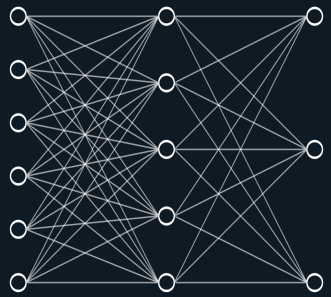
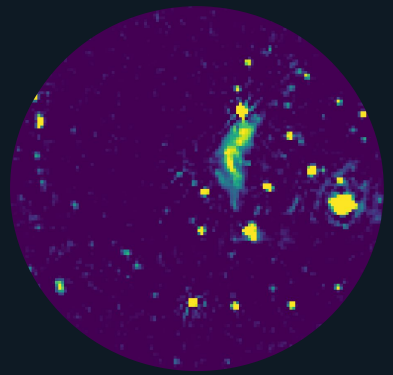
Classification of diffuse radio emission from LoTSSDR2



Markus Bredberg
26th of January 2026
Supervisor: Emma Tolley

SKACH EPFL



Classification of diffuse radio emission from LoTSSDR2

Credit: David Chifiriuc

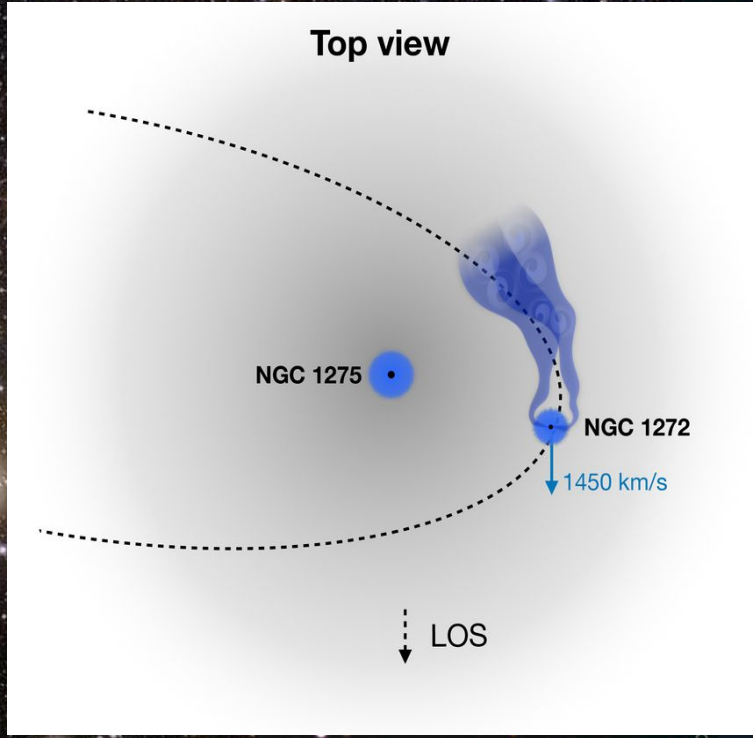
Credit: J.-C. Cuillandre, Euclid, ESA, NASA

Credit: radio: VLA 230–470 MHz, Gendron-Marsolais+17;
optical: SDSS, gri bands, Abolfathi+18

Credit: Marie-Lou Gendron-Marsolais (Université de Montréal), Julie Hlavacek-Larrondo (Université de Montréal), Maxime Pivin Lapointe - Own work, CC BY 4.0,

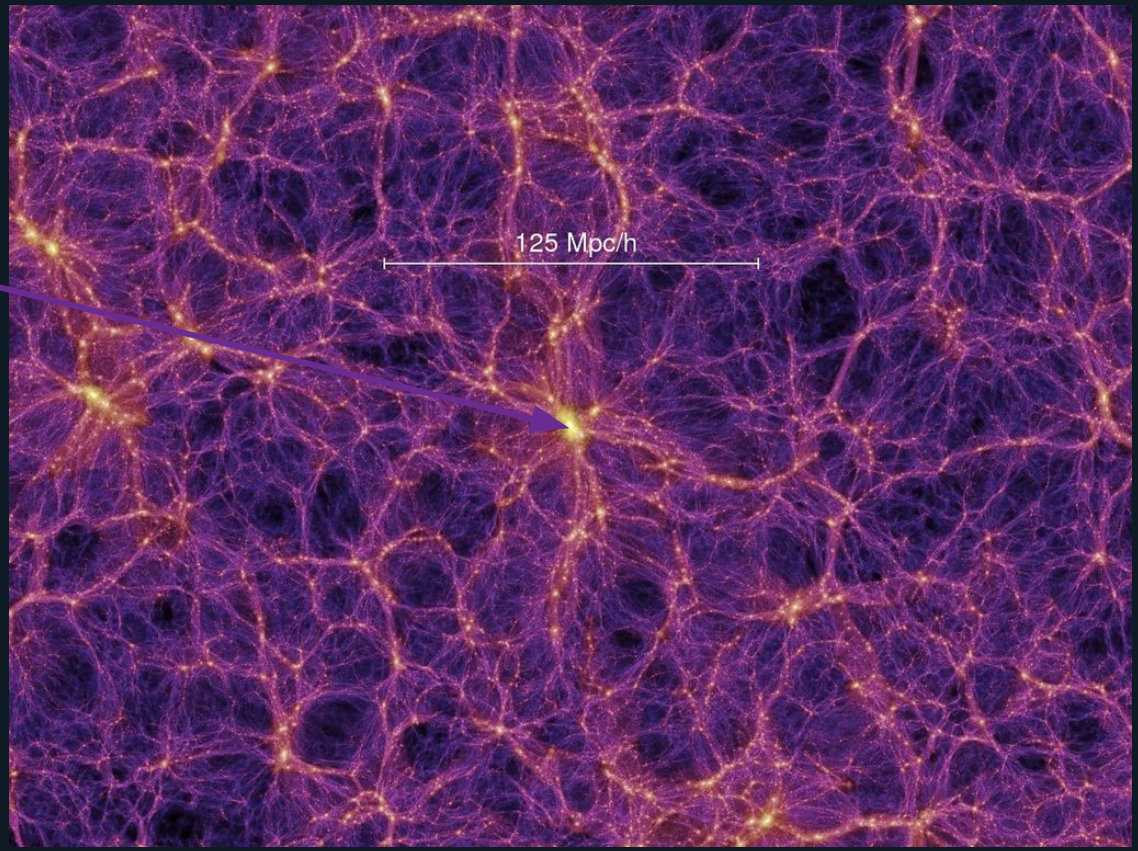
Credit: Gendron-Marsolais et al 2021 ApJ 911

