Persistence Detection in Slew Dark Data Using Autoencoders

Presenter: Amirmohammad Chegeni amirmohammad.chegeni@unipd.it

Sezione di Padova

Department of Physics and Astronomy "Galileo Galilei" April 15th, 2025

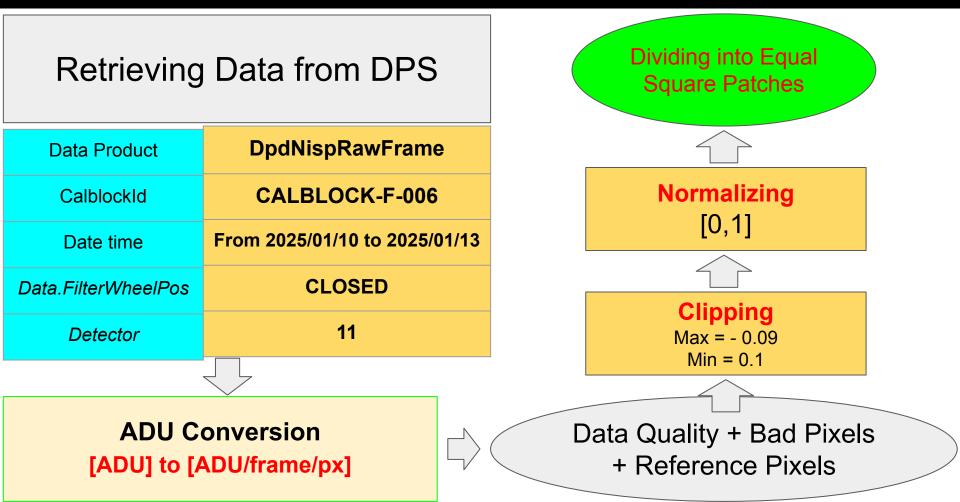
Outline

Data Preprocessing Pipeline (CALBLOCK-F-006)

Training Process

- Autoencoders
- Training data and Optimization
- Reconstruction loss
- Latent Space Analysis
 - UMAP for Dimension reduction
 - GMM for Clustering UMAP Space
 - Number of Clusters Challenge
 - Data Diversity Vs XAI
- Slew dark Persistence Unsupervise Recognition across Channels
- Segmentation Pipeline
- Next Steps ...

Data Processing Pipeline

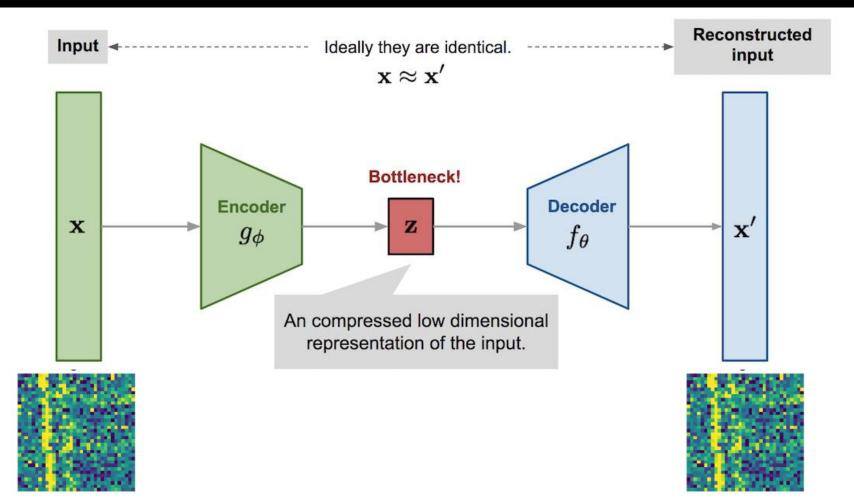


136*136 Patches

68*68 Patches

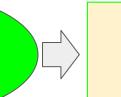
				-					-	_		_			STATISTICS STATISTICS	MINISTER MINISTER	MARCEL MARCHINE		-		PARTIES INVERSE				R. T. Street	NAME OF OCCUPANT	INTER DATE	CALLSON DECISION
																				NUCLEAR ADDRESS	Restored Parabolis			and provident of				
e sette	e dia ma	- <u>1</u> -19		e source	No.		and the state		and a state	e de re				ar - Ar		REALING BUILDER	ADADESS PRESIDE		1522-052 BAS4823 1	NUMBER OF STREET	PERSONAL ENCODES	BRANNIN BRANN	A STRAFTS Hours	NGO ESSENCIES IN				
a da. Redstarie		a dina dina Secondaria	1978) 1979	a digi sebah Sebah sebah seb		9.44 19.44 19.44		an an Ara Salar									CARGONS COTORS		Reference Inconcerna I	and an owned	homents medican	BRANKS DARESA	a activity at the st	and prevente at	COMPANY CONTRACTOR	Constanting Designation		
	WE S		2.80								1.000										ENTRY ECONS	Internet Descent			COLUMN BOULDER			
		0	2		<u></u>		the second second							19- CL-2						200101 220200	Sectors Address							
												an tress				Exception Interaction					station arrange	Address streets			NAMES OF TAXABLE			
																					CARD SEE				1992			
		. Carro					a construction			4.49.3							NUMBER OF STREET		Accessingly approvide 1	STREETS MODELSMA	10193-1021 (SOTO2)	Restances Meansure	a sources missi	an areas b	COLUMN LEAGUER	Control Constant		
				1. A. 1. 2. 2.										44 C	ESCOURS ENGLISH		CONTRACT CONTRACTOR	NUMBER OF STREET	Distantia Internetia	Provide and the	COLUMN 1	straught photogra	a sources save	No. PROJECT	CALCURE BELIEVE	Land to the Design of the		
						12.44				4																		
													2.05			Reality Provides	NOTIONS DESIGN		HEADING SHEROLDH	and the second	REAL PROPERTY.	NUMBER OF STREET			Northernesi	Contraction and Contraction		
	Service of the servic									s. Electric							Andreas Modeland		CONSIGNATION INCOMENT	COLUMN STRUCTURE	ADVISION REPORT	Entering Colores				Ascense Rossons		
		63.68		-	ante antes antes antes		an and a second										assess access		INCOME DESCRIPTION		BROADER RECENT							
Contraction of the second seco	Color Colored															6605500 6339588	RADACTO RECORD		Increased Internet	NUMBER OF STREET, STRE	RECEIM DISECTOR	ADDRESSED BROAD		alla Ballanad B				
						and the																						
																				1996 A 1998	SECURIA PROPERTY		2 23-24-252 142-6-2	BOR ZERESA R				
NUM							and a second									STREET, STREET,	ADDRESS CONTRACTOR		Instants Berrinted		ALCONO. IN CONTROL	BEDRESS RECEIPTS			Notes and		0	
	AND THE OWNER	Market of Arriston						E CONTRACTO			And a second second					STOKING STOCK	STREET, STREET		ENGINEER ENGINEER		ADDINESS DESCRIPTION		S REPARTS MERCO		INCOME DISPLACE	MANUTA INFIDIA		
									i prine La com	- (²⁰	ur j						STREET, STREET,			20226	BALLANK BALLAND	Encoded Document	C DURNER PROD		100000	Records Provide		
			1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.							and the second		N. Sector	Received and															
										12							PERSONAL INCOME		Processor Resource 1		BORNESSE RESIDENT			AND PERSONNEL IN				
						A CONTRACTOR										economic provides					Stores destan	DOMENTS REPORT						
						1996-1													Restance Management		Marine Partice							
																Internetia strensial			Designation Restaurant I									
			100 May 1															Entrances Exception			Arrown Division							
												1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.	Tuesday and			BARGERS BREIGHT				COURSE NOTION								
									CONTRACTOR OF										Reserved Referred 1		Services Manager							
Second Second	Contract Contract		and a second second	24 AN 1999	respectively.	a service and	and the second	and the second	and the second of	(1) A. (8746)																	and a	

Training Process - Autoencoders



Training Process - Training data and Optimization

Total Data Shape (192960,34,34)



