Automatic Modulation Recognition

Seminar CyberExcellence UMONS 2023

Ir Alexander Gros

Electromagnetism and Telecommunication Department
Faculty of Engineering
University of Mons

alexander.gros@umons.ac.be

May 10, 2023

Outline

Outline

Subject: what exactly ?

AMR → Automatic modulation recognition → a long history !

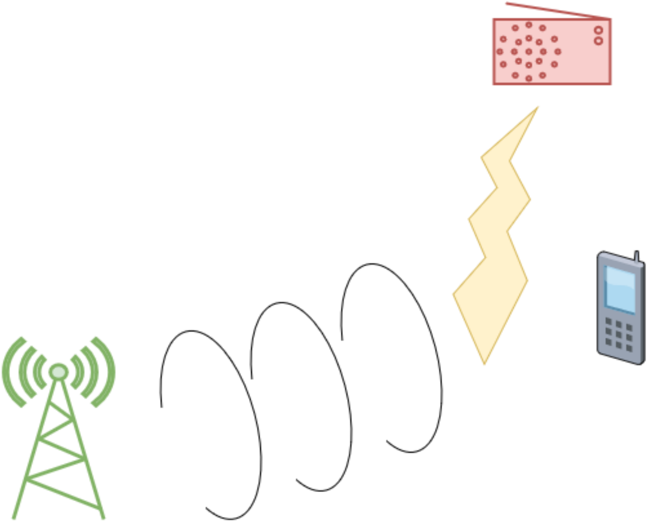
- Spectrum awareness and monitoring → RF scene analysis
- CR adaptive modulation/demodulation
- Military → electronic warfare (EW) → interference avoidance
- Increase spectrum efficiency (modulation cohabitation)
- improve or prevent jamming attacks
- other

Link to security ?

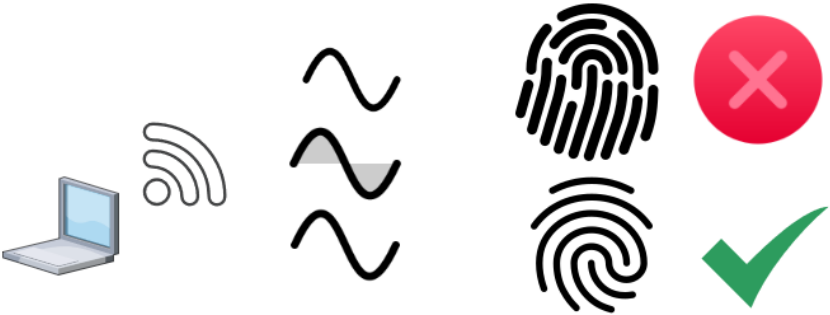
Monitoring examples:

- Drone dropping false Wifi dongle on/close to building
- Drones blocking airports or dropping items over prison
- Detection and monitoring of introduced malicious IoT devices

Jamming



Fingerprinting



State of the art

How to perform Modulation Recognition ?

1. Decision trees based on statistics -> classical military approach
2. Decision theoretic approach (likelihood based classifiers -> cumulative distribution functions (CDF))
3. Feature based approach (spectral features, cyclostationarity combined with Machine learning (ML): KNN SVM GA)
4. Deep learning (CNN, LSTM, Transformers, ...)

How AMR has been achieved here:
→ Fusion of signal decomposition and Convolutional Neural Networks (CNN)

State of the art

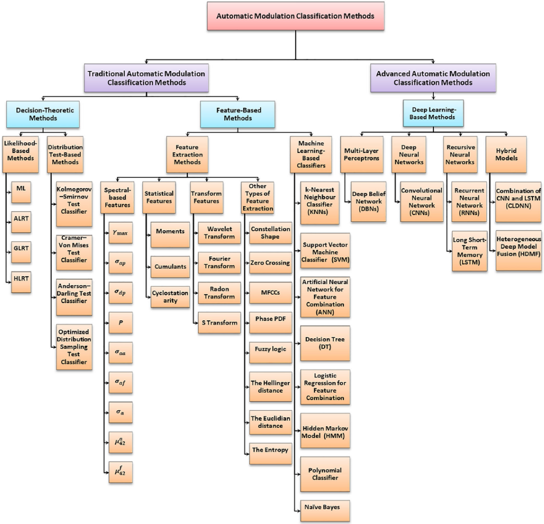


Figure: [1]