Credit: Gendron-Marsolais et al 2021 ApJ 911



The presence of large-scale magnetic fields

Galaxy clusters: ~μG

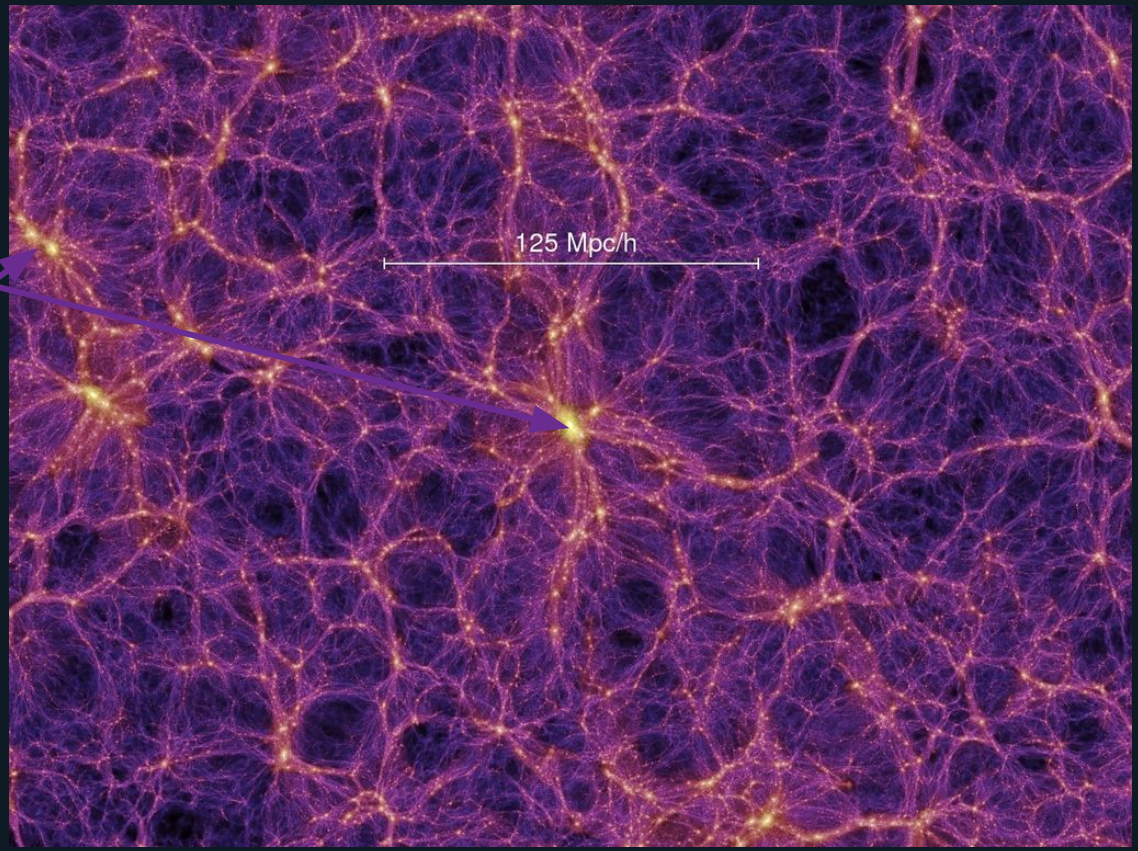


Credit: Millennium run

The presence of large-scale magnetic fields

Galaxy clusters: ~μG

Intercluster bridges: ~.1μG



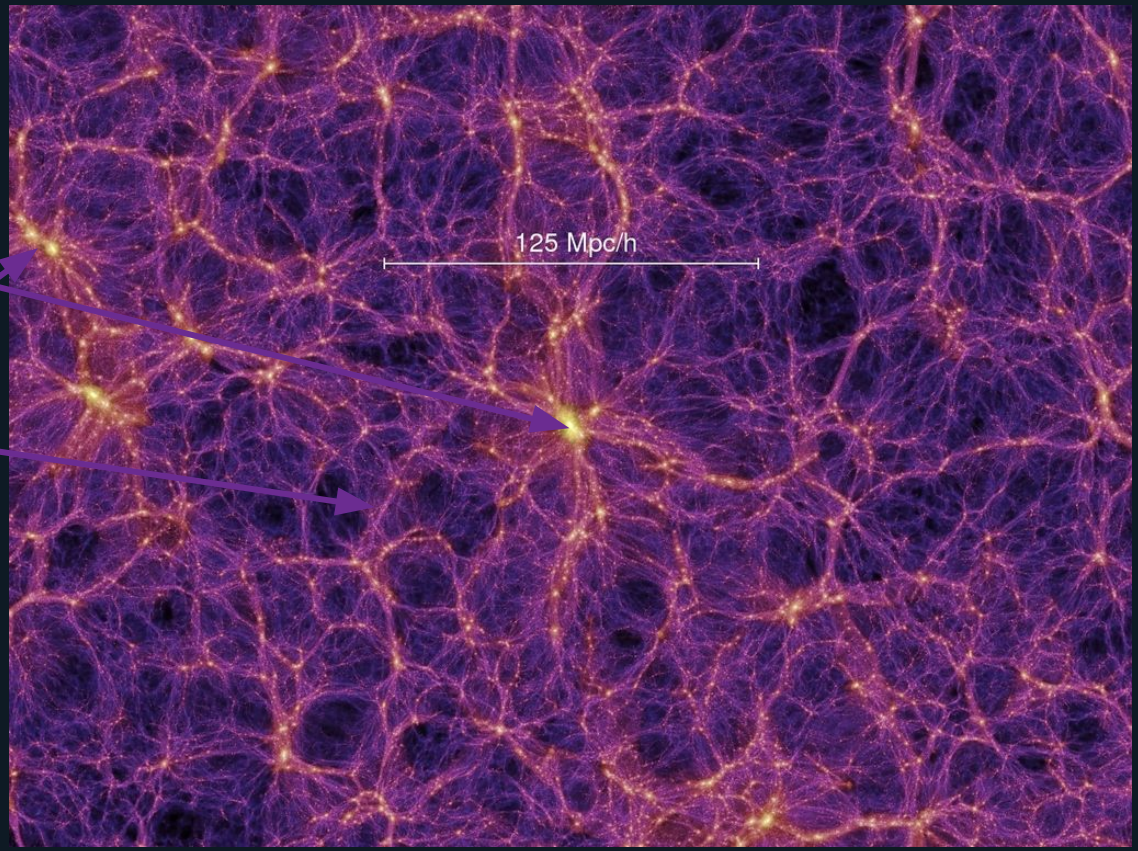
