Density Core Regime Selection in Experimental Self Mixing Optical Feedback Signals Using DBSCAN

Başak Ersöz
Politecnico di Torino
Torino, Italy
basak.ersozyildirim@polito.it

Priyanka Chaudhary
Università di Bari Aldo Moro
Bari, Italy
priyanka.chaudhary@uniba.it

Maurizio Dabbicco
Università di Bari Aldo Moro
Bari, Italy
maurizio.dabbicco@uniba.it

Massimo Brambilla
Politecnico di Bari
Bari, Italy
massimo.brambilla@poliba.it

Lorenzo Columbo
Politecnico di Torino
Torino, Italy
lorenzo.columbo@polito.it

Paolo Bardella
Politecnico di Torino
Torino, Italy
paolo.bardella@polito.it

***Abstract*—Self mixing optical feedback interferometry enables compact sensing because target back reflection perturbs the laser cavity and produces measurable voltage like waveforms. In experimental conditions, self mixing signals are highly sensitive to alignment drift, feedback fluctuations, and speckle decorrelation , so acquired datasets often contain mixed operating regimes where regime inconsistent traces coexist with stable measurements. This is especially critical for sensing pipelines that must not only classify valid targets but also reject non informative acquisitions, represented here by a Blank class.**

We acquire self mixing waveforms under speckle illumination generated by a nanofiber diffuser, conceptually related to speckle driven single pixel imaging architectures, and consider a six target benchmark comprising five letters (U, B, A, R, I) and a Blank rejection class. Each waveform is mapped to a compact six feature descriptor built from global statistics of the raw signal and its discrete time derivative. After z-score standardization, descriptors are embedded in two dimensions using t distributed stochastic neighbor embedding and filtered using DBSCAN to retain dense regime consistent cores while rejecting sparse outliers. By sweeping the DBSCAN operating point, we construct four filtered datasets spanning a retention robustness tradeoff and evaluate standard classifiers under a fixed stratified 70/15/15 protocol. Results show that conservative core selection improves stability and reduces the dominant confusion between U and Blank, while more permissive settings increase retention at the cost of higher cross class confusions.

***Index Terms*—Self mixing interferometry, optical feedback, classification, t-SNE, DBSCAN, machine learning**

I. Introduction

Optical feedback interferometry in the self mixing configuration provides compact readout because a fraction of the emitted field is reinjected into the laser cavity and modulates the laser signal in a way that depends on the external target [1]. In real measurements, the same sensitivity that enables high resolution sensing also makes the waveform statistics strongly dependent on alignment, feedback strength, and speckle decorrelation [2]. As a consequence, experimental datasets commonly contain mixed operating regimes, where stable measurements coexist with regime inconsistent traces that degrade downstream inference [3].

In practical sensing, the system must both identify informative targets and reject non informative acquisitions arising from background regions, weak feedback, or unstable regimes. We model this requirement explicitly by including a Blank class. Blank is challenging because under imperfect optical feedback conditions its waveforms can overlap with true targets, especially for low feedback letters such as U. For this reason, regime consistency and rejection reliability must be evaluated directly, not only overall accuracy.

This paper proposes a data driven regime core selection stage based on density in a learned two dimensional manifold. We map each waveform to a compact six feature descriptor and use t distributed stochastic neighbor embedding for visualization [4]. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is then applied to retain dense cores interpreted as regime consistent measurements and reject sparse outliers interpreted as unstable regimes [5]. The DBSCAN operating point acts as a selectivity control knob that trades retention for robustness. We validate this mechanism on an experimental U, B, A, R, I and Blank dataset acquired under speckle illumination in an optical feedback setup.

II. Experimental setup and dataset

A. Optical feedback acquisition under speckle illumination

The experimental optical feedback chain is summarized in Fig. 1. A semiconductor laser operating at 830 nm illuminates a printed target through a nanofiber diffuser mounted on a motorized translation stage. Translating the diffuser produces a sequence of distinct speckle illuminations. For each stage position, a waveform is acquired while the back reflected field from the target is reinjected into the laser cavity in the self mixing configuration [1]. A CMOS camera records the speckle pattern associated with the same diffuser position. The acquisition is performed point by point along the scan, which intentionally exposes the measurement process to small alignment drift and feedback fluctuations, leading to mixed regimes in the collected waveforms.

