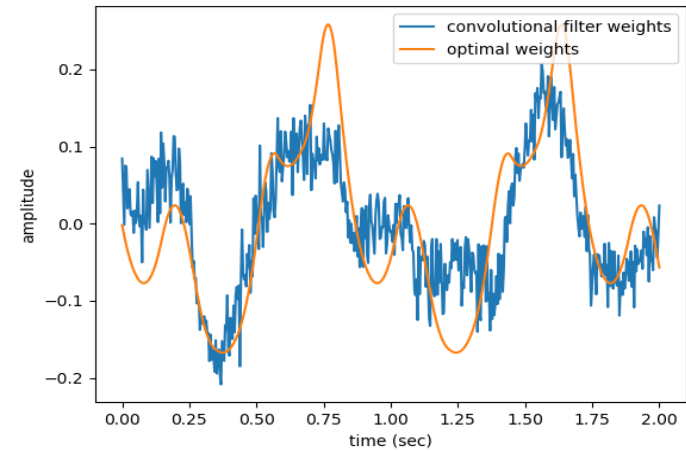
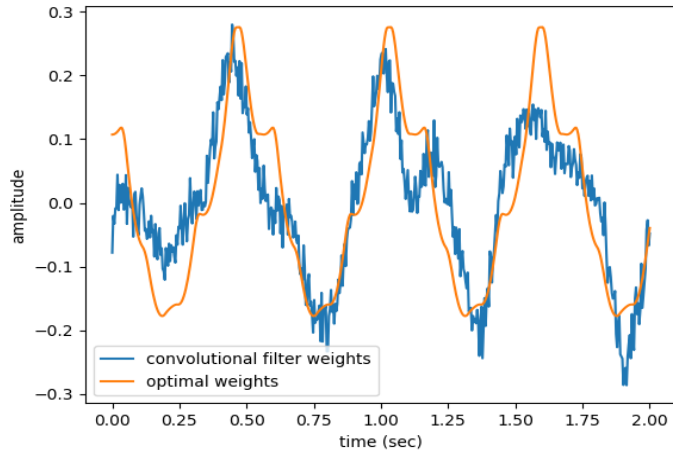
K = 512 échantillons



III. Détection & classification de bruits de propulseur

Autres résultats

- Apprentissage d'une banque de filtres bruités même avec un nombre de signature M petit.



IV. Architectures multi-résolution explicables

I. Introduction

II. État de l'art de la classification acoustique

III. Détection & classification de bruits de propulseur

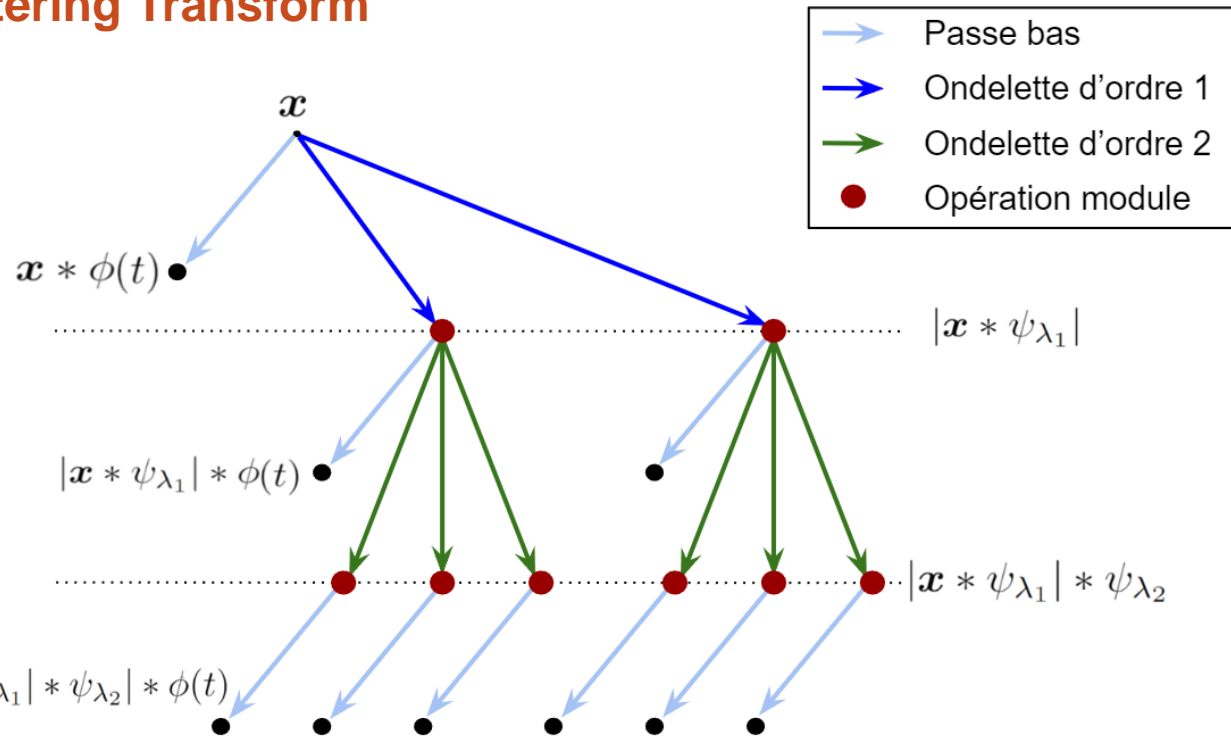
IV. Architectures multi-résolution explicables