Credit: Millennium run

The presence of large-scale magnetic fields

Galaxy clusters: ~μG

Intercluster bridges: ~.1μG

Cosmic filaments: <nG



Credit: Millennium run

The presence of large-scale magnetic fields

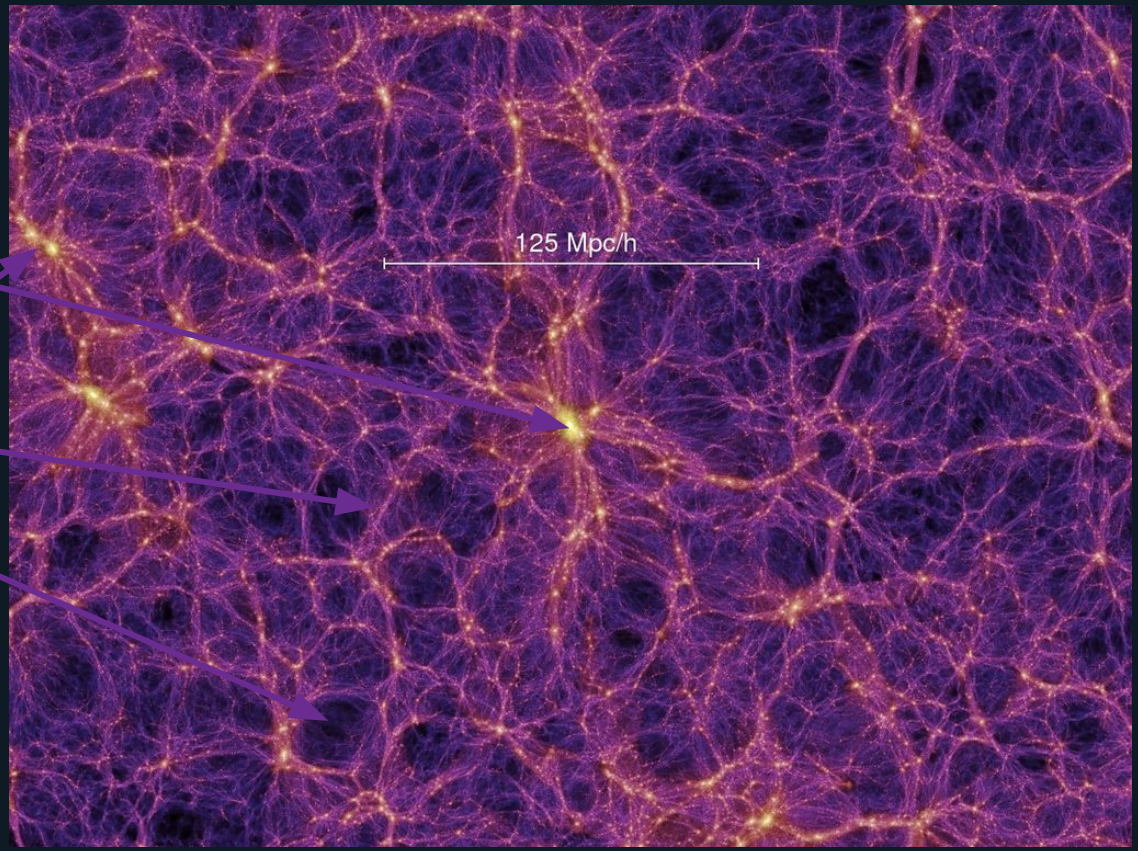
$\beta = p_{thermal}/p_{magnetic} \sim 100$