70% Training data 15% Validation data 15% Test data

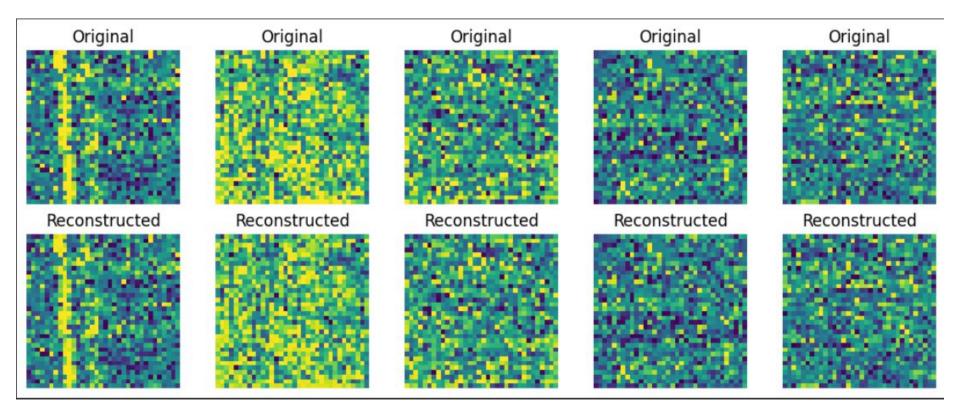
Consider a Simple AE Architecture

Layer (type)	Output Shape	Param #
input_layer (InputLayer)	(None, 136, 136, 1)	0
conv2d (Conv2D)	(None, 136, 136, 16)	160
<pre>max_pooling2d (MaxPooling2D)</pre>	(None, 68, 68, 16)	0
conv2d_1 (Conv2D)	(None, 68, 68, 32)	4,640
<pre>max_pooling2d_1 (MaxPooling2D)</pre>	(None, 34, 34, 32)	0
conv2d_2 (Conv2D)	(None, 34, 34, 64)	18,496
conv2d_3 (Conv2D)	(None, 34, 34, 64)	36,928
up_sampling2d (UpSampling2D)	(None, 68, 68, 64)	0
conv2d_4 (Conv2D)	(None, 68, 68, 32)	18,464
up_sampling2d_1 (UpSampling2D)	(None, 136, 136, 32)	0
conv2d_5 (Conv2D)	(None, 136, 136, 1)	289

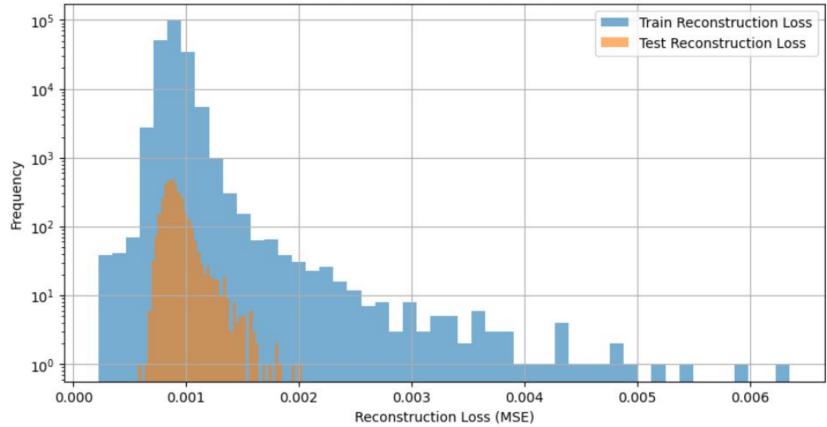
Optimize the AE parameters using Optuna library

ModelCheckpoint EarlyStopping ReduceLROnPlateau

Training Process - Reconstruction loss

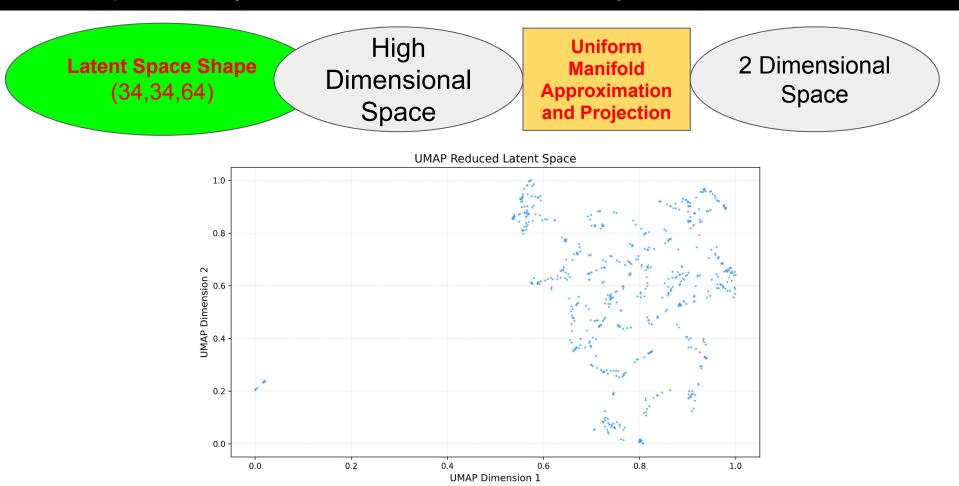


Training Process - Reconstruction loss

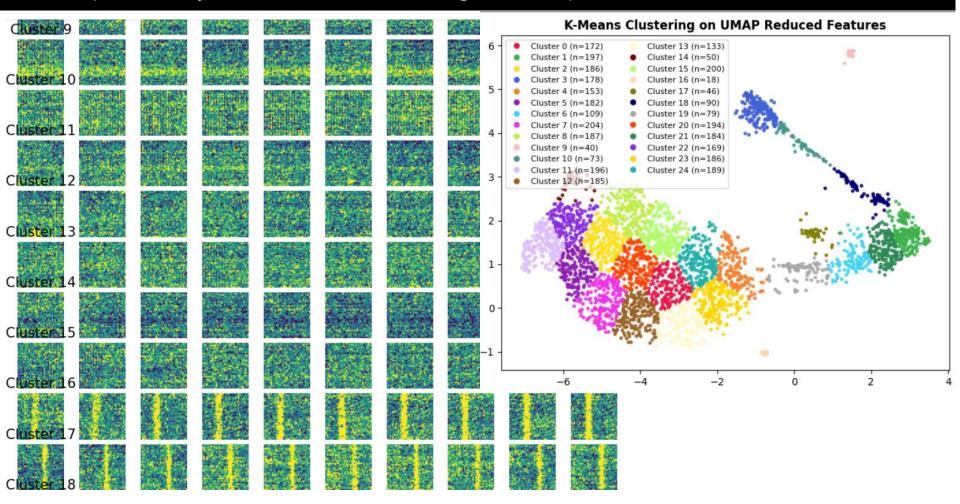


Reconstruction Loss Distribution

Latent Space Analysis - UMAP as A Dimensionality Reduction Tool

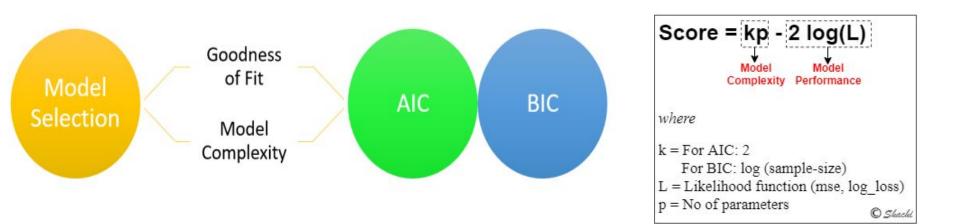


Latent Space Analysis - GMM for Clustering UMAP Space

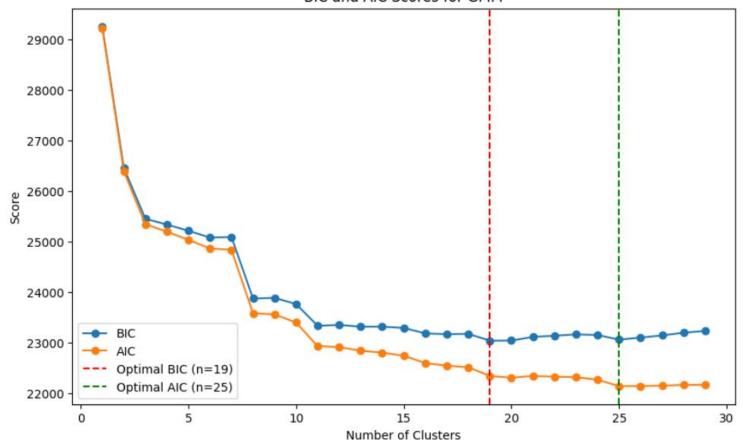




AIC = Akaike Information Criterion BIC = Bayesian Information Criterion



Latent Space Analysis - GMM for Clustering UMAP Space



BIC and AIC Scores for GMM

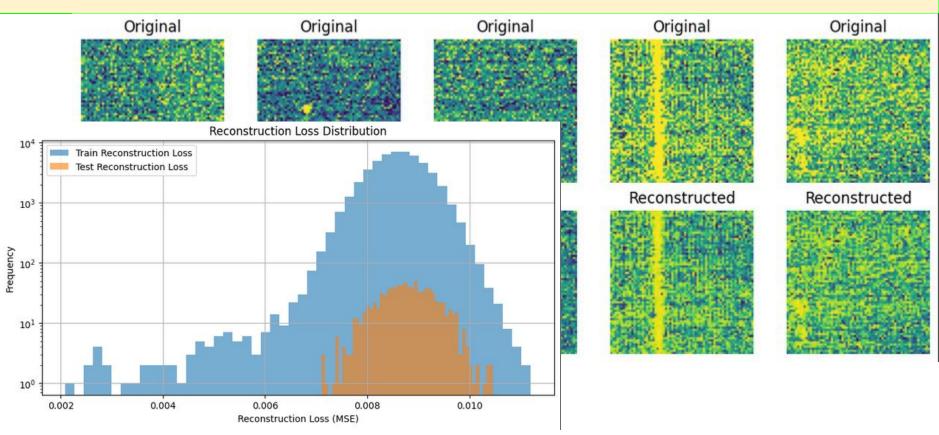
Feature Clustering - Identifying Clusters (Red Edgecolor Patches)