V. Conclusions et perspectives

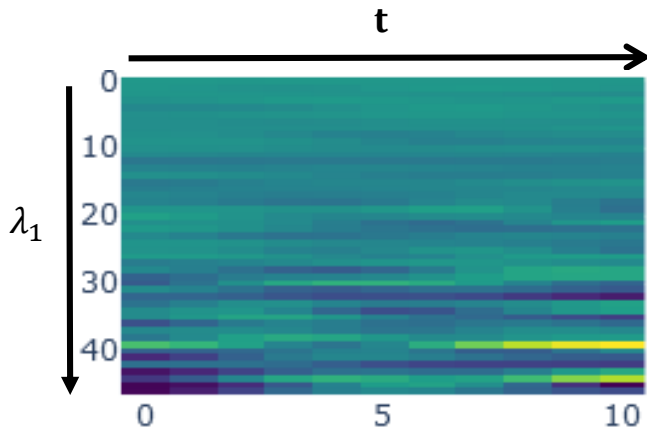
IV. Architectures multi-résolution explicables

Algorithme de la Scattering Transform

- Permet d'analyser le spectre de modulation du signal à plusieurs ordres grâce à des décompositions sur différentes échelles temporelles.
- Calcul à partir d'une cascade de convolutions de filtres d'ondelettes et des opérations module.
- Adaptée aux signaux cyclostationnaires de cavitation de pales



- Représentation de $|x * \psi_{\lambda_1}| * \phi(t)$ avec des fenêtres temporelles comme un scalogramme.



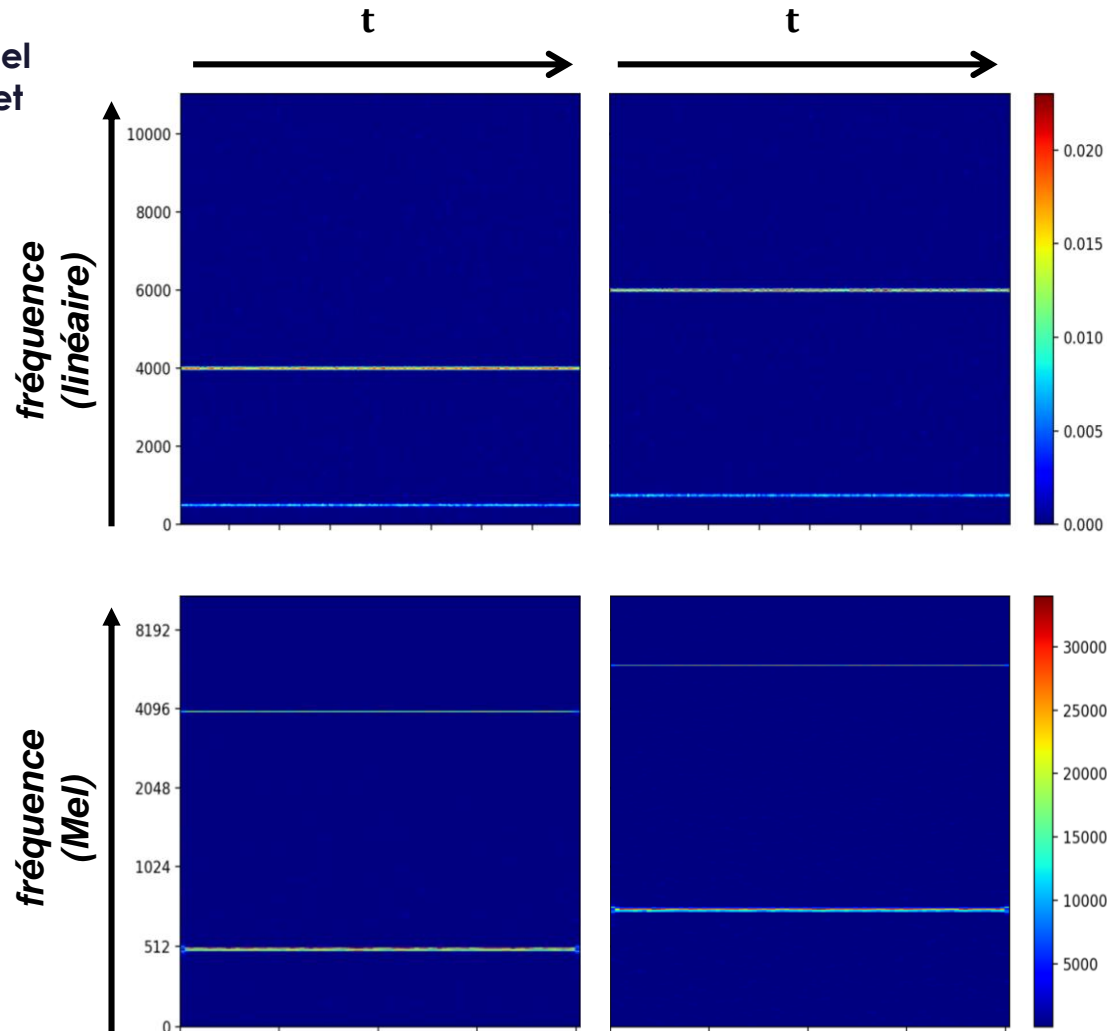
IV. Architectures multi-résolution explicables

Algorithme de la Scattering Transform

- Définit une représentation en échelle Mel localement invariante aux translations et stable aux déformations temporelles.

$$\begin{aligned} \|Sx_\tau - Sx\| &\leq C \sup_t |\tau'(t)| \cdot \|x\| \\ x_\tau(t) &= x(t - \tau(t)) \\ |\tau'(t)| &< 1, \sup_t |\tau(t)| \ll T \end{aligned}$$

- Kymatio : calcul efficace avec des convolutions dans le domaine fréquentiel.
- Basée sur des ondelettes complexes fixes.