Galaxy clusters: ~μG

Intercluster bridges: ~.1μG

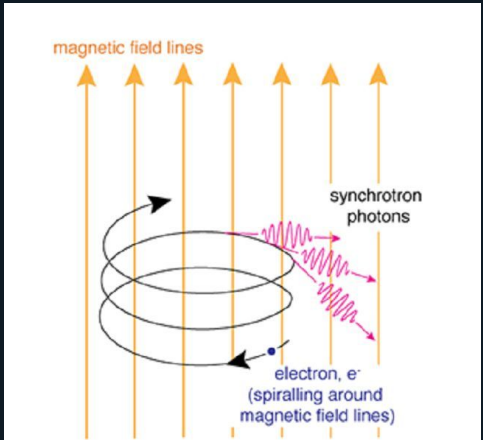
Cosmic filaments: <nG

Cosmic voids: 10^{-16}-10^{-10}?



Credit: Millennium run

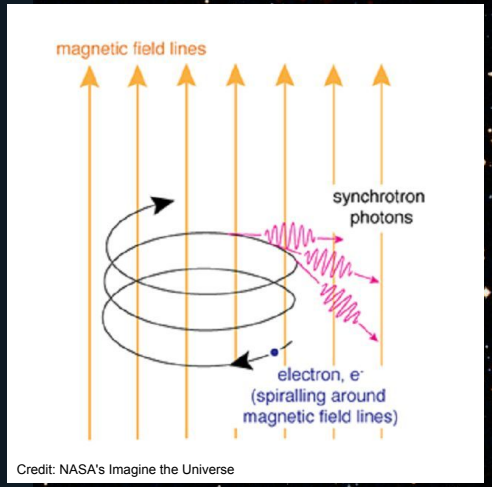
Synchrotron radiation is extended and complex



Credit: NASA's Imagine the Universe

Synchrotron radiation is extended and complex

Time to cover: ~Gyr
Dissipation time: ~10-100 Myr
→ In-situ re-acceleration



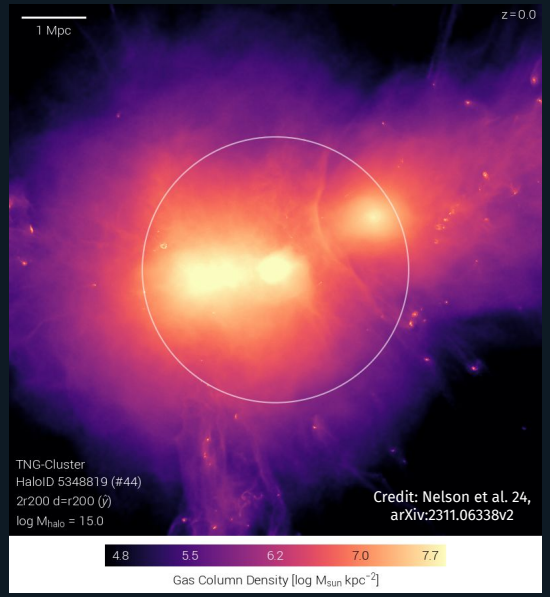
Credit: NASA's Imagine the Universe

← Size of the Milky Way

~Mpc

Credit: SARAO (Abell 3667)