Outline

What is BEMD

EMD:

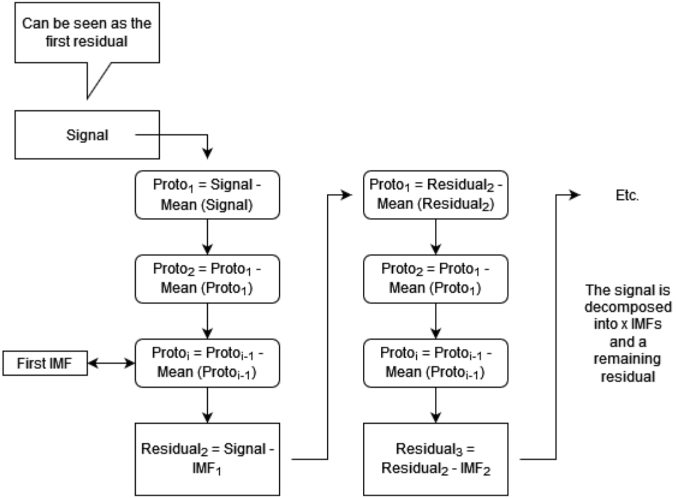
- stands for Empirical Mode Decomposition
- invented by N.Huang in 1998 [2]
- no predetermined basis function
- we obtain Intrinsic Mode Functions (IMFs) $\rightarrow$ sifting process $\rightarrow$ it is an algorithm
- applications: biomedical, natural phenomena analysis, mechanical, image, speech processing
- scarcely used in telecoms $\rightarrow$ opportunity in AMR

In digital telecoms: 2 variables $\rightarrow$ complex signal (IQ)
this justifies the use of Bivariate EMD (BEMD) : [3]

Multivariate EMD methods

- Complex Empirical Mode Decomposition [4]
- Rotation Invariant Complex Empirical Mode Decomposition [5]
- Bivariate EMD [3] [6]
- Bivariate EMD for Unbalanced Signals [7]
- Turning tangent EMD (2T-EMD) [8]
- EMD for Trivariate Signals [9]
- Multivariate EMD (and 3A-EMD = Active Angle Averaging) [10] [11]
- Fast Multivariate Empirical Mode Decomposition [12]

EMD decomposion flow



BEMD Sifting Algorithm [3]

for $1 \leq k \leq N$ **do**

Project the complex valued signal $x(t)$ on direction φ_k (Plane P)

$\rightarrow p_{\varphi_k}(t) = \mathrm{Re}(e^{-i\varphi_k} x(t))$

Extract the locations $[t_j^k]$ of the maxima of $p_{\varphi_k}(t)$

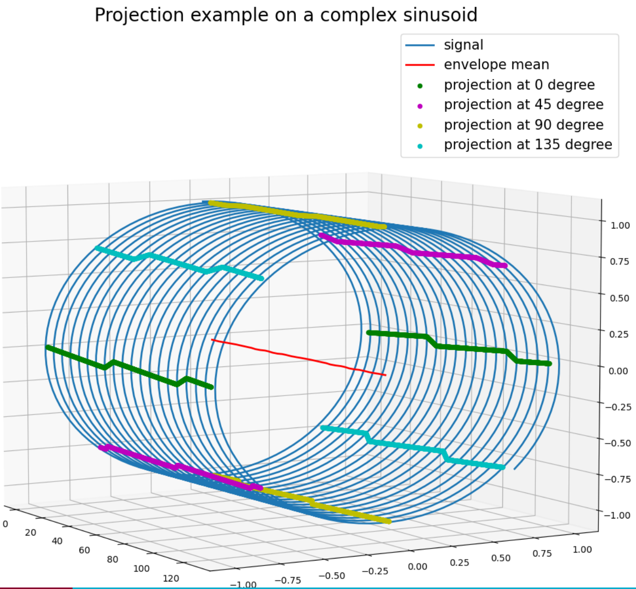
Interpolate the set $(t_j^k, x(t_j^k))$ to obtain the envelope curve in direction $\varphi_k : e_{\varphi_k}(t)$