GMM Clustered Patches (n_clusters=19)

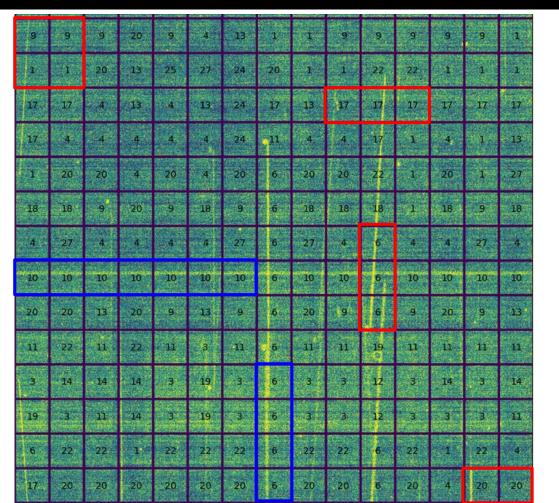
18 18 11 7 2 7 7 11 11 10 11 11 11 1 2 2 11 3 11 5 11 3 11 11 2 35 11 1 11 1
14 3 5 0 15 3 3 3 3 10 3 3 3 3 2 2 3 3 3 0 3 3 3 3 3 2 15 3 5 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
18 14 14 18 15 14 14 14 14 14 10 8 14 0 3 3 15 15 14 18 14 14 0 7 5 14 8 2 2 16 3 0 0 0 14 0 14 8 0 8 8 0 3 3 3 8 3 8 16 3 3 3 8 8 3 0 0 8
18 18 18 15 15 18 18 18 18 16 18 18 16 16 15 15 18 18 7 18 18 18 5 5 16 2 15 13 7 7 7 48 18 18 18 5 18 13 18 7 16 8 13 5 5 5 13 13 16 16 18 18 18 18 7
18 14 18 14 15 18 18 14 14 10 14 14 14 8 14 15 15 18 18 8 14 0 5 8 3 8 2 15 5 14 3 0 14 0 7 0 8 7 14 5 3 8 3 3 8 6 16 0 3 8 8 8 3 0 8 0
1 1 9 1 15 10 1 18 18 10 18 18 18 18 14 15 15 18 10 18 18 18 18 18 7 5 2 15 7 18 18 14 18 18 7 7 7 13 13 16 16 5 13 16 13 5 7 0 13 7 18 18 7 18 18 1
1 1 9 9 15 18 18 18 10 10 18 18 7 13 5 15 15 10 18 18 18 18 5 16 5 7 2 15 16 5 18 18 18 18 18 18 16 13 5 7 3 16 16 3 3 7 5 16 8 3 5 7 18 7 18 7 18 18
1 1 9 12 17 12 1 10 18 10 14 18 7 5 18 15 15 1 1 14 18 18 18 14 18 14 2 2 14 14 7 14 18 18 18 5 16 5 5 16 5 16 8 16 7 5 5 5 16 14 14 14 18 18 14 1
1 1 9 9 17 12 9 18 14 10 5 5 14 16 5 15 15 14 14 8 0 14 14 14 5 8 2 15 8 5 5 16 8 8 3 3 3 8 14 8 3 3 8 3 3 8 16 3 16 3 16 3 8 8 5 16 14 14
1 1 9 9 17 12 9 9 18 10 0 14 5 5 5 15 15 18 18 5 14 14 0 5 5 5 15 15 5 0 8 5 0 0 5 8 16 5 8 5 8 16 5 5 5 5 8 8 8 5 5 5 5 0 0
9 9 9 17 12 9 9 9 1 18 18 18 18 14 15 15 18 18 18 18 18 18 7 18 18 2 15 18 18 18 18 18 18 18 18 18 3 0 16 18 3 16 3 7 18 18 18 16 16 16 7 13 18 18 18 18 18 18 18 18
9 9 9 9 9 12 9 1 9 9 18 18 18 18 18 15 15 16 18 18 18 18 18 18 18 7 2 15 18 18 18 18 18 18 15 7 7 18 18 7 3 3 3 7 5 18 18 18 18 7 18 18 14 18 18 18 1
9 9 9 9 12 12 9 1 1 9 18 18 18 14 18 15 15 18 18 5 7 18 7 13 7 16 2 15 13 7 16 5 7 7 5 16 3 5 5 5 16 16 3 16 13 16 13 5 16 13 16 7 7 14 7
9 9 9 9 1712 9 9 9 9 1 1 7 14 5 15151414 5 5 14 14 13 5 5 2 2 5 16 5 5 5 5 5 5 8 8 5 7 8 16 3 16 5 5 5 5 5 8 0 5 5 14 5 5 1
9 9 9 112 9 9 1 1 1 7 5 2 15414 5 7 18 7 16 5 2 2 5 16 5 5 5 5 5 5 5 5 3 3 5 5 5 5 5 5 5 16 5 5 7 5 14
9 9 9 9 17 12 9 9 9 1 1 1 1 1 16 13 15 15 7 16 13 7 7 13 13 13 7 2 2 3 3 7 7 13 5 7 13 3 13 13 13 13 13 16 15 13 7 13 13 13 13 13 17 7 13 13 13
9 9 9 9 12 12 9 9 9 1 1 1 1 1 7 7 15 15 18 7 5 7 13 7 13 7 5 2 15 3 3 13 13 13 13 13 16 16 16 16 16 16 16 13 7 7 7 7 13 13 7 18 13 7 1
9 9 9 12 17 12 9 9 9 1 1 1 1 14 14 15 15 14 18 14 18 18 14 0 18 18 2 15 8 3 14 0 14 14 0 0 0 14 0 0 0 8 0 0 0 14 14 18 18 0 14 14 18 18 14 14 14 1
9 9 9 9 17 12 9 1 9 1 1 1 1 7 5 15 15 18 7 16 5 7 16 5 5 5 2 15 3 3 7 5 7 5 16 13 3 16 16 5 5 16 3 16 5 13 7 13 5 16 7 7 16 13 13 7
9 9 9 12 12 9 9 9 1 1 1 1 18 5 15 15 5 0 5 5 0 5 8 0 5 2 15 3 3 5 5 0 5 5 5 16 5 5 5 8 16 3 5 0 7 14 5 0 0 0 5 14 0 0 0
9 9 9 9 17 12 9 9 9 1 18 1 1 1015 15 7 7 13 7 14 5 16 5 16 2 15 16 3 7 7 7 7 16 16 16 16 16 16 16 16 7 7 18 16 7 7 7 18 18 7 7 7
9 9 12 17 12 9 9 9 1 1 1 1 1 10 15 15 14 14 5 14 18 5 8 0 16 2 15 16 16 7 7 7 7 5 5 16 16 16 7 5 16 3 16 16 7 7 5 5 5 14 18 18 14 7 7 1
9 9 9 17 12 9 9 9 1 1 1 1 1 1 1 17 15 18 18 18 18 18 18 18 18 2 15 0 3 14 14 14 14 5 5 5 13 16 16 5 16 3 5 5 14 18 16 14 14 14 4 18 5 14 14 1
9 9 9 9 17 12 9 9 9 1 1 1 1 1 18 15 15 0 0 8 0 0 16 8 5 8 2 15 3 3 0 0 0 0 8 8 8 16 16 8 8 3 3 16 8 0 0 8 8 8 0 14 0 0 8 8 0
1 9 9 9 17 12 9 1 1 1 18 1 1 18 18 1 2 7 14 5 5 0 5 16 0 5 2 15 16 3 18 14 5 5 16 16 16 8 8 5 3 8 3 5 5 14 7 8 0 8 16 18 5 5 7 5 0
9 9 9 9 17 12 9 1 9 1 14 1 1 18 14 15 2 8 8 8 8 8 8 8 8 8 2 15 3 3 14 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8
9 9 9 12 17 12 9 1 9 1 1 1 9 14 18 17 15 0 8 8 0 8 5 8 8 8 2 15 3 3 14 5 0 0 8 8 8 8 5 13 0 8 8 10 8 0 8 5 8 0 1 7 0 8 8 0
9 9 9 12 17 12 9 9 9 1 1 1 1 10 10 15 15 5 5 16 16 16 16 16 16 2 2 3 3 13 7 13 18 16 16 16 16 16 16 16 16 16 16 16 5 16 16 7 16 5 18 13 7 7 16
9 9 9 12 17 12 9 9 1 1 1 1 9 1 1 17 15 5 7 5 5 7 0 5 5 2 2 16 3 5 0 7 5 5 8 5 16 5 5 5 5 5 10 7 7 5 14 0 14 7 5 14 14 0 1
9 9 9 12 17 12 9 1 9 1 1 9 9 1 1 17 15 18 14 16 16 14 5 16 16 16 2 15 16 16 14 14 14 14 5 18 14 16 5 0 5 7 0 7 10 14 14 7 14 5 1 7 18 14 14 14 1
9 9 9 9 17 12 9 1 1 1 14 1 1 1 18 15 15 8 8 3 8 8 16 16 8 8 2 15 3 3 8 0 0 7 8 8 8 16 16 8 8 0 16 0 10 0 0 5 7 8 14 5 0 8 0 0 9
9 9 9 12 17 12 9 9 1 1 1 9 1 1 1 17 15 16 16 8 16 16 5 16 3 3 2 15 3 3 16 5 16 16 16 5 8 16 16 16 16 16 16 16 16 16 16 16 16 16
9 9 9 12 9 9 9 9 9 9 9 9 9 9 12 6 6 6 6 6 6 6 6 6 2 2 6 6 6 6 6 6 6
9 9 12 17 12 9 9 9 1 1 9 9 1 17 15 14 5 8 5 18 14 5 8 8 2 15 16 7 14 14 7 18 14 5 5 0 14 14 8 14 5 18 10 18 18 14 14 18 10 18 14 14 18 10 18 14 14 18 18 1
9 9 9 12 17 12 9 1 1 1 1 1 1 1 1 1 17 15 0 8 8 5 14 14 5 16 8 2 15 3 3 14 5 8 0 0 5 8 8 5 5 0 0 5 14 10 14 14 0 14 14 10 4 0 14 14 14 0
9 9 9 12 15 12 9 9 1 1 1 1 1 1 1 1 15 16 5 16 7 18 7 16 3 3 13 15 3 3 16 7 16 16 16 16 13 16 16 16 7 10 7 5 7 5 13 10 18 5 5 16 5 1
9 9 9 9 17 12 9 9 9 1 1 1 9 9 1 15 15 5 8 5 18 18 13 7 5 2 2 3 16 18 18 18 18 18 18 18 18 18 18 18 18 18
9 9 9 9 17 12 9 9 1 9 9 9 9 1 1 15 15 16 16 7 18 18 18 7 16 16 2 2 16 13 18 18 7 7 7 7 7 18 18 18 18 10 18 7 10 18 18 18 18 18 18 18 18 18 18 18 18 18
9 9 9 12 12 12 9 9 1 1 9 9 9 9 1 15 15 7 14 14 14 18 14 14 5 5 2 15 5 5 1 18 14 18 18 18 14 14 10 1 10 1 1 10 9 10 10 1 18 18 1 1 18 18 14 14 14
9 9 12 12 12 9 9 9 1 1 1 1 1 1 2 15 8 14 14 14 14 14 8 8 2 15 0 14 1 14 14 14 14 0 0 14 14 14 14 14 14 14 1 18 14 14 14 1 1 18 14 14 14 1 14 14 14 14 14 14 14 14 14 1
9 9 9 12 12 12 9 9 9 9 1 9 1 9 1 1 15 15 10 10 18 10 10 10 10 10 2 15 10 10 1 1 10 10 10 10 10 10 10 10 10 10
9 9 12 15 12 1 1 1 1 1 1 1 1 1 18 13 2 15 13 13 16 13 13 13 13 13 13 2 15 16 3 18 18 13 13 13 13 13 13 13 13 13 13 13 13 13
9 9 9 17 12 9 9 9 1 1 1 1 13 16 15 2 16 16 16 7 7 7 5 16 16 2 15 5 5 18 18 13 7 7 7 5 18 18 18 7 18 18 7 10 18 14 13 7 18 1 18 7 7 7 13
12 9 9 12 17 12 1 1 1 1 1 1 18 16 16 15 2 3 8 8 8 5 16 16 3 3 2 15 16 16 10 14 16 8 13 0 5 8 5 0 8 14 5 5 10 7 14 5 16 5 18 16 5 5 5 3 4
9 9 12 17 12 1 1 1 1 18 18 16 16 16 2 15 3 3 16 3 8 16 16 3 3 2 15 16 16 14 5 16 16 16 5 5 16 16 16 16 16 16 16 16 16 16 16 16 16
9 9 9 12 17 12 9 9 1 10 7 5 16 16 7 15 17 1 7 16 8 16 7 16 16 5 2 15 18 18 18 18 18 13 7 18 7 5 18 18 7 7 18 18 18 10 1 1 18 18 18 9 7 7 18 7 7
14 18 9 9 17 12 9 1 14 0 8 8 8 0 2 15 0 8 8 8 0 0 8 8 0 2 15 0 0 14 14 0 0 14 14 14 14 14 14 0 0 0 14 10 18 14 14 14 0 1 0 14 0 1 8 4
5 8 0 9 2 10 0 0 8 8 8 8 8 8 8 15 2 8 8 8 8 8 8 8 8 8 8 8 8 15 15 8 8 0 8 8 8 8 14 5 8 8 8 16 8 0 10 18 18 0 0 0 10 8 8 0 8 8
13 7 7 16 15 10 18 5 5 16 5 16 16 16 16 2 2 16 16 8 5 5 16 8 5 5 2 15 13 13 7 18 5 5 7 5 16 5 5 5 5 7 5 7 10 18 18 7 5 7 10 5 7 5 5 16
16 8 0 0 2 2 0 0 0 8 8 0 8 8 16 2 15 16 8 8 8 0 0 8 0 0 2 15 0 8 0 18 0 0 14 0 8 0 0 14 18 0 0 14 10 18 18 14 14 14 1 0 1 14 0 8
7 18 18 18 2 18 18 18 14 7 18 18 3 16 15 15 18 18 18 18 18 18 18 7 5 18 2 15 18 7 18 1 18 18 18 18 18 18 18 18 18 18 18 1
8 14 14 15 14 14 8 8 0 0 14 14 8 0 15 15 0 0 8 0 0 8 8 8 0 2 15 8 8 14 14 0 0 14 14 14 14 14 14 14 14 14 14 14 14 14
5 5 14 14 2 14 14 14 14 14 14 14 5 14 14 14 15 15 14 14 5 14 14 14 0 14 14 2 15 14 0 14 5 14 14 14 14 14 14 14 14 14 14 14 14 14
8 0 0 0 15 2 14 0 0 0 0 0 0 8 2 15 0 0 0 0 14 0 0 2 2 0 0 0 0 0 14 0 14 14 14 14 14 14 14 14 0 0 1 14 14 14 14 1 1 18 0 0 0
16 16 13 13 15 13 13 15 16 3 7 13 16 16 15 13 13 16 13 13 13 13 13 13 2 7 3 3 13 13 13 13 13 13 13 13 13 13 13 13