Galaxy cluster mergers are extremely energetic



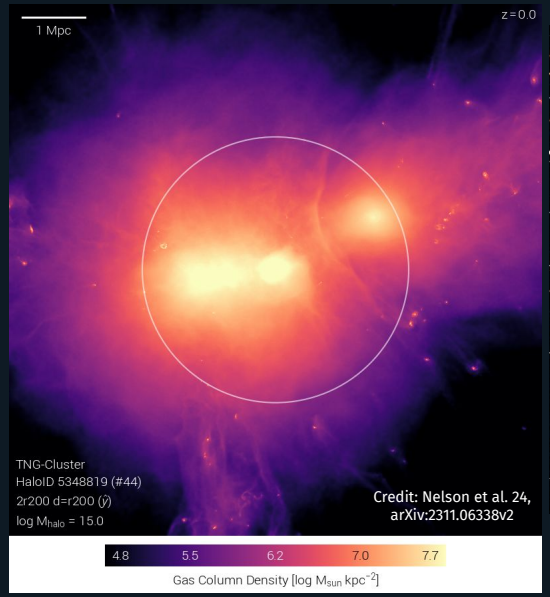
Rotation

Turbulence

Gas compression

Magnetic field growth

Galaxy cluster mergers are extremely energetic



Rotation

Turbulence

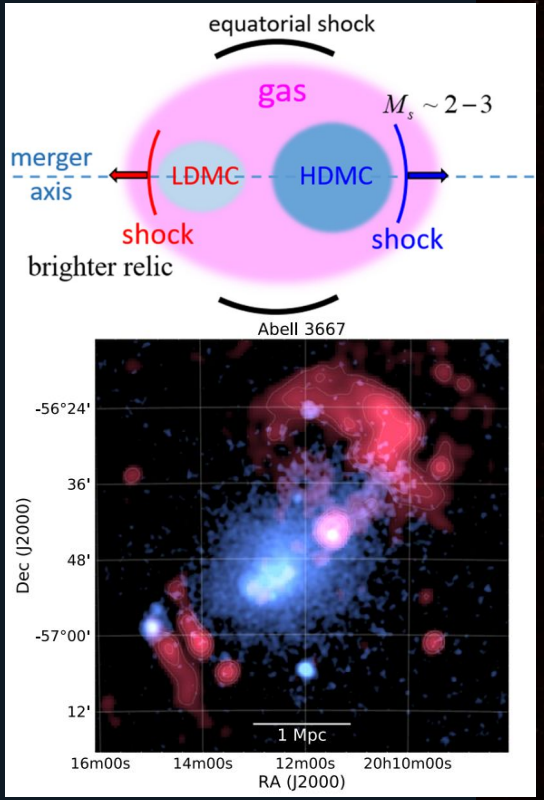
Gas compression

Magnetic field growth

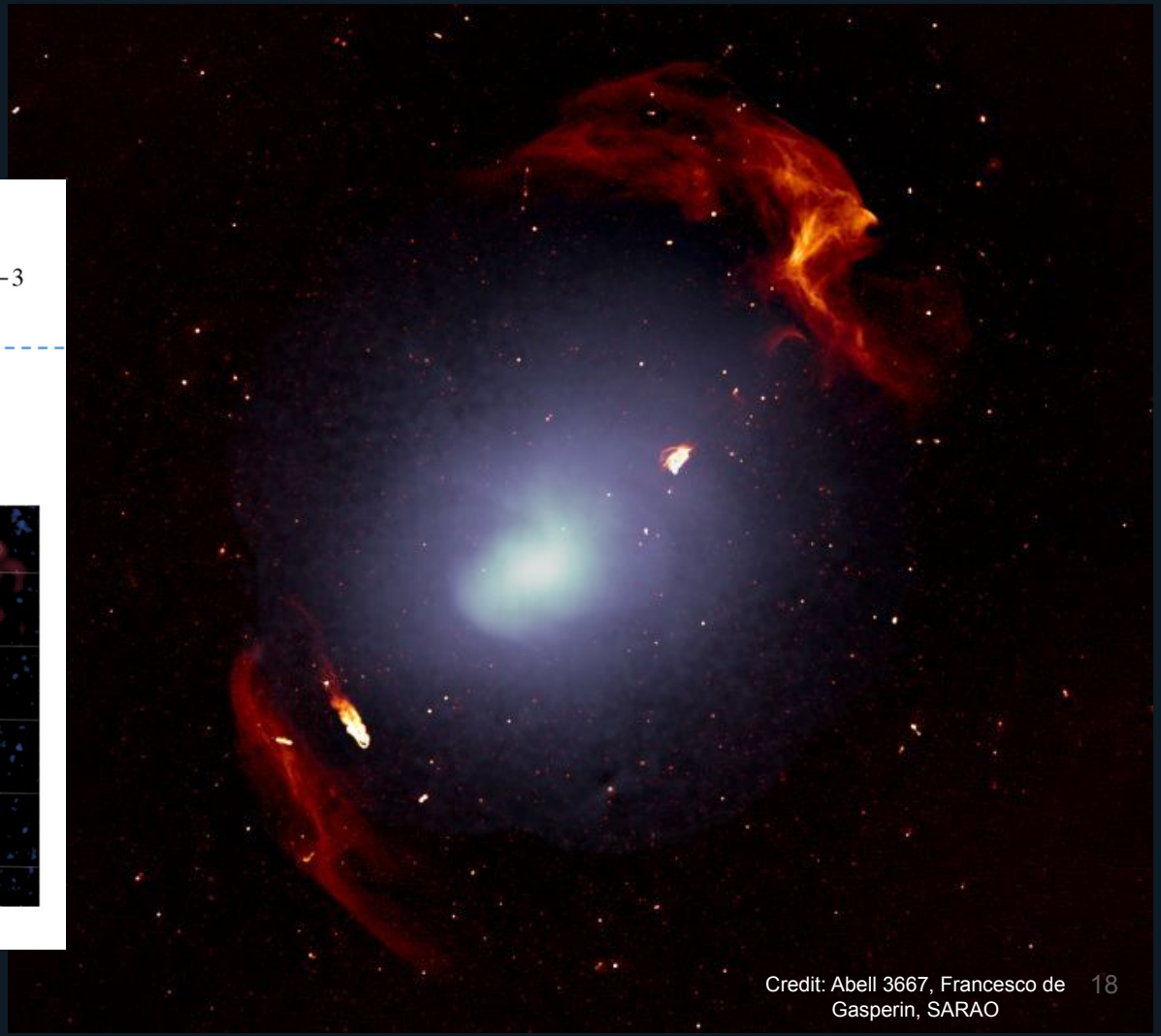
Radio relics are shock-accelerated

Credit: Abell 3667, Francesco de Gasperin, SARAO

Radio relics are shock-accelerated



Credit: MWA 170–231 MHz and ROSAT PSPC (Hurley-Walker et al. 2017; Voges et al. 1999)

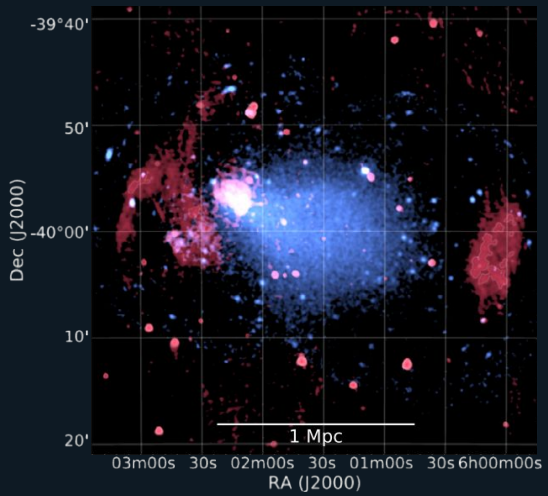


Credit: Abell 3667, Francesco de Gasperin, SARAO