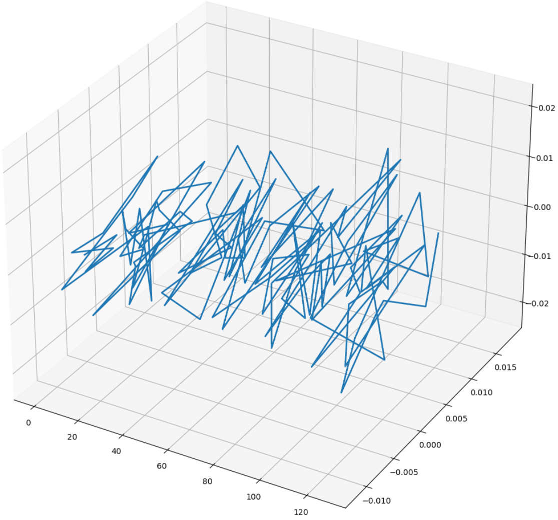
end for

Compute the mean of all envelope curves $m(t) = \frac{1}{N} \sum_k e_{\varphi_k}(t)$

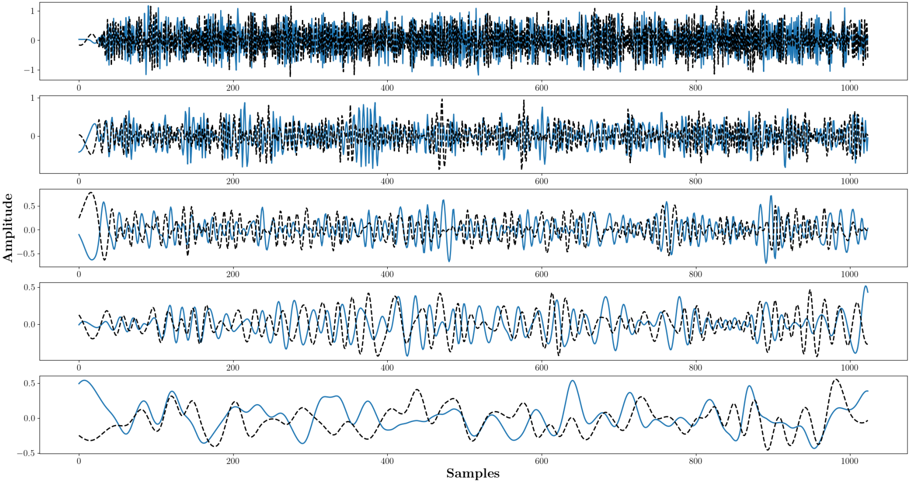
Subtract the mean

Projections example



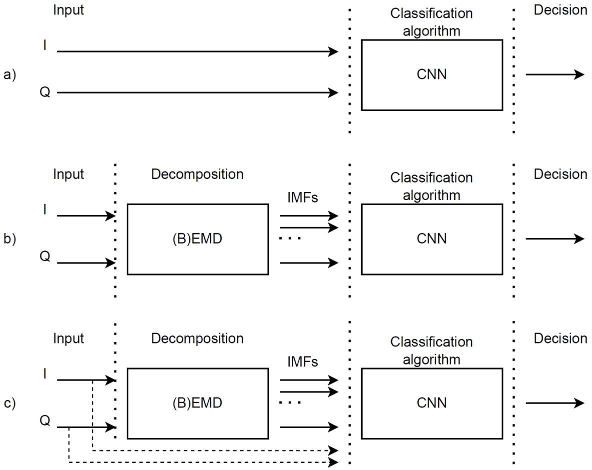

0.02
0.01
0.00
−0.01
−0.02
0.015
0.010
0.005
0.000
−0.005
−0.010
0
20
40
60
80
100
120