IV. Architectures multi-résolution explicables

Architecture de Scattering Transform Apprise (LST) CNN d'ordre 1

- Utilisation de la structure de l'algorithme de transformée en ondelette rapide pour apprendre un ensemble de F filtres de tailles réduites à différentes échelles de résolution au lieu de Q version d'une même ondelette mère par octave.

$$\tilde{S}[x](s, f) = |x_{\dagger s} \star \psi_f| \star \phi_K$$

$$x_{\dagger s+1} = (x_{\dagger s} \star \psi_{LF}) \star \phi_2$$

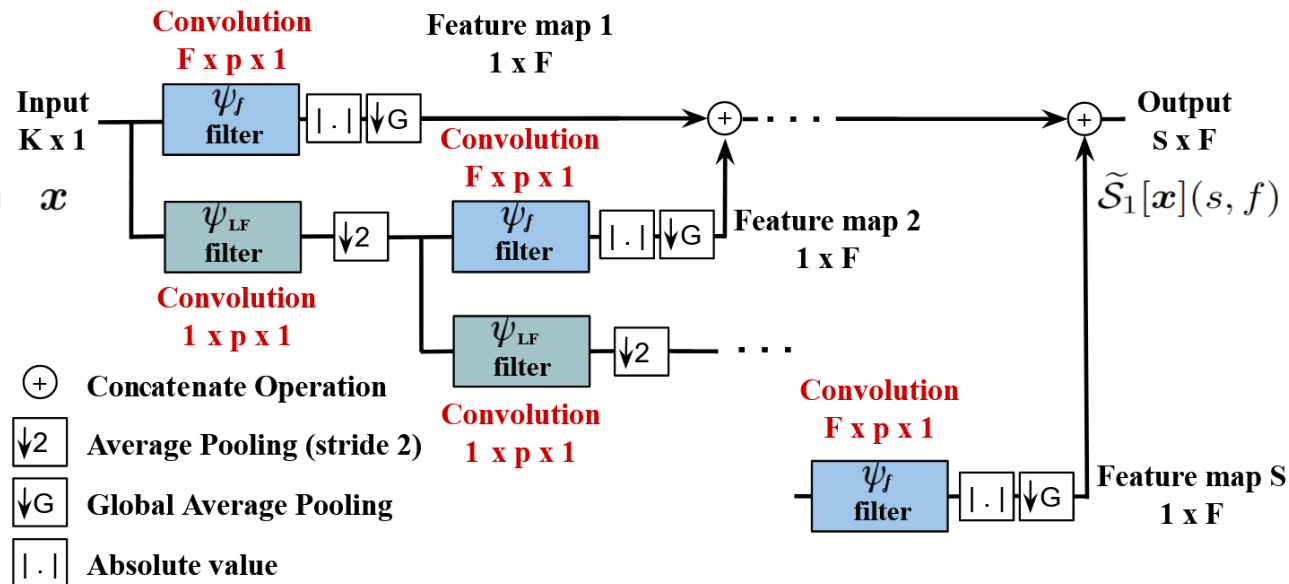
- Forme d'apprentissage frugal comparé aux CNN 1D grâce aux skip connections.

- Possibilité d'obtenir une représentation dépendant du temps en utilisant des opérations average pooling locales, pour obtenir :

$$\tilde{S}[x](s, f, t) =$$

$$|x_{\dagger s} \star \psi_f| \star \phi_T(t)$$

- Représentation similaire à un Spectrogram d'échelle Mel multi canal appris.



IV. Architectures multi-résolution explicables

Résultat de généralisation avec une architecture LST d'ordre 1

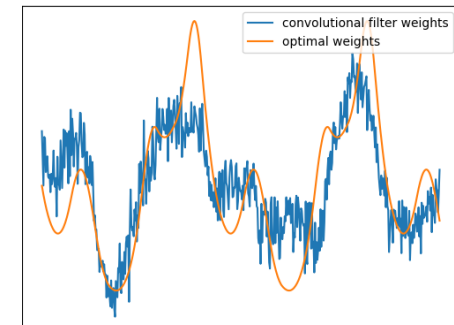
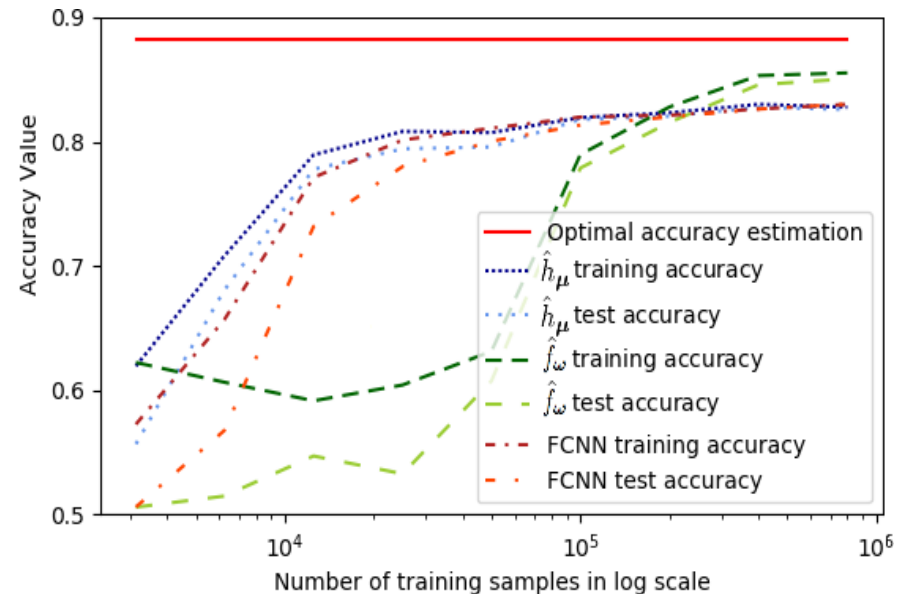
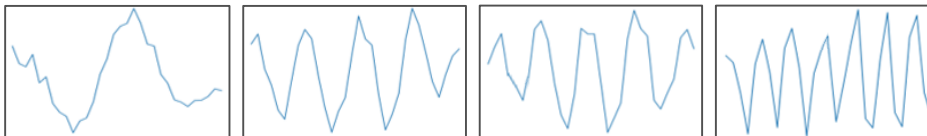
- Apprentissage des filtres adaptés dans un espace appris de dimension réduite:

$$f_{\theta}(x) = \lambda_0 + \sum_{m=1}^M \lambda_m \Phi_{e, \gamma_2} \left(\sum_{k=1}^K \varphi_{m,k} \Phi_{x^2, \gamma_1}(x(k)) \right)$$

$$\hat{h}_{\mu}(x) = \alpha_0 + \sum_{m=1}^M \alpha_m \Phi_{e, \gamma_2} \left(\sum_{\lambda \in \Lambda} \beta_{m,\lambda} \tilde{\mathcal{S}}[\Phi_{x^2, \gamma_1}(x)](\lambda) \right)$$