Classification of diffuse radio emission from LoTSSDR2

Synchrotron radiation probes relativistic charged particles and magnetic fields

Radio Relic



MWA 170–231 MHz and ROSAT PSPC (Hurley-Walker et al. 2017; Voges et al. 1999)

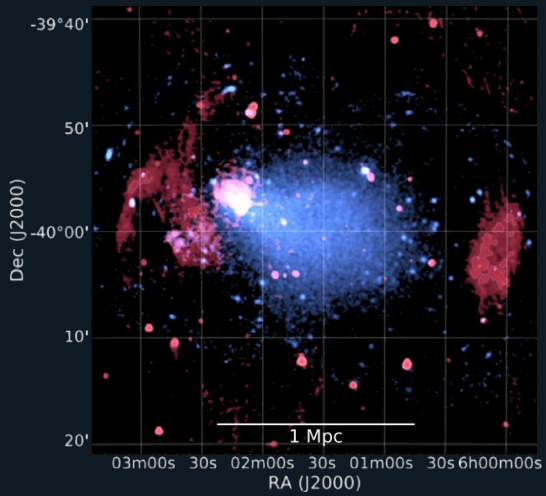
Periphery in merger

≳ 1 Mpc

~ 30-70 % polarisation

Synchrotron radiation probes relativistic charged particles and magnetic fields

Radio Relic



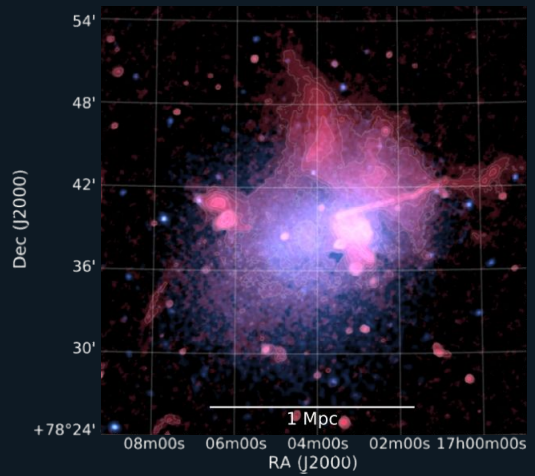
MWA 170–231 MHz and ROSAT PSPC (Hurley-Walker et al. 2017; Voges et al. 1999)

Periphery in merger

≳ 1 Mpc

~ 30-70 % polarisation

Radio Halo



LOFAR 120–170 MHz and XMM-Newton 0.4–1.3 keV (van Weeren et al. in prep)