Testing Clustering Algorithm

Apply the algorithm on 1st exposure of 2025/01/18 and detector 11



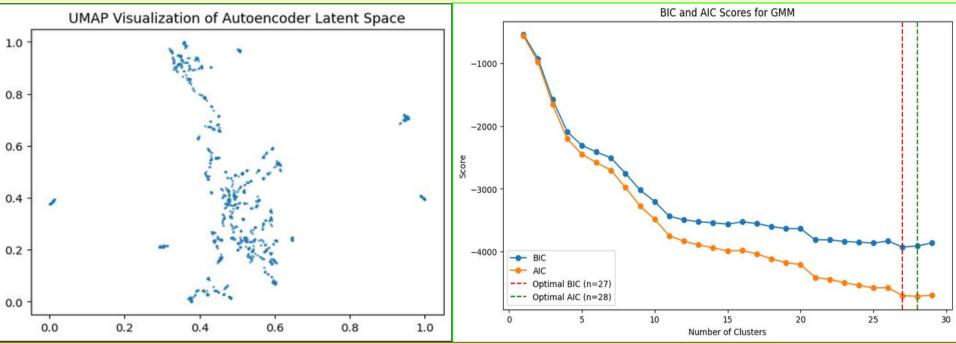
How much our clustering is effective!