- Représentation des filtres temporels $b_{m,k}$, de tailles K dans un nouvel espace appris de taille $S \times F$, avec S tel que :

$$2p > \frac{K}{2^S} \geq p$$
$$S \approx \ln_2(K) - \ln_2(p)$$



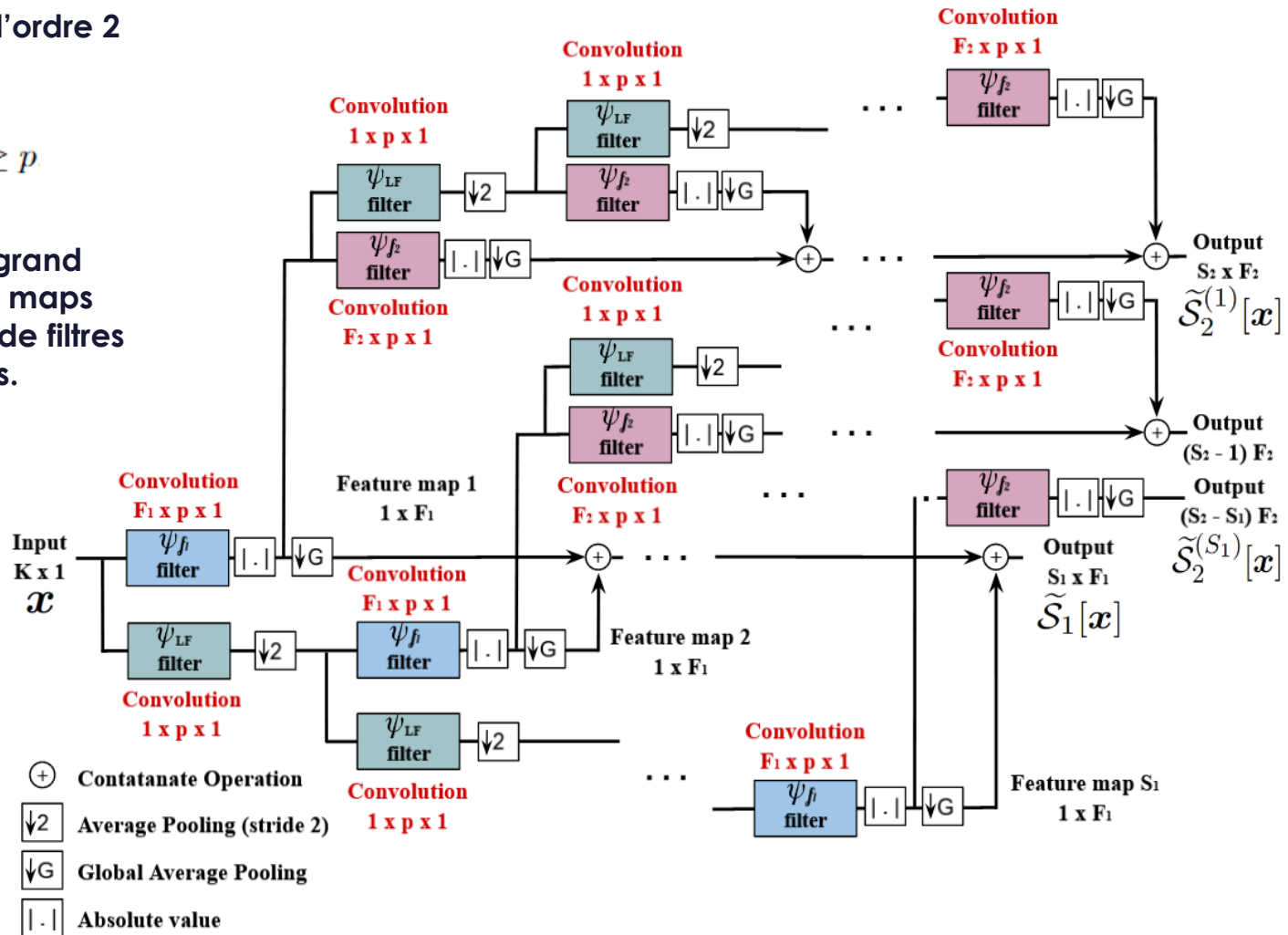
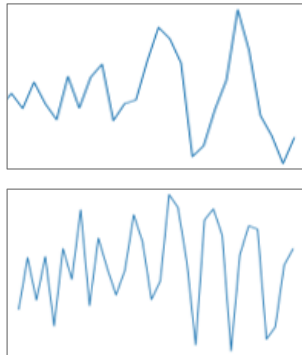
IV. Architectures multi-résolution explicables

Architecture de Scattering Transform apprise (LST) CNN d'ordre 2

- Décomposition à l'ordre 2 avec S_2 tel que :

$$2p > \frac{K}{2^{S_2-s-1}} \geq p$$

- Création d'un très grand nombre de feature maps limitant le nombre de filtres pouvant être appris.
- Comportement d'apprentissage imprévisible.