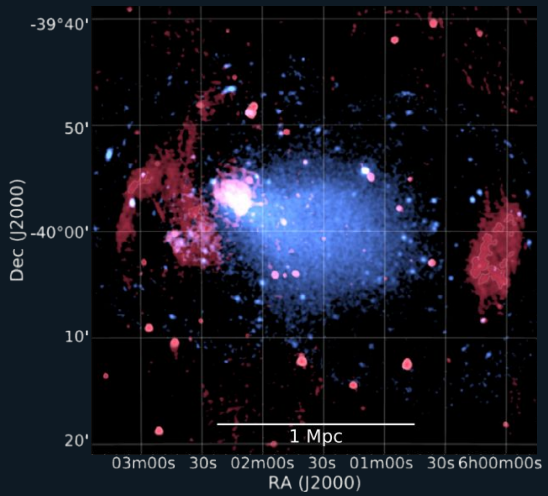
Centre of cluster

~300 kpc to ~2 Mpc

Large cluster mass

Synchrotron radiation probes relativistic charged particles and magnetic fields

Radio Relic



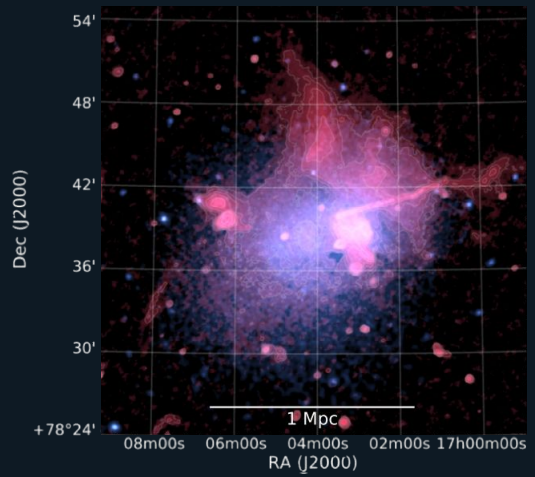
MWA 170–231 MHz and ROSAT PSPC (Hurley-Walker et al. 2017; Voges et al. 1999)

Periphery in merger

≳ 1 Mpc

~ 30-70 % polarisation

Radio Halo



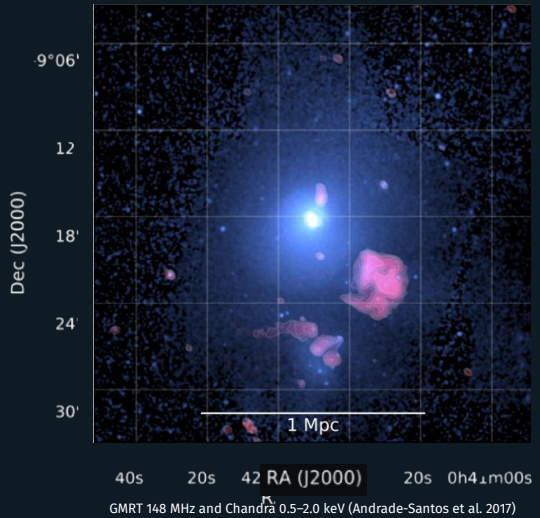
LOFAR 120–170 MHz and XMM-Newton 0.4–1.3 keV (van Weeren et al. in prep)

Centre of cluster

~300 kpc to ~2 Mpc

Large cluster mass

Revived Fossil Plasma



GMRT 148 MHz and Chandra 0.5–2.0 keV (Andrade-Santos et al. 2017)

Anywhere in cluster

~300 kpc

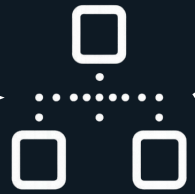
AGN origin, and ultra-steep radio spectra with breaks

- Growth and evolution of magnetic fields
- Cosmic-ray particle acceleration
- Dynamics, demographics and evolution of clusters

How to study diffuse cluster radio emission

X-ray and SZ preselection

Radio observation of galaxy cluster



Data “manipulation”

- Point source subtraction
- Tapering
- Multi-frequency analysis
- etc.

Specific explanation

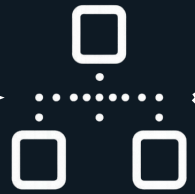
PSZ2 G122.30+54.52

- $M = 4.5 \times 10^{14} M_{\odot}$
- r_{500}=1.1 Mpc
- No diffuse emission

How to study diffuse cluster radio emission

X-ray and SZ preselection

Radio observation of galaxy cluster



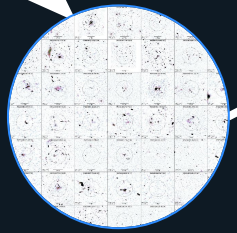
Data “manipulation”

- Point source subtraction
- Tapering
- Multi-frequency analysis
- etc.

Specific explanation

PSZ2 G122.30+54.52

- $M = 4.5 \times 10^{14} M_{\odot}$
- r_{500}=1.1 Mpc
- No diffuse emission



Statistical sample

General explanation

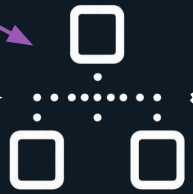
- Mini-halos more common in CCC, through sloshing.
- Duty cycle?
- B-field(M, z)?