Apply GMM for clustering and the number of clusters is 29

Latent Space Clustering

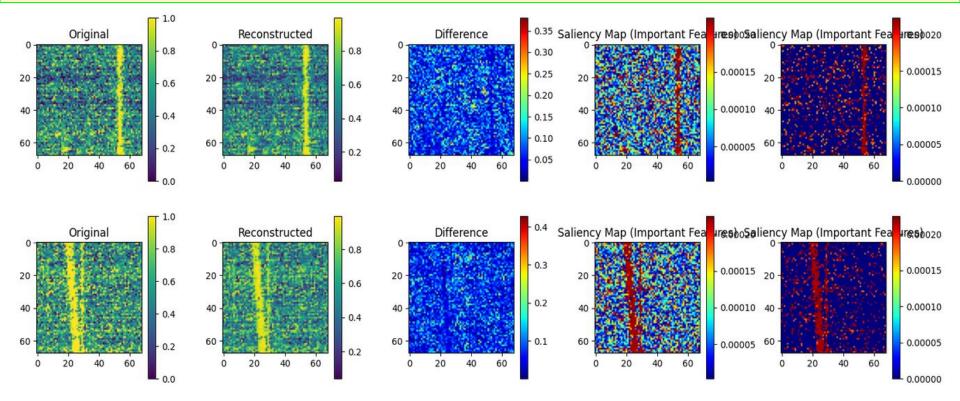
- 1) Apply UMAP algorithm on latent space with **n_neighbors = 3**
- 2) Apply GMM for clustering the latent space and measure the BIC and AIC

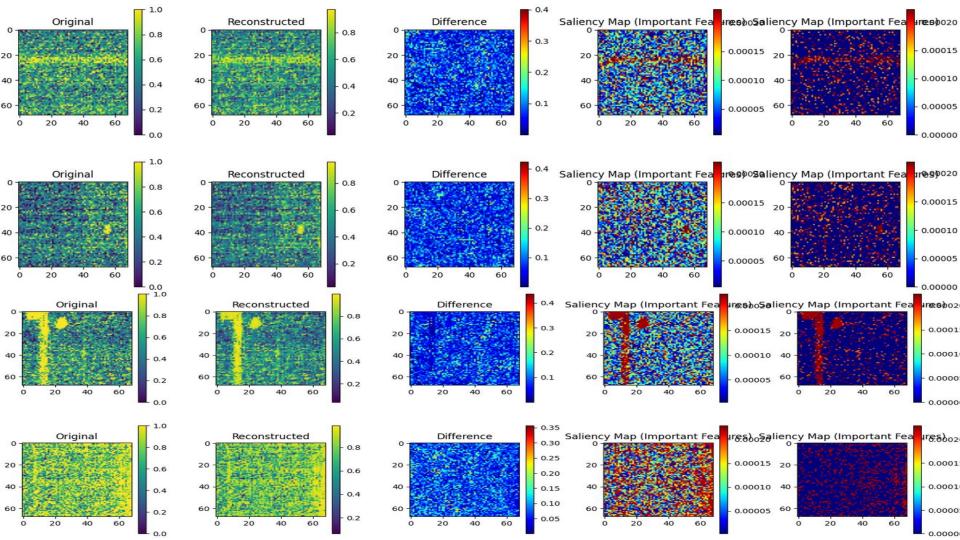


Do we need a large number of clusters for effective clustering? Does this requirement arise from the training process, or is it a result of high data diversity?

Does the machine learn using the correct features in the data?

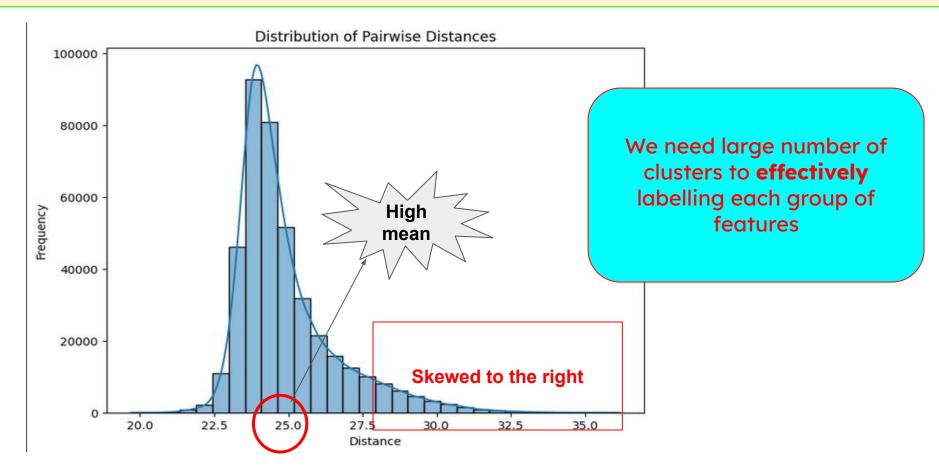
- 1) Apply Gradient-based method to find the most effective pixels in training
- 2) **Saliency map** highlights the most influential pixels





Calculating data diversity

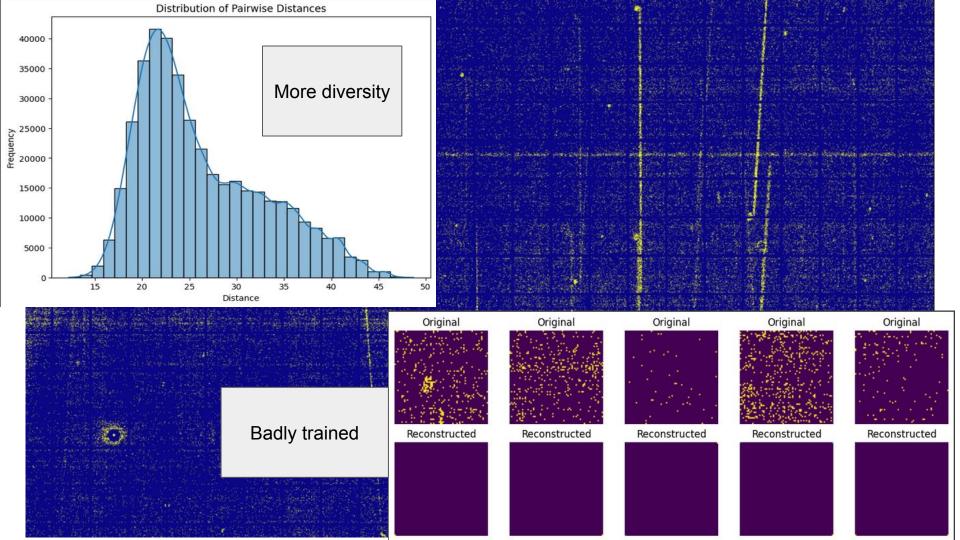
1) Computes the pairwise Euclidean distances between rows of data



Solution 1: Removing Background to reduce the data diversity

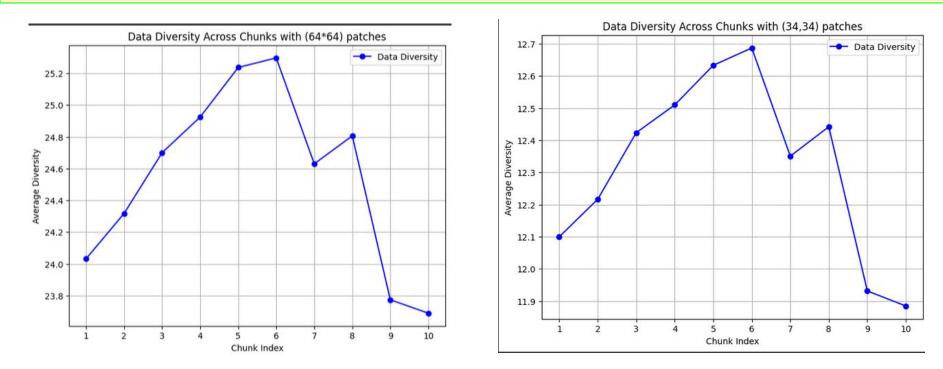
- 1) Removing data background should be a in a way that important features of cosmics retained!
- 2) Subtracting pixels with values lower that **0.95** (the data is normalized between 0 and 1)
- 3) Retaining pixels with values greater than Median + (MAD*0.5)
- 4) Subtracting smaller regions (clusters of connected pixels) that have fewer than **min_neighbor = 2** neighboring pixels

```
if mad:
   med = np.median(part)
   mad value = np.median(np.abs(part - med)) / 0.6745 # Compute MAD
   # Define positive and negative thresholds
   pos thr = med + (mad value*0.5)
   neg thr = med - (mad value*0.5) # Fix: Should subtract instead of adding
   # Fix: Use bitwise operators instead of 'and'
   intense mask = (part > pos thr)
   # Retain only the intense areas
    part = np.where(intense mask, part, 0) # Set background to 0
if remove mean background:
   mask part = part > threshold mean
   part = part * mask part
if remove neighbor:
   labeled clusters, num labels = label(part) # Label connected components
    refined part = np.copy(part)
   for label val in range(1, num labels + 1): # Labels start from 1
        cluster pixels = np.where(labeled clusters == label val)
        if len(cluster pixels[0]) < min neighbor:</pre>
            refined part[labeled clusters == label val] = 0 # Remove small clusters
```