V. Conclusions et perspectives

I. Introduction

II. État de l'art de la classification acoustique

III. Détection & classification de bruits de propulseur

IV. Architectures multi-résolution explicables

V. Conclusions et perspectives

V. Conclusions et perspectives

Conclusions

- Développement d'un modèle simple de bruit d'hélice et d'un simulateur pour la génération de données permettant de comparer des solutions retenus dans l'état de l'art.
- Proposition d'une méthode d'explicabilité liée à la théorie de l'approximation via des architectures bayésiennes intrinsèquement explicables.
- Adaptation de la structure de la Scattering Transform pour concevoir des architectures CNN 1D capables d'apprendre des représentations internes explicables.

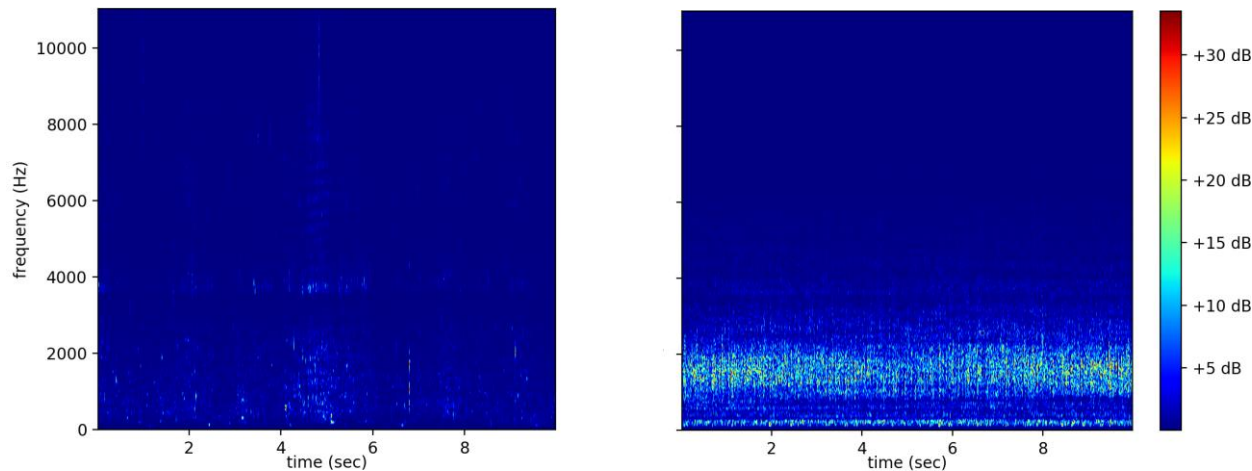
Publications

- Propeller Noise Detection with Deep Learning, Thomas Mahiou, Lionel Fillatre, Laurent Deruaz-Pepin, IEEE ICASSP 2020, Spain.
- Explainable Deep Learning Detection of Gaussian Propeller Noise with Unknown Signal-to-Noise Ratio, Thomas Mahiou, Lionel Fillatre, Laurent Deruaz-Pepin, IEEE MLSP 2021, Australia.
- Learning based Scattering Transform for Explainable Underwater Acoustic Classification, Thomas Mahiou, Lionel Fillatre, Laurent Deruaz-Pepin, IEEE EUSIPCO 2022, Serbia.

V. Conclusions et perspectives

Comparaisons de différentes architectures LST CNN

- Comparaison sur des jeux de données Open Sources contenant des enregistrements acoustiques de navires.



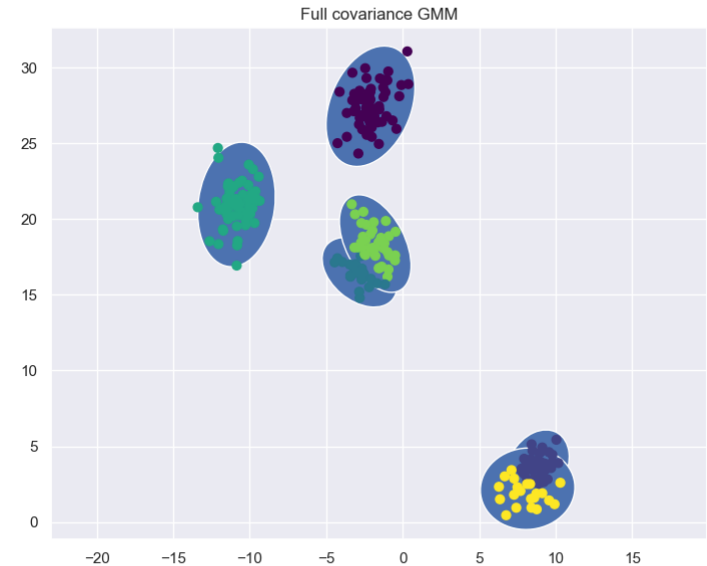
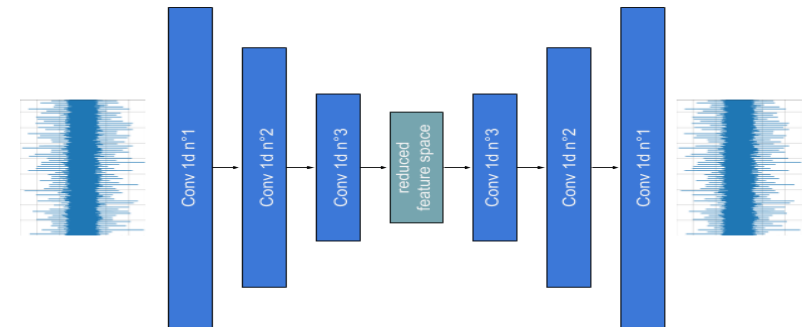
Pistes de recherche

- Modification de la back propagation pour inclure Q filtres de tailles différentes par octaves afin d'obtenir une meilleur résolution fréquentielle sans apprendre des filtres supplémentaires.
- Stratégie d'initialisation avec des filtres d'ondelettes classiques.