Example: QAM16 decomposition



Outline

Methodology flows



CNN process

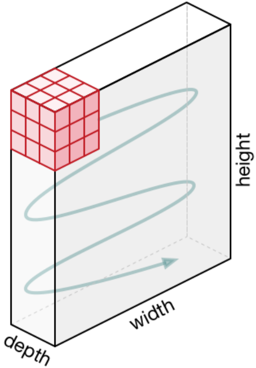
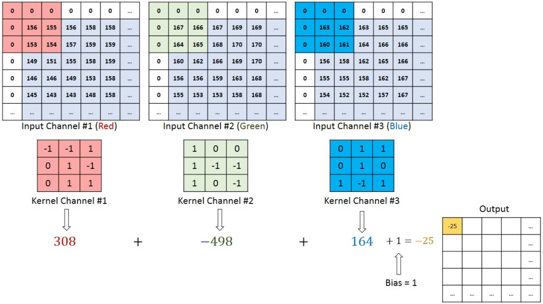


Figure: https://towardsdatascience.com/
a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way

CNN process



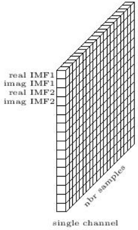
[1]

---

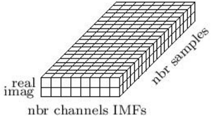
[1] https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

Overall accuracy improvement for each mode and w./w.o. original signal

	EMD	EMD +	BEMD	BEMD +
3D mode	0.7%	1.3%	2%	0.88%
2D mode	-12%	-4.1%	-10.3%	-3.8%

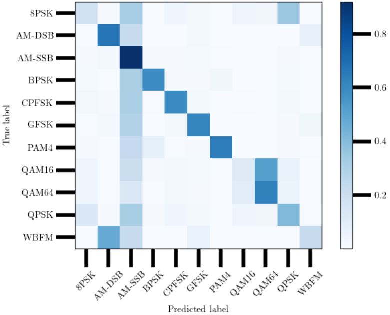


2D mode

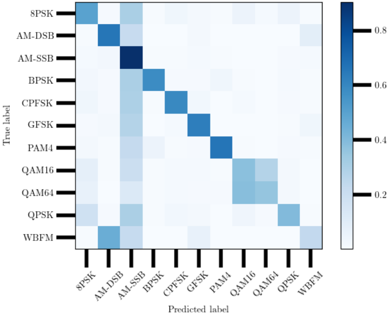


3D mode ("weighted recomposition")