Solution 2: Dividing into different channels without removing background

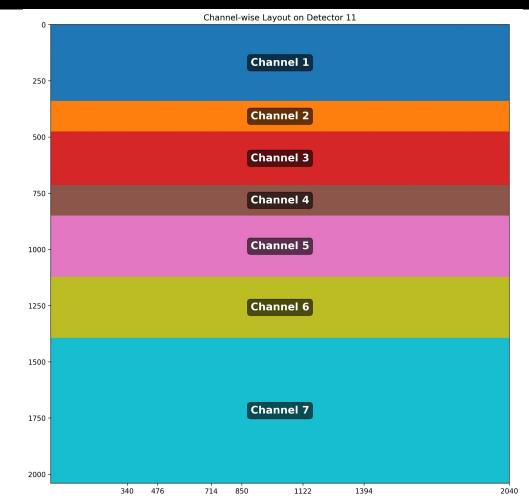
- 1) Considering two different size patches including (900, 64, 64) and (3600, 34, 34)
- 2) Splitting each image into 10 horizontally channels to reduce the complexity in the latent space
- 3) Applying BIC and AIC algorithms on each channel to find the effective number of clusters (Minimum numbers of clusters, maybe we need more)
- 4) Applying GMM on latent space and find the clusters including cosmics



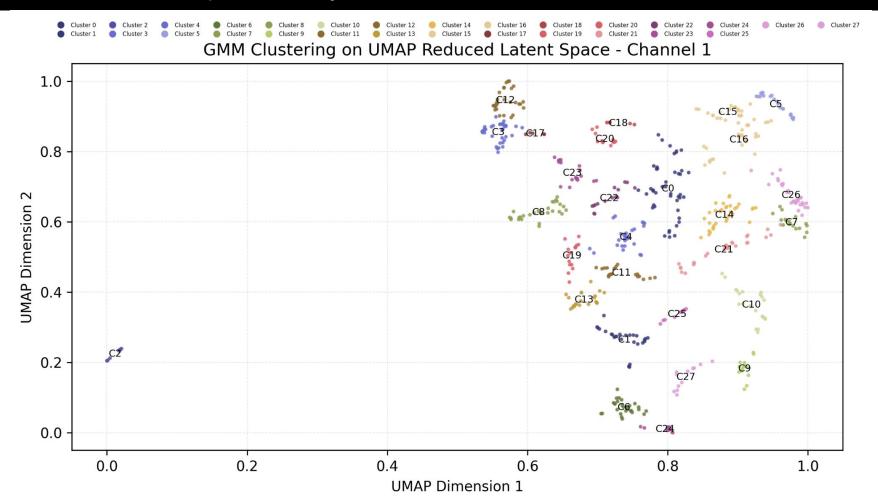
Channel-Wise Analysis Pipeline

- 1) Channel-Wise Horizontally Separation
- 2) Making latent space for each channel and applying umap to project high-dimensional latent space to 2D space
- 3) Applying BIC and AIC algorithms on each channel to find the effective number of clusters (This is defining Minimum numbers of clusters, maybe we need more)
- 4) Applying GMM on latent space and find the clusters including persistences
- 5) Labelling persistences clusters based on the NISP persistences features in spectro images
- 6) Segmenting persistences from each channel
- 7) Making masking maps based on this segmentation and concatenating all of the making maps

Channel-Wise Layouting



Channel 1 Latent Space Analysis

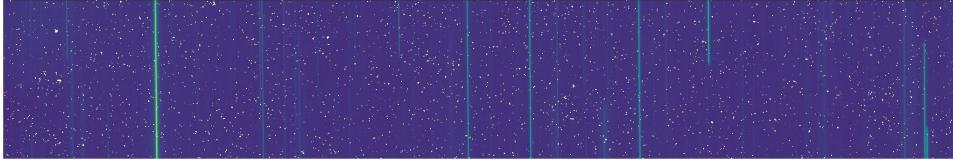


Channel 1 : Slew dark Persistence Unsupervise Recognition

GMM Clustered Channel 1 (n_clusters=28)

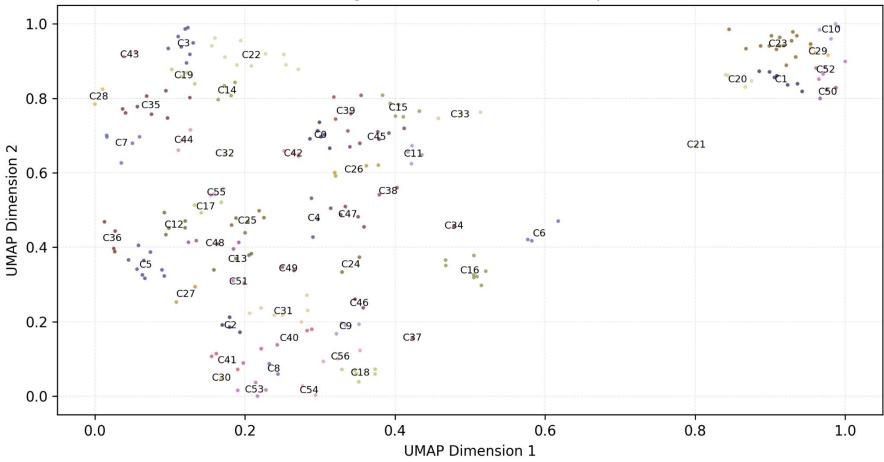
16
15
15
16
16
2
15
5
16
16
15
16
16
15
16
16
15
15
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
16
1

RGS180 and Tilt is 0 - Channel 1



Channel 2 Latent Space Analysis

K-Means Clustering on UMAP Reduced Latent Space - Channel 2



Channel 2 : Slew dark Persistence Unsupervise Recognition

GMM Clustered Channel 2 (n clusters=57)

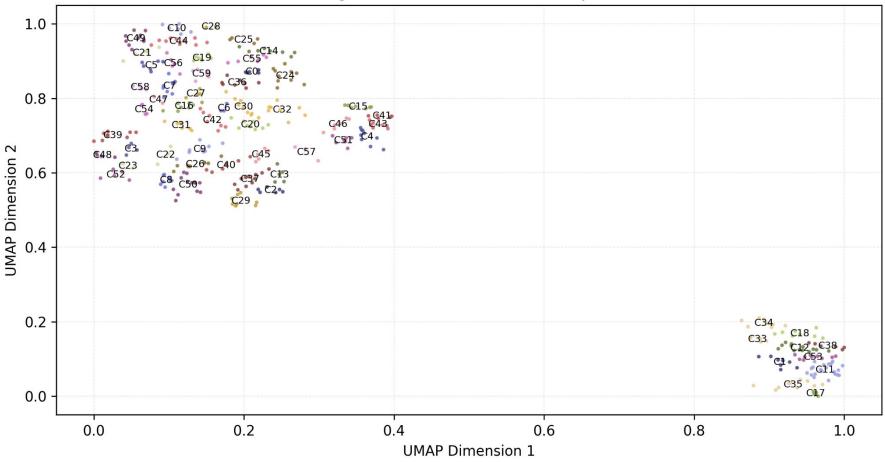
23 23 23 29 52 1 23 23 23 10 3 19 22 35 14 6 6 22 22 19 22 22 14 44 3 22 34 16 22 22 35 3 22 35 7 44 28 35 57 7 43 35 19 3 28 35 7 35 7 43 14 14 3 3 3 19 44 29 1 23 23 57 1 23 1 23 1 23 10 39 0 15 39 45 6 21 33 45 39 39 33 15 45 0 47 11 16 0 0 11 15 15 0 47 38 38 42 26 24 47 47 4 4 47 26 42 42 45 11 26 42 26 11 15 39 0 45 45 15 29 23 23 23 52 1 23 52 1 23 52 23 10 55 48 25 32 55 16 16 55 12 48 48 13 17 36 12 13 4 16 25 49 51 5 17 25 25 5 5 27 27 13 5 5 36 49 48 5 5 12 13 5 5 36 25 51 12 36 12 48 25 17 50 20 20 52 1 1 5 50 20 10 20 33 46 9 8 16 16 18 53 2 40 56 56 41 2 40 37 16 31 31 18 54 8 31 2 9 56 41 31 46 18 41 41 31 9 9 40 2 40 18 53 30 31 18 54 53 24 31 18 53

RGS180 and Tilt is 0 - Channel 2



Channel 3 Latent Space Analysis

GMM Clustering on UMAP Reduced Latent Space - Channel 3



Channel 3 : Slew dark Persistence Unsupervise Recognition

GMM Clustered Channel 3 (n_clusters=60)