V. Conclusions et perspectives

Apprentissage non supervisé

- Possibilité d'apprendre de façon non supervisé des représentations internes sur des jeux bien plus vastes.
- Apprentissage par VAE / GAN.
- Classifications via des distances au sein d'une base de données.
- Classification à base de clusters avec GMMs, afin de rejeter des données éloignées de la distribution du jeu d'entraînement.



Prise en compte du contexte opérationnel

- Prise en compte d'information lors du pistage...

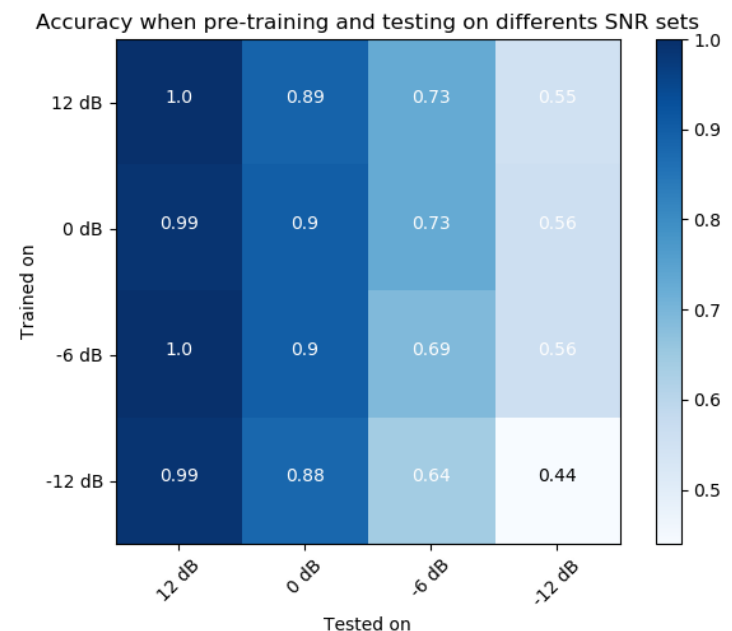
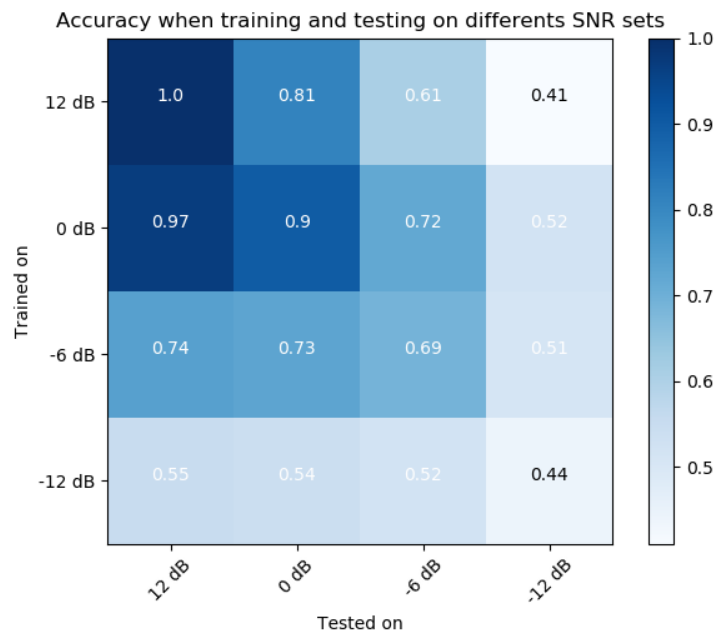
**Merci pour votre
attention**

Résultats sur la classification multi-classes

- Comparaisons de différentes architectures classiques.

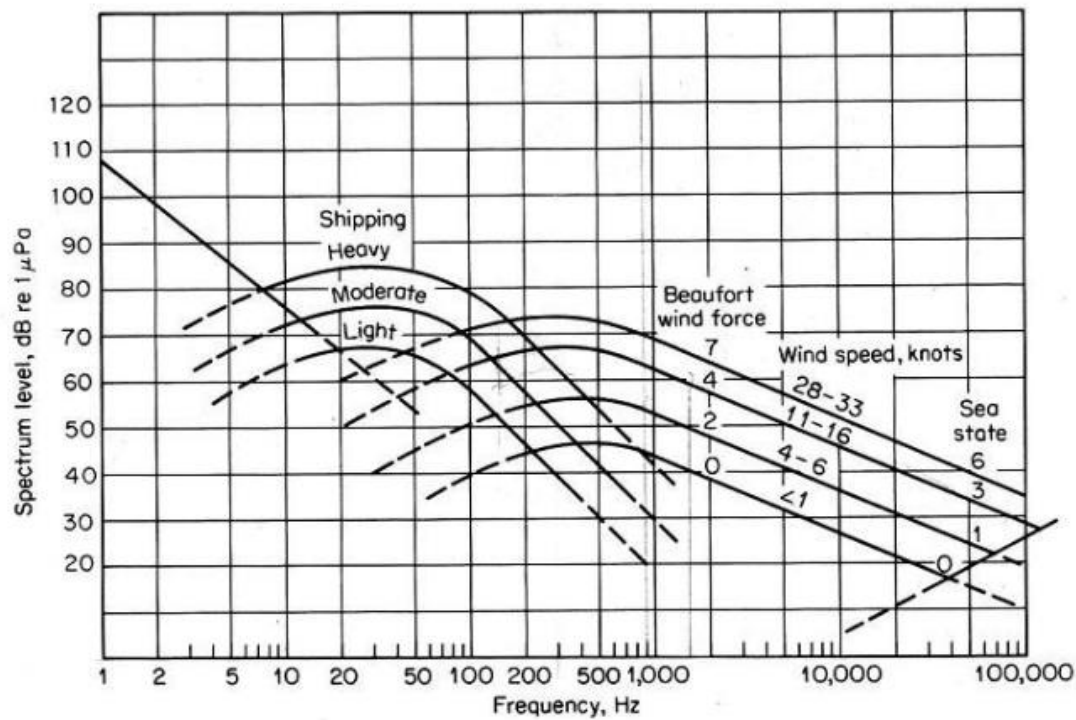
Architecture	Précision de classification Entraînement (%) $\pm \sigma$	Précision de classification Test (%) $\pm \sigma$
FCNN M11	94,75 \pm 0.09	94,29 \pm 0.36
EnvNet-v2	88,79 \pm 0.14	85,92 \pm 0.63
LST-CNN	94,14 \pm 0.08	93,68 \pm 0.16
CNN 2D	94,17 \pm 0.02	94,02 \pm 0.28