B. Six target benchmark with Blank rejection

We consider six labeled conditions: five letters (U, B, A, R, I) and a Blank rejection class. Blank samples correspond to non letter regions or non informative acquisitions. The

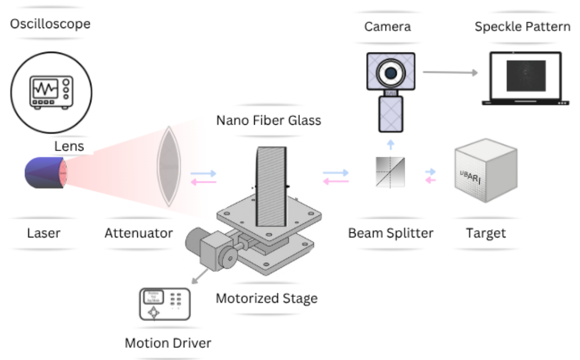

Fig. 1. Experimental optical feedback measurement chain. The attenuated laser output is split to illuminate the target through the nanofiber diffuser on a motorized stage and provide monitoring with an external photodiode. Light back reflected from the target is reinjected into the laser cavity in the self mixing configuration and the waveform is acquired by the oscilloscope. A CMOS camera records the speckle pattern.

dataset contains 12000 waveforms, balanced across the six classes with 2000 samples per class. Unless otherwise stated, all results use a fixed stratified split of 70 percent training, 15 percent validation, and 15 percent test.

III. Feature representation and regime core selection

A. Compact six feature descriptor

Direct comparison of raw waveforms is unreliable under drift and feedback fluctuations. We therefore map each waveform to a time independent descriptor built from global statistics computed on the raw signal and on its discrete time derivative. We use three robust statistics: root mean square, variance, and peak to peak amplitude. Computing the same statistics on the derivative emphasizes fast variations and complements amplitude dominated information from the raw waveform. The resulting six dimensional descriptor is denoted f6.

Before any distance based analysis and before training linear models, each feature dimension is standardized using z score normalization computed on the training set only. The training mean and standard deviation are then used to normalize training, validation, and test features.

B. Manifold embedding and DBSCAN filtering

Standardized f6 descriptors are embedded into two dimensions using t distributed stochastic neighbor (t-SNE) [4]. DBSCAN is then applied in the embedded space to retain dense cores and reject sparse outliers [5]. We interpret dense cores as regime consistent measurements and outliers as regime inconsistent measurements.

Figure 2 shows an example embedding with DBSCAN core and noise separation for Dataset A. The DBSCAN operating point is defined by epsilon and minPts. Increasing epsilon or decreasing minPts makes the filter more permissive, retaining more samples but admitting less consistent measurements.

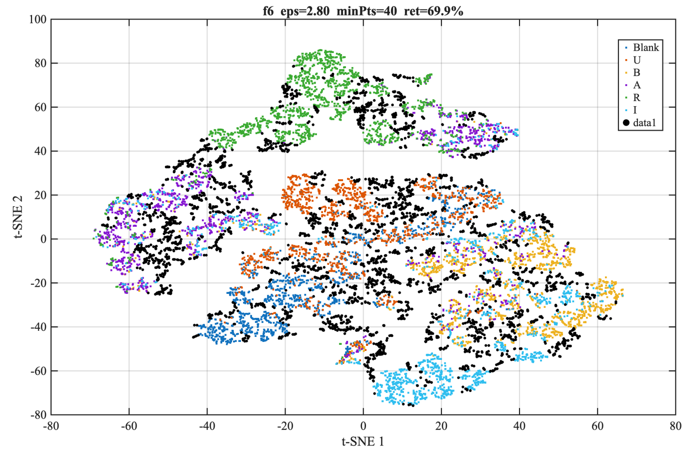

Fig. 2. Example standardized f6 t distributed stochastic neighbor embedding with DBSCAN core and noise separation for Dataset A. Colored points indicate targets and black points are rejected as noise.

TABLE I
DBSCAN operating points A, B, C, D with retention ratio rho and test performance for logistic regression and random forest using f6 (70 15 15 split).

	DBSCAN		LR		RF	
Set	epsilon	rho	Acc	Macro F1	Acc	Macro F1
A	2.80	0.699	0.7154	0.6915	0.8482	0.8469
B	3.00	0.860	0.7207	0.7055	0.8345	0.8341
C	2.80	0.945	0.6812	0.6660	0.8171	0.8166
D	3.00	0.969	0.7070	0.6946	0.8171	0.8172

IV. Protocol, metrics, and results

A. Operating points and evaluation protocol

To quantify the selectivity tradeoff, we define four DBSCAN operating points and construct four filtered datasets A, B, C, D by sweeping epsilon and minPts. Each filtered dataset is then evaluated with standard supervised baselines under the same fixed stratified 70/15/15 split. We report accuracy and Macro F1 [6]. Macro F1 weights all classes equally and is therefore appropriate when the key failure mode is class specific, such as Blank rejection.