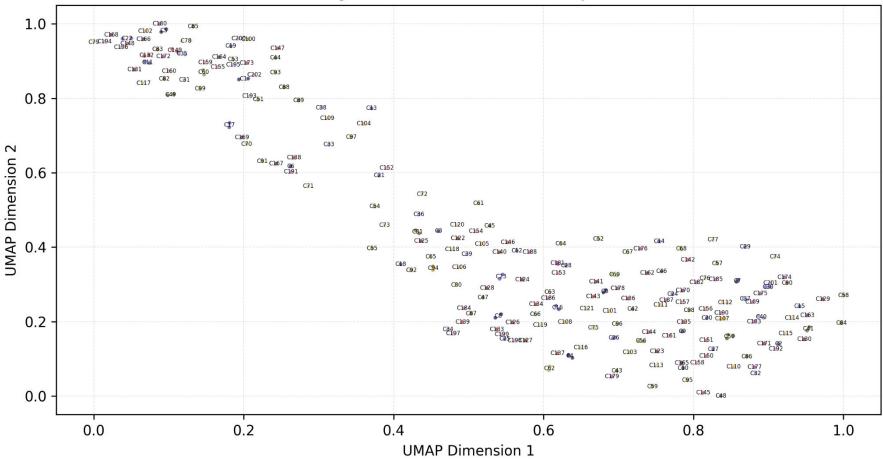
11 53 11 38 35 17 11 53 12 12 12 12 12 49 44 5 21 48 10 36 19 10 49 14 56 56 49 44 39 21 27 49 10 44 44 49 71 25 5 49 44 16 11 11 53 11 17 17 11 1 53 33 33 1 1 22 26 48 52 50 50 9 40 50 50 26 26 50 3 3 9 26 26 50 9 50 8 9 40 9 9 8 50 31 9 22 22 26 26 8 8 50 26 50 8 50 8 50 8 8 50 50 6 1 11 11 53 18 35 11 12 12 18 18 18 12 0 55 48 52 14 55 36 36 24 55 14 0 39 39 16 31 24 36 55 24 11 11 53 53 35 35 11 38 38 12 1 38 12 43 43 85 35 41 51 46 46 43 39 57 57 4 41 41 46 46 15 15 51 46 4 15 46 51 15 43 4 43 4 4 12 11 53 1 35 17 35 12 1 12 35 34 12 39 52 19 19 7 7 56 56 59 79 54 31 58 19 58 58 16 44 28 11 11 11 11 35 17 11 11 53 33 33 33 34 34 40 23 52 29 13 37 37 29 45 13 3 23 40 31 29 29 37 37 37 45 29 3 31 31 42 42 20 30 32 32 42 30 32 30 20 24 32 30 42 38 18 18 17 38 1 18 34 34 18 34 34 34 48 52 32 20 9 30 32 20 42 6 20 39 39 42 31 6 20 30 30 20

RGS180 and Tilt is 0 - Channel 3



Channel 4 Latent Space Analysis

GMM Clustering on UMAP Reduced Latent Space - Channel 4



Channel 4 : Slew dark Persistence Unsupervise Recognition

GMM Clustered Channel 4 (n_clusters=203)

166 5 172 11 17 72 5 149 35 1 1 53 195 51 88 70 6 199 153 141 121 87 63 178 16 26 62 4 96 75 127 186 8 0 197 66 108 26 133 101 198 16 75 62 119 134 124 136 4 4 69 26 193 184 126 143 139 34 25 8 78 132 83 131 49 148 22 149 11 117 31 60 160 33 109 169 138 81 72 3 125 105 92 188 146 140 54 71 54 0 36 94 94 12 181 128 162 23 28 23 80 187 123 47 69 39 120 181 134 45 51 72 81 18 65 106 178 122 21 152 78 200 132 85 17 79 180 19 60 155 44 100 82 144 97 91 191 37 185 112 7 30 98 107 40 7 151 27 110 130 182 37 107 41 15 157 135 171 114 84 107 192 50 86 50 189 24 163 174 115 112 56 90 30 2 20 41 201 49 170 164 202 194 196 49 168 102 159 95 95 13 173 93 89 104 167 46 116 14 165 142 76 111 10 88 29 41 179 56 161 55 67 144 42 50 37 150 57 129 9 111 165 158 190 137 48 176 58 77 56 59 52 183 103 175 95 88 74 177 345

RGS000 and Tilt is 0 - Channel 4



RGS180 and Tilt is 4 - Channel 4

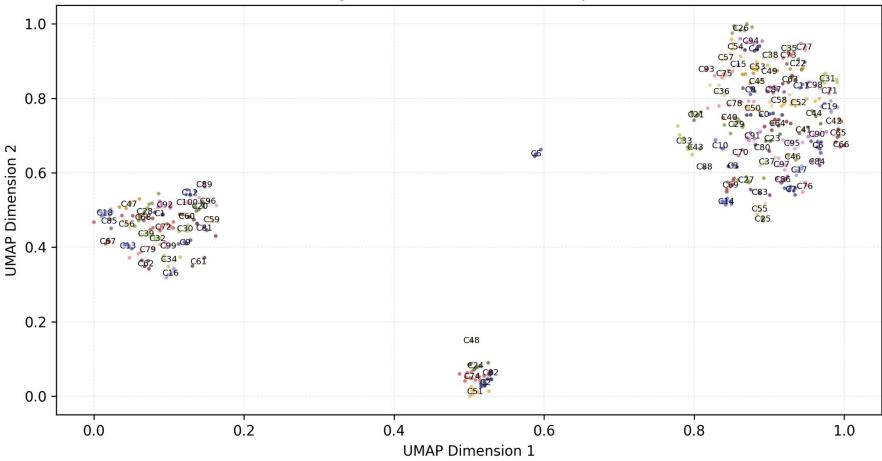


RGS000 and Tilt is -4 - Channel 4

RGS180 and Tilt is 0 - Channel 4

Channel 5 Latent Space Analysis

GMM Clustering on UMAP Reduced Latent Space - Channel 5

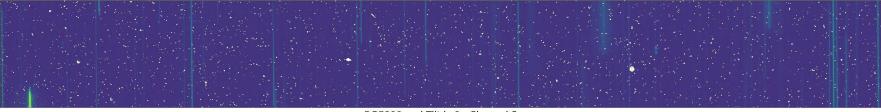


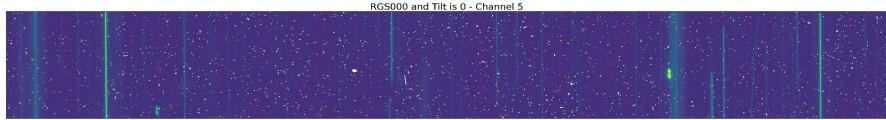
Channel 5 : Slew dark Persistence Unsupervise Recognition

GMM Clustered Channel 5 (n clusters=101)

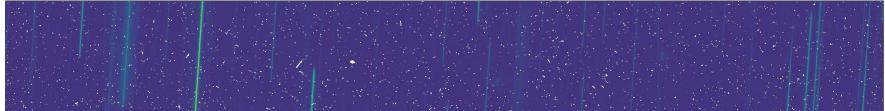
 17
 39
 25
 28
 62
 62
 99
 60
 26
 30
 100
 100
 40
 16
 43
 85
 20
 63
 77
 55
 22
 13
 55
 77
 41
 52
 63
 77
 40
 52
 63
 77
 41
 52
 63
 77
 41
 52
 63
 77
 41
 52
 63
 77
 41
 52
 63
 77
 41
 52
 63
 77
 41
 52
 63
 77
 41
 52
 63
 77
 41
 52
 63
 77
 41
 52
 63
 77
 41
 52
 63
 77
 41
 52
 63
 77
 41
 52
 63
 77
 41
 52
 63
 77
 41
 52
 63
 77
 41
 52
 63
 77
 41
 52
 63
 77
 41
 52
 63
 77
 41
 52
 63
 41
 52
 51
 77
 41
 52
 63
 41
 52
 <

RGS180 and Tilt is 0 - Channel 5



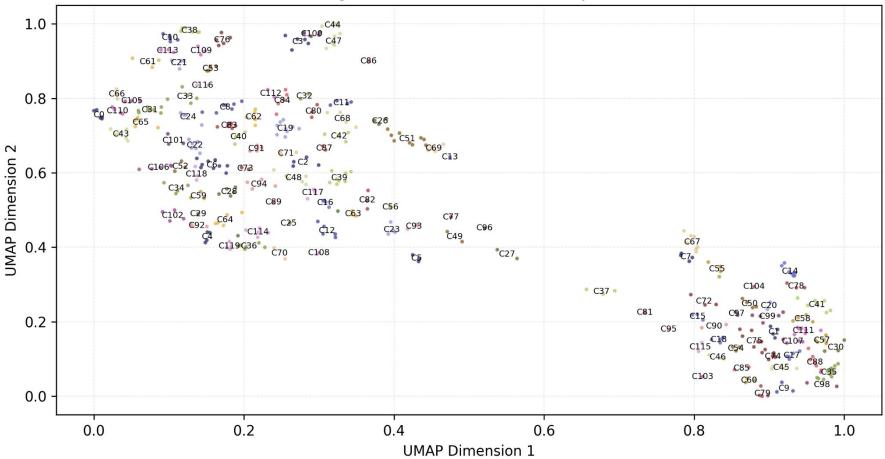


RGS180 and Tilt is 4 - Channel 5



Channel 6 Latent Space Analysis

GMM Clustering on UMAP Reduced Latent Space - Channel 6



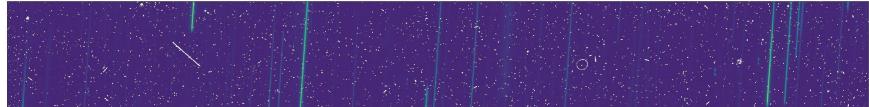
Channel 6 : Slew dark Persistence Unsupervise Recognition