- Entraînement de modèles FCNN 1D sur des SNR différents.



Bruit ambiant

- Bruit quasi constant dans toute les directions.



Annexe. Théorie de l'apprentissage

Apprentissage automatique supervisé

- Recherche d'une fonction $g_\theta(\mathbf{x})$ par la minimisation du risque empirique $\mathcal{R}_N(\theta)$ sur un jeu d'entraînement $\mathcal{S} = \{(\mathbf{x}^{(1)}, y^{(1)}), \dots, (\mathbf{x}^{(N)}, y^{(N)})\}$ où $(\mathbf{x}, y) \sim P$.

Contrôle du risque

$$\mathcal{R}_N(\theta) := \frac{1}{N} \sum_{i=1}^N \mathcal{L}(g_\theta(\mathbf{x}^{(i)}), y^{(i)})$$

- Convergence de $g_\theta(\mathbf{x})$ vers le détecteur optimal $\delta(\mathbf{x})$ mesuré par l'excès de risque $\mathcal{R}(g_\theta) - \mathcal{R}^*$.

$$\mathcal{R}(f) := \mathbb{E}_{(\mathbf{x}, y) \sim P} [\mathcal{L}(f(\mathbf{x}), y)]$$

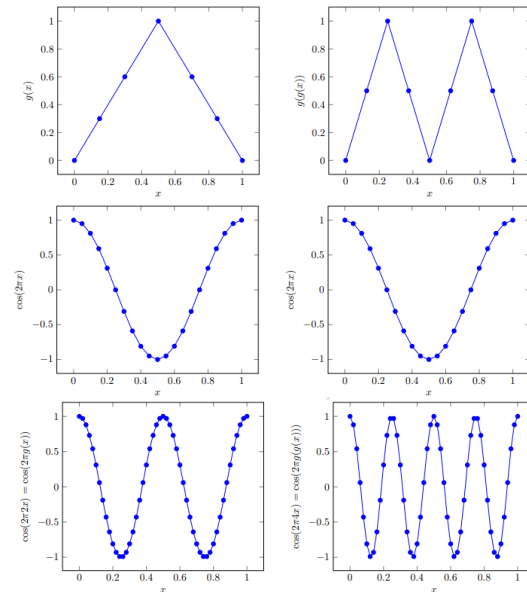
$$\mathcal{R}^* = \min_f \mathcal{R}(f) = \mathcal{R}(\delta)$$

- Décomposition en différents types d'erreurs que l'on va essayer d'influencer avec l'architecture, le nombre d'exemples d'entraînement, etc...

$$\begin{aligned} 0 \leq & \mathcal{R}(g_\theta) - \mathcal{R}^* & = & \text{(excès de risque)} \\ & \mathcal{R}(g_{\theta^*}) - \mathcal{R}^* & + & \text{(erreur d'approximation)} \\ & \mathcal{R}_N(g_{\theta^*}) - \mathcal{R}(g_{\theta^*}) & + & \text{(erreur de généralisation 1)} \\ & \mathcal{R}_N(g_{\theta'}) - \mathcal{R}_N(g_{\theta^*}) & + & \leq \epsilon \\ & \mathcal{R}_N(g_\theta) - \mathcal{R}_N(g_{\theta'}) & + & \text{(erreur d'optimisation)} \\ & \mathcal{R}(g_\theta) - \mathcal{R}_N(g_\theta) & + & \text{(erreur de généralisation 2)} \end{aligned}$$

■ Erreur d'approximation

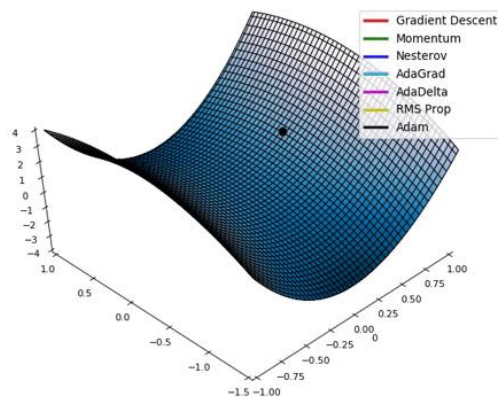
- Erreur $\mathcal{R}(g_{\theta^*}) - \mathcal{R}^*$, avec $g_{\theta^*} = \arg \min_{g_{\theta} \in \mathcal{G}} \mathcal{R}(g_{\theta}) \in \arg \min_f \{\mathcal{R}(f)\}$.
- Capacité d'approximation de la classe de fonction \mathcal{G} liée à l'architecture du réseau. Théorie sur l'approximation de fonction continue par morceaux (Cybenko), Fourier, mais également feature map invariantes (Yarotsky).