B. Blank rejection as a sensing requirement

In practical sensing, a system must not only classify valid targets but also reject non informative measurements arising from background regions, weak feedback, or unstable operating regimes. We model this requirement explicitly with a Blank class. Because Blank can overlap with true targets under imperfect optical feedback, we report confusion analyses and a pooled Blank versus Target view to quantify rejection reliability.

C. Main findings

Table I shows that random forest achieves the best overall performance across operating points. The best result is obtained for Dataset A with accuracy 0.8482 and macro F1 0.8469. The pooled Blank rejection view in Table II summarizes the same outcome as a two class decision that

TABLE II

BLANK REJECTION SUMMARY FOR RANDOM FOREST ON DATASET A. LETTERS ARE POOLED INTO A SINGLE TARGET CLASS (U,B,A,R,I). ROWS ARE TRUE AND COLUMNS ARE PREDICTED.

	Pred Blank	Pred Target
True Blank	172	33
True Target	35	1018

matches sensing pipelines where non informative measurements must be rejected. As the DBSCAN operating point becomes more permissive from A to D, retention increases but macro F1 decreases, indicating that the additional retained samples include a higher fraction of regime inconsistent measurements that increase cross class confusions. The dominant residual error remains the confusion between U and Blank, making the rejection boundary the most sensitive failure mode under imperfect optical feedback. This behavior is consistent with prior evidence that experimental misalignment amplifies inference errors in related optical pipelines [7].

V. CONCLUSION

We presented an experimental optical feedback benchmark with a practically motivated Blank rejection class and demonstrated that DBSCAN controlled density core extraction provides a simple data driven mechanism to identify regime consistent subsets in self mixing measurements. By tuning the DBSCAN operating point, retention can be traded for robustness. Results show that conservative core selection improves stability and reduces U versus Blank confusions, supporting regime core selection as an effective preprocessing step for sensing pipelines operating under alignment drift and feedback fluctuations.

ACKNOWLEDGMENT

This study was carried out within the MIRABILIS project (CUPM D53D23002780006), funded by European Union Next Generation EU within the PRIN 2022 program. This manuscript reflects only the authors views and opinions and the Ministry cannot be considered responsible for them.

REFERENCES

[1] D. M. Kane and K. A. Shore, *Unlocking Dynamical Diversity: Optical Feedback Effects on Semiconductor Lasers*. Hoboken, NJ: Wiley-IEEE Press, 2005.

[2] S. Ottonelli, M. Dabbicco, F. De Lucia, M. Di Vietro, and G. Scamarcio, "Laser-self-mixing interferometry for mechatronics applications," *Sensors*, vol. 9, no. 5, pp. 3527–3548, 2009. [Online]. Available: https://www.mdpi.com/1424-8220/9/5/3527

[3] T. Fukui, Y. Kohno, R. Tang, Y. Nakano, and T. Tanemura, "Single-pixel imaging using multimode fiber and silicon photonic phased array," *Journal of Lightwave Technology*, vol. 39, no. 3, pp. 839–844, Feb 2021. [Online]. Available: https://opg.optica.org/jlt/abstract.cfm?URI=jlt-39-3-839

[4] S. Gare, S. Chel, M. Kuruba, S. Jana, and L. Giri, "Dimension reduction and clustering of single cell calcium spiking: Comparison of t-sne and umap," in *2021 National Conference on Communications (NCC)*, 2021, pp. 1–6.

[5] D. Deng, "Dbscan clustering algorithm based on density," in *2020 7th International Forum on Electrical Engineering and Automation (IFEEA)*, 2020, pp. 949–953.

[6] A. Luque, A. Carrasco, A. Martín, and A. de las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognition*, vol. 91, pp. 216–231, Jul. 2019.

[7] B. Ersöz, P. Chaudhary, P. Bardella, M. Dabbicco, L. Columbo, and M. Brambilla, "Inaccuracy amplified: Compressed sensing under experimental misalignment," in *2025 International Conference on Numerical Simulation of Optoelectronic Devices (NUSOD)*, 2025, pp. 51–52.