Confusion matrices

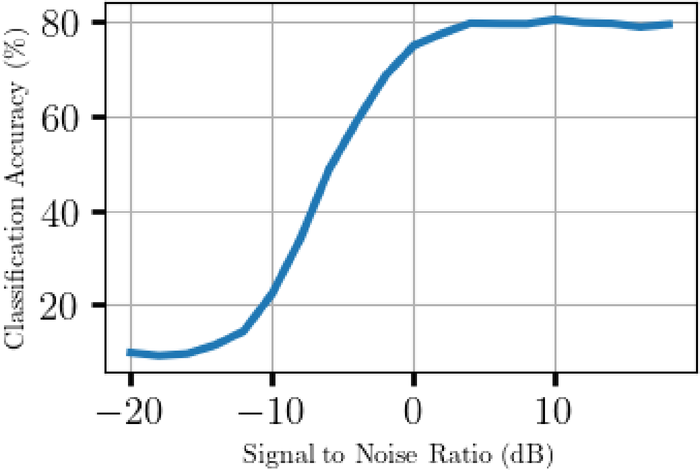


original result using IQ signal

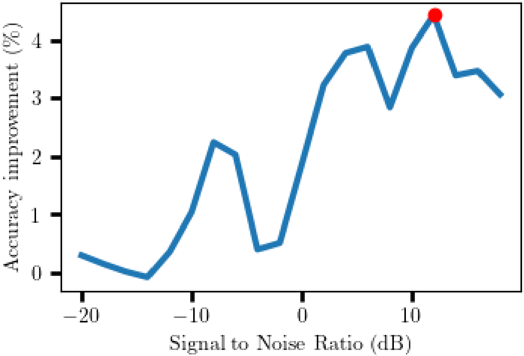


new method using IMFs

Overall accuracy depending on SNR



Accuracy improvement (%) for all modulations depending on SNR



- 2 % overall accuracy improvement
- up to 4.4 % improvement

Outline

Challenges

- How can we improve the extracted IMFs ? Can we get better ones by applying more projections and sifting more ?
- Problem: decomposition is computationally expensive. Solution:
- → code has been modified, Sift (6+r imfs, 3 sifts, 4 projections)
- → takes 26 minutes instead of 7 hours (for 110000 data tensors)
- Can an additional projection step improve the classification ?
- → but CNNs do not like short and long data, they like square images !
- → solution ? use the Vector diagrams of the BEMD decompositions
- Can we improve the AI architecture itself ?
- Ultimate question: decompositions prior to AI architecture → improves accuracy → how does it give more information

Outline

Assumption and parameters ?

Assumption: increasing the number of siftings and projections would give more refined intrinsic mode functions, increasing therefore the quality of the AI architectures input, and thus the classification accuracy.

Decomposition parameters:

- number of projections
- number of siftings
- type of used interpolation

Cubic-spline vs Linear interpolation

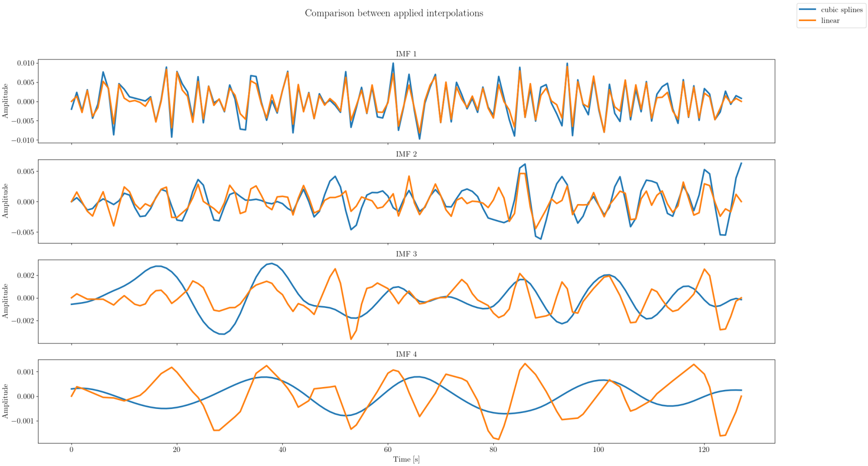


Table: Overall accuracy depending on decomposition parameters

interpolation	siftings	projections	accuracy %	approx time (min)
cubic	3	4	53,86	84
		16	54,05	310
		64	53,67	1012
	10	4	53,96	269
		16	53,94	907
		64	53,76	3917
linear	3	4	51,92	39
		16	52,93	138
		64	53,71	676
	10	4	50,73	134
		16	50,61	530
		64	50,86	2302