How to study diffuse cluster radio emission

X-ray and SZ preselection

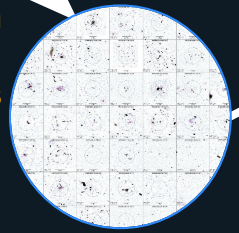
Radio observation of galaxy cluster

Slow



Data “manipulation”

- Point source subtraction
- Tapering
- Multi-frequency analysis
- etc.



Statistical sample

Specific explanation

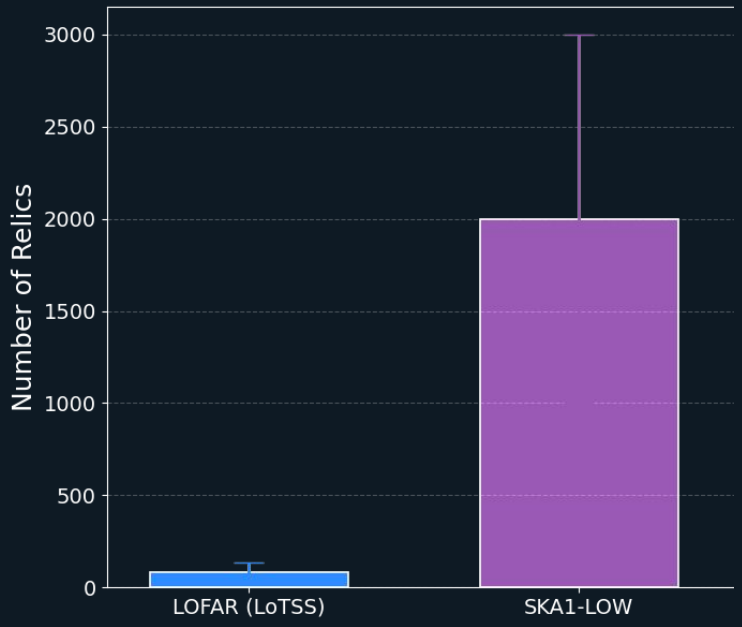
PSZ2 G122.30+54.52

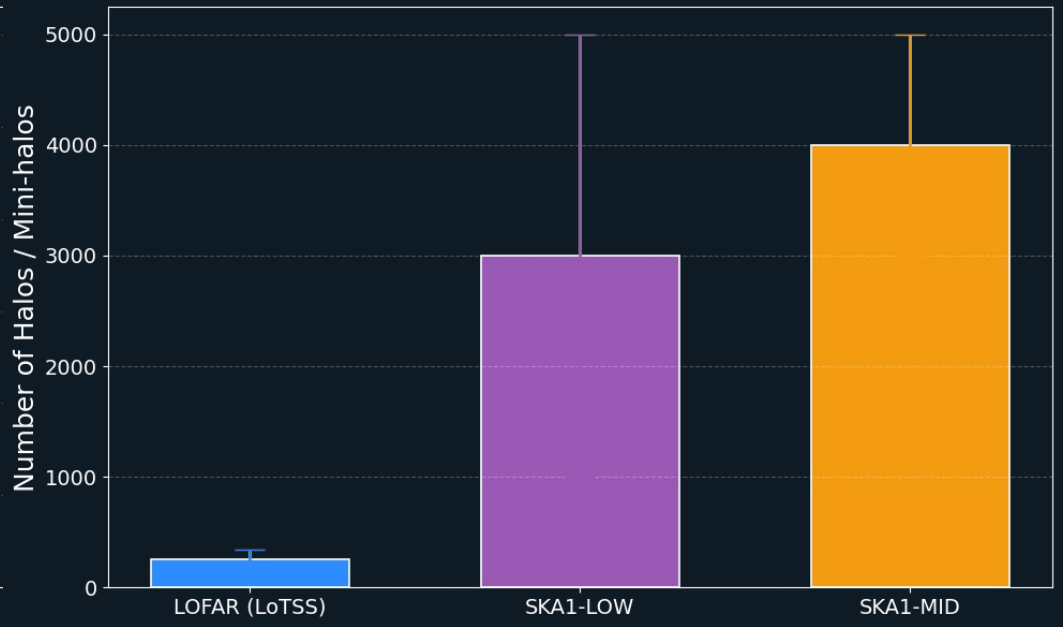
- $M = 4.5 \times 10^{14} M_{\odot}$
- r_{500}=1.1 Mpc
- No diffuse emission

General explanation

- Mini-halos more common in CCC, through sloshing.
- Duty cycle?
- B-field(M, z)?

Greater abundance calls for greater order





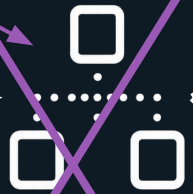
Credit: Cassano et al. 2014, arXiv:1412.5940
Botteon et al. 2022, https://doi.org/10.1051/0004-6361/202143020
SKAO 2021, SKA1 Scientific Use Cases (Rev. 04)
Gitti et