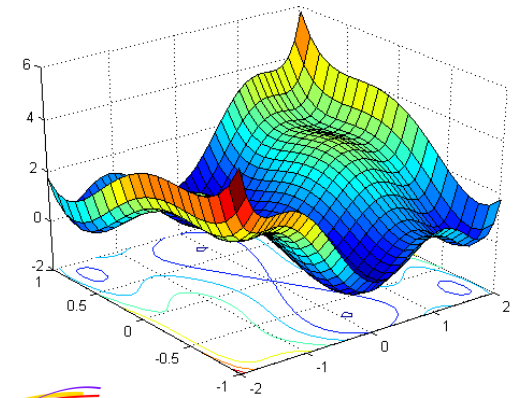
■ Erreur d'optimisation

- Erreur $\mathcal{R}_N(g_{\theta}) - \mathcal{R}_N(g_{\theta'})$, avec $g_{\theta'} = \arg \min_{g_{\theta} \in \mathcal{G}} \mathcal{R}_N(g_{\theta})$.
- Optimisation non-convexe et convergence liée au point de départ, la taille des mini-batches ou encore la fonction de perte utilisée.

$$\mathcal{R}_N(g_{\theta}) := \frac{1}{N} \sum_{i=1}^N \mathcal{L}(g_{\theta}(x^{(i)}), y^{(i)})$$



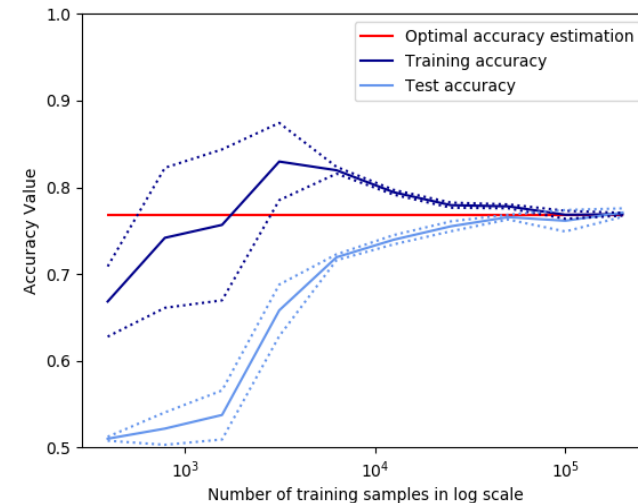
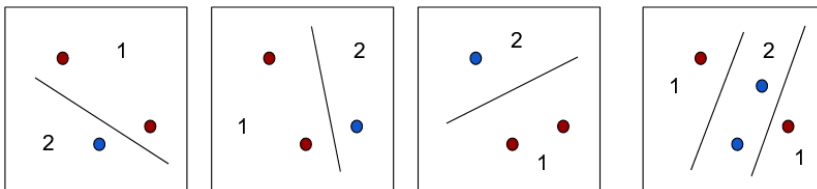
- Augmenter N aboutit généralement à des paysages plus lisses pour le risque empirique.
- Importance de la SGD pour sa régularisation implicite (Roberts).



■ Erreur de généralisation

- Erreur $\mathcal{R}(g_\theta) - \mathcal{R}_N(g_\theta)$, liée à des données indépendantes mais suivant toutes la même distribution $(\mathbf{x}, y) \sim P$.
- Capacité des architecture à surapprendre un problème.
- Existence de bornes liées à la capacité de pulvérisation d'une architecture : on a avec une probabilité $1 - \delta$:

$$\forall g_\theta \in \mathcal{G}, \mathcal{R}(g_\theta) \leq \mathcal{R}_N(g_\theta) + 2\sqrt{2 \cdot \frac{h \log(\frac{2eN}{h}) + \log(\frac{2}{\delta})}{N}}$$



- Ne suffit pas à expliquer certains comportements contre-intuitifs observés dans l'apprentissage profond.
- L'erreur de généralisation est entièrement déterminée par le triplet (P, D_N, g_θ) .
- Besoins de repenser la généralisation (Zhang et al.) avec des mesures de complexité.

Mesures de complexité

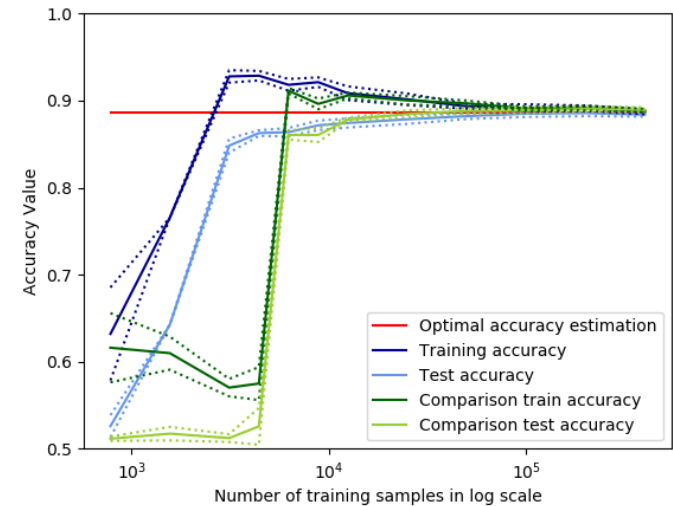
- Tentative de quantifier la complexité de la fonction que représente un réseau (prédicteur de son degré de généralisation) :

- Extension de mesures classiques : Dimension VC, Rademacher, Norme des poids ...

- PAC Bayes et Sharpness

$$\approx \max_{|\nu_l| \leq \alpha(|w^{(l)}|+1)} \mathcal{R}_N(f_{w+\nu}) - \mathcal{R}_N(f_w)$$

- Observation empirique : Fisher-Rao, distance des poids à l'initialisation, Vitesse de convergence, Ampleur du bruit du gradient ...



- Théorise l'influence des choix de paramètres de réseau et d'apprentissage sur la généralisation.
- Possibilité de vérifier la cohérence de ces mesures en comparant leur évolution théorique et empirique avec une GridSearch sur un jeu de donnée (Neyshabur et al.).
- Certains aspects, comme les propriétés d'invariance des architectures, sont encore difficilement quantifiables.