Conclusion

- The parameters have very little effect on the overall accuracy of the classifier
- It seems to be an unfavorable result in the sense that we can not improve the results considerably by refining the decomposition
- But it also means that it is not necessary to use high numbers of projections and siftings that increase the decomposition times drastically in order to get good results.
- using linear interpolation gives more IMFs

Outline

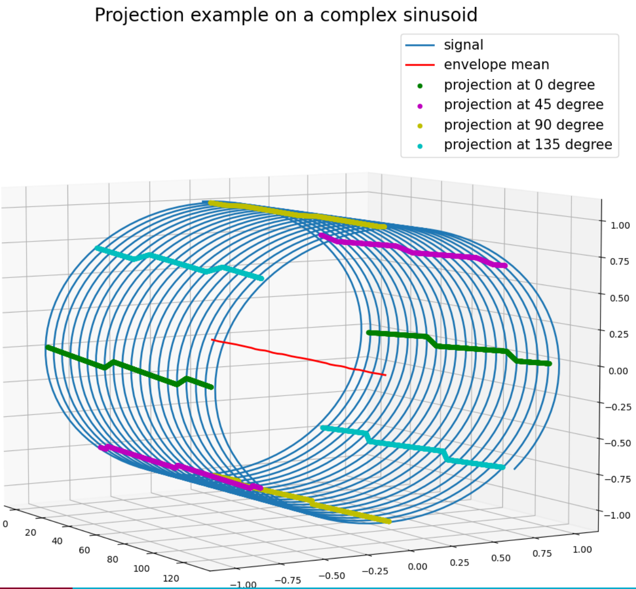
The data transformation

We start from IQ samples:
$S = \{s_1, s_2, s_3, ..., s_N\}$
S a complex set denoted as a measurement,
s_i a complex value of the signal sampling point,
N the number of sampling points per measurement

Originally: we create a two length-N real vector:
$M = \begin{pmatrix} \Re S \\ \Im S \end{pmatrix}$

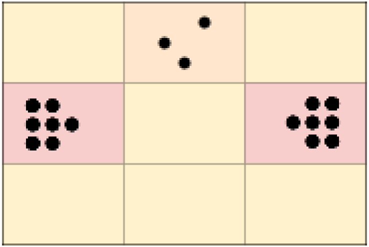
N points in the complex plane are represented as:
$C = \{(\Re s_1, \Im s_1), (\Re s_2, \Im s_2), ..., (\Re s_1, \Im s_N)\}$

Adding a new projection direction

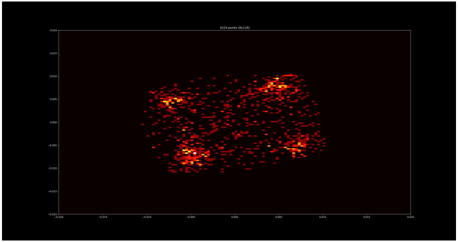


2D histogram or Density kernel

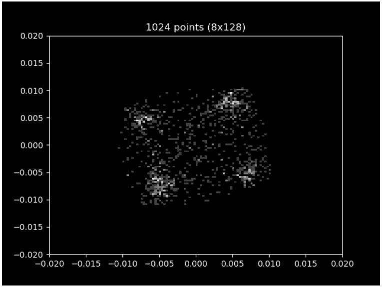
We simply count the number of samples contained in the bin.
Possibility to use colors in order to highlight densities.



Colored constellations

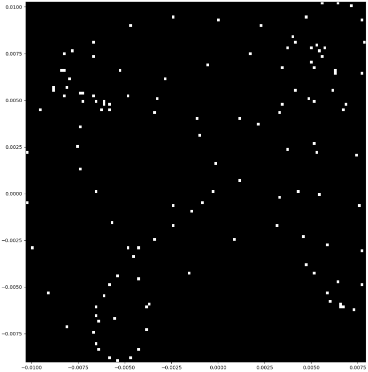


(a) colored QPSK

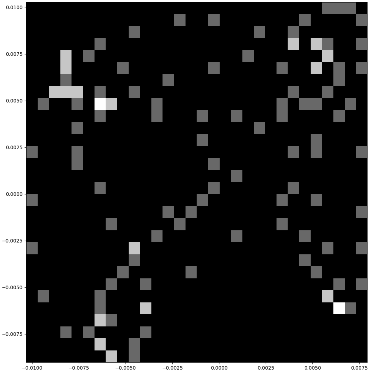


(b) gray QPSK

Effect of bin size (or density kernel)