GMM Clustered Channel 6 (n_clusters=120)

RGS000 and Tilt is 0 - Channel 6



RGS180 and Tilt is 4 - Channel 6

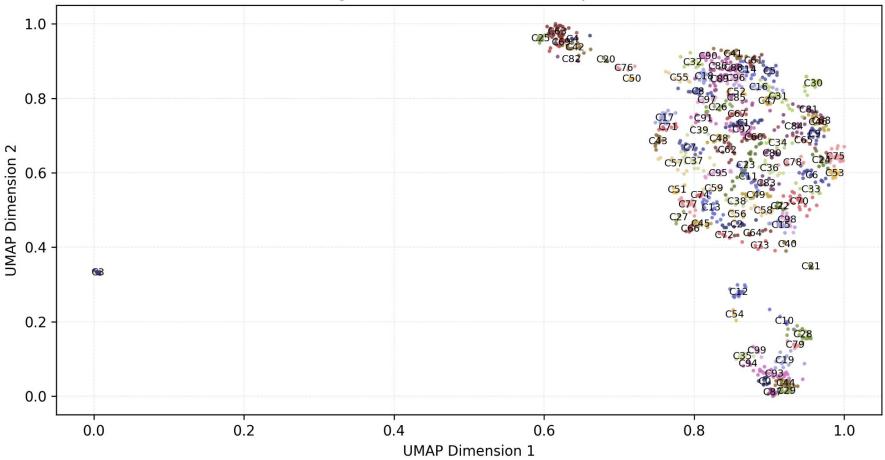


RGS180 and Tilt is 0 - Channel 6



Channel 7 Latent Space Analysis

GMM Clustering on UMAP Reduced Latent Space - Channel 7

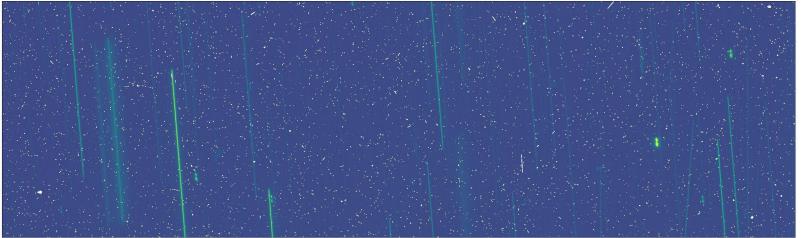


Channel 7: Slew dark Persistence Unsupervise Recognition

GMM Clustered Channel 7 (n clusters=100)

 44
 42
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 95
 <td

RGS000 and Tilt is -4 - Channel 7



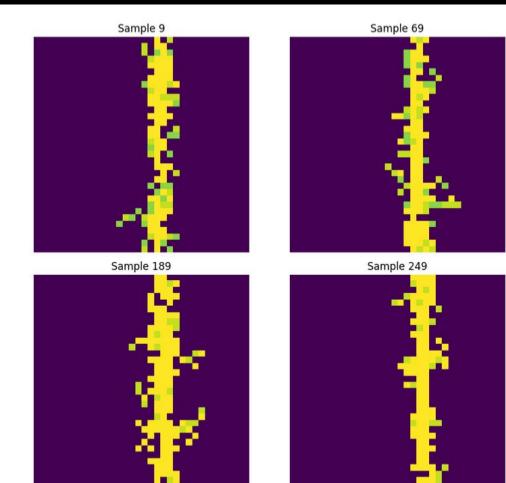
Segmentation Pipeline

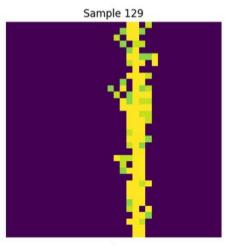
Step	What It Does	Effect
🔟 Thresholding (Mean)	Keeps pixels above a brightness threshold	Removes background
Standard Deviation Filtering	Keeps only high-contrast regions	Removes weak features
Connected Component Analysis	Groups connected bright pixels	Identifies separate objects
🖪 Small Cluster Removal	Removes small noisy blobs	Keeps only significant clusters

GMM Clustered Channel 1 (n_clusters=4)

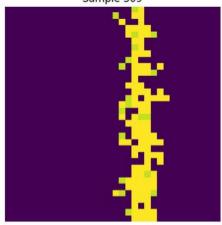
26 5 5 24 1 26 5 4 14	
4 4 4 18 4 12 12 4 4 1	1 1 1 1 1 1 1 1 1 1
19 8 20 13 20 25 8 8 8	22 37 37 38 30 40 30 30 30 30 30 30 30 30 30 30 30 30 30
1, 3 1 19 1 B B P Z	
16 11 16 16 10 B 15 16 1	<u>2</u> 15 2 3 2 10 10 3 35 77 71 73 57 75 8 28 7 10 70 77 3 37 37 37 37 37 37 37 37 37 37 37 3

Segmentation Examples





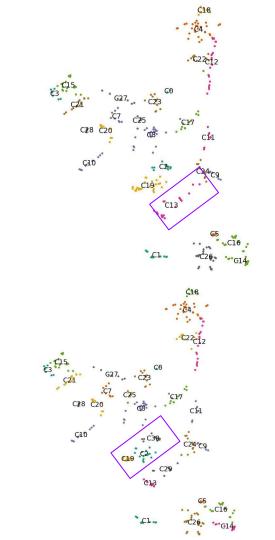
Sample 309

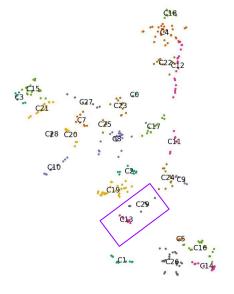


Next Steps ...

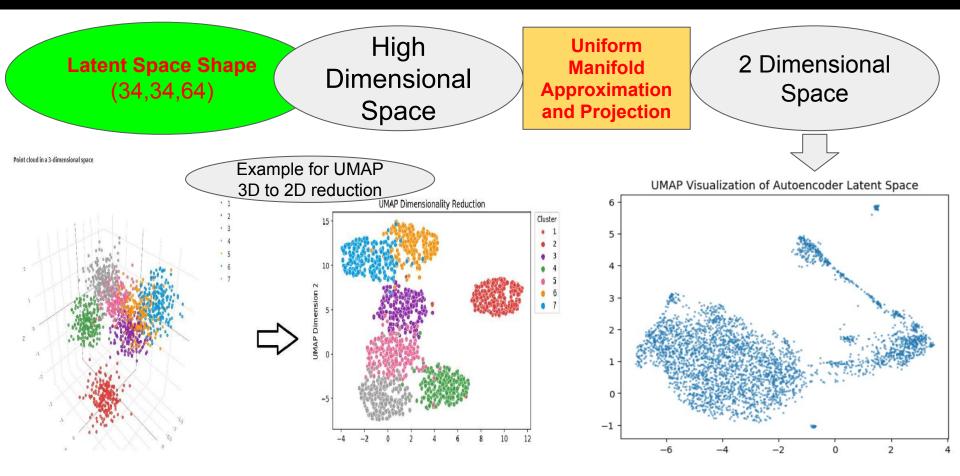
- 1) Merging latent space of 7 channels
- 2) Labeling clusters including persistances and define a neighboring area for each cluster
- 3) Applying this approach on rest of exposures and consider any sample as persistence if they are in the neighboring area of persistence clusters
- 4) Applying segmentation on persistence samples and labelling them for supervise training
- 5) Training A UNet using the training data from segmented persistence samples
- 6) Apply the trained Unet on whole samples to improve the segmentation
- 7) ...







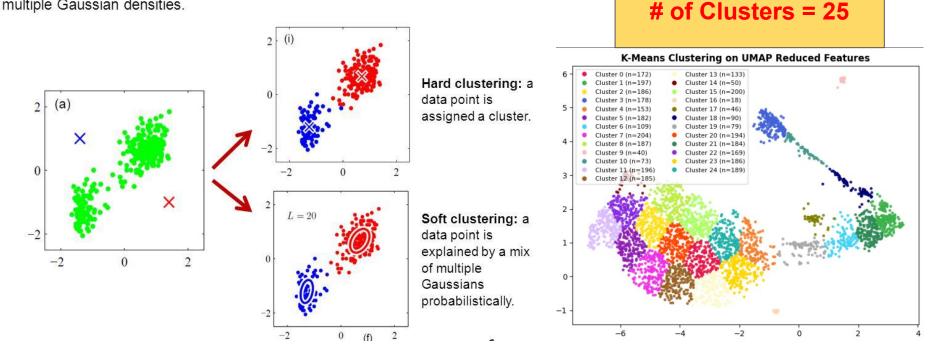
Latent Space Analysis - UMAP as A Dimensionality Reduction Tool



Latent Space Analysis - K-means for Clustering UMAP Space

ondon K-means vs GMM

Two standard methods are k-means and Gaussian Mixture Model (GMM). K-means assigns data points to the nearest clusters, while GMM represents data by multiple Gaussian densities.



K-means