(a) QPSK, bin size = 128



(b) QPSK, bin size = 30

First attempts

First attempts using the vector diagrams are disappointing !

Adopt previous architecture but adapt input shape for the constellation.

- vectorial diagram of the signals IQ values $\rightarrow$ The training accuracy was 66% and the validation accuracy only 9%.
- vectorial diagram of one of the IMFs $\rightarrow$ training accuracy was 93% and the validation accuracy only 14%.

The network is heavily overfitting !

Outline

Hypermodel optimization

Table: A: 56,25 % / VA: 53,72 %

Value	Best Value So Far	Hyperparameter
5	5	conv_blocks
192	224	filters_0
96	224	filters_1
32	224	filters_2
60	90	Dense units
0.0046402	0.0099645	learning_rate
224	224	filters_3
256	256	filters_4
30	30	tuner epochs

Hyper-parameters using the IQ signal alone and without transformation to vector diagram

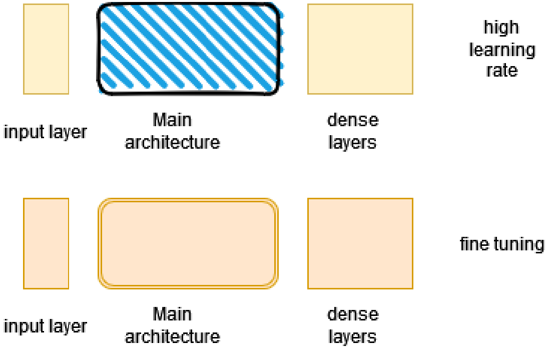
Hypermodel optimization

Table: A: 71,88 % / VA: 56,33 %

Value	Best Value So Far	Hyperparameter
5	5	conv_blocks
160	128	filters_0
64	64	filters_1
128	128	filters_2
50	40	Dense units
0.0049548	0.00011296	learning_rate
224	256	filters_3
64	128	filters_4
30	30	tuner epochs

Hyper-parameters using the IQ signals of the IMFs and without transformation to vector diagram

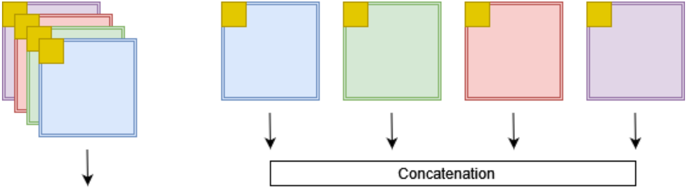
Transfer learning



Tested: VGG16, Resnet, Xception → bad accuracies due to initial layer compression ?

Layer fusion

Also known as 'Hierarchically Deep Convolutional Neural Network', 'Tree-CNN' or 'Concatenated networks'.



- may eliminate the data loss due to the averaging over the layers
- layers analyzed individually → if the information lays in the signal in its whole then the knowledge is lost
- validation accuracy less than 20 % → disappointing

Outline

Explainable AI

We need to understand what part of our signal excites the blackbox model !

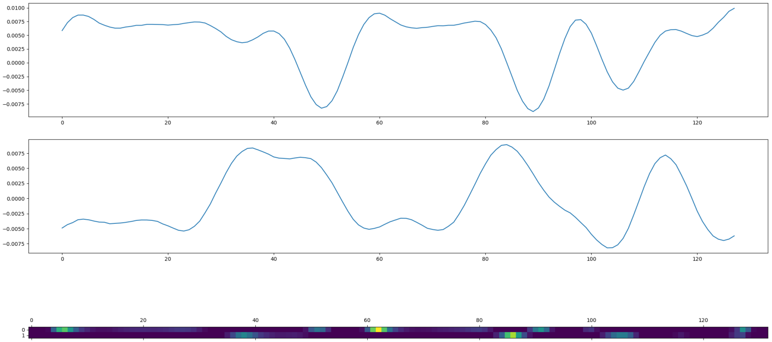
Gradient-weighted Class Activation Mapping (Grad-CAM), after inference of a sample, the weights of the last convolutional layer highlight important regions of the analyzed tensor.

(a) heatmap of the dog output

(b) heatmap of the cat output

Figure: multi-classification problem : animals

GRAD-CAM on IQ signal of QPSK modulation



Conclusion

- Solutions to memory and computational power issues could be found
- It has been proven in some extend that decomposing a signal can improve the accuracy
- A collection of tools (although not optimized) has been created and tested
- In order to highlight what information excites the blackbox model it is needed to use explainable AI (XAI) methodologies
- We also need to understand the XAI outputs

References [x] I

[1] M. Abdel-Moneim, W. El-Shafai, N. El-Salam, E.-S. El-Rabaie, and F. Abd El-Samie, "A survey of traditional and advanced automatic modulation classification techniques, challenges and some novel trends," *International Journal of Communication Systems*, 07 2021.

[2] N. Huang, Z. Shen, S. Long, M. Wu, H. Shih, Q. Zheng, N.-C. Yen, C.-C. Tung, and H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, pp. 903–995, 03 1998.

[3] G. Rilling, P. Flandrin, P. Goncalves, and J. M. Lilly, "Bivariate empirical mode decomposition," *IEEE Signal Processing Letters*, vol. 14, no. 12, pp. 936–939, 2007.

References [x] II

[4] T. Tanaka and D. P. Mandic, "Complex empirical mode decomposition," *IEEE Signal Processing Letters*, vol. 14, no. 2, pp. 101–104, 2007.

[5] M. U. Bin Altaf, T. Gautama, T. Tanaka, and D. P. Mandic, "Rotation invariant complex empirical mode decomposition," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, vol. 3, pp. III–1009–III–1012, 2007.

[6] J. Lilly and S. Olhede, "Bivariate instantaneous frequency and bandwidth," *Signal Processing, IEEE Transactions on*, vol. 58, 02 2009.

[7] A. Ahrabian, N. U. Rehman, and D. Mandic, "Bivariate empirical mode decomposition for unbalanced real-world signals," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 245–248, 2013.

References [x] III

[8] J. Fleureau, J.-C. Nunes, A. Kachenoura, L. Albera, and L. Senhadji, "Turning tangent empirical mode decomposition: A framework for mono- and multivariate signals," *IEEE transactions on signal processing : a publication of the IEEE Signal Processing Society*, vol. 59, pp. 1309–1316, 03 2011.

[9] N. ur Rehman and D. P. Mandic, "Empirical mode decomposition for trivariate signals," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1059–1068, 2010.

[10] R. N. and M. D. P., "Multivariate empirical mode decomposition," *Proc. R. Soc. A.4661291–1302*, 2010.

References [x] IV

[11] J. Fleureau, A. Kachenoura, J.-C. Nunes, L. Albera, and L. Senhadji, “3A-EMD: A Generalized Approach for Monovariate and Multivariate EMD.,” in *Information Sciences, Signal Processing and their Applications*, (Kuala Lumpur, Malaysia), pp. 300 – 303, May 2010.

[12] X. Lang, Q. Zheng, Z. Zhang, S. Lu, L. Xie, A. Horch, and H. Su, “Fast multivariate empirical mode decomposition,” *IEEE Access*, vol. 6, pp. 65521–65538, 2018.

Thank you for your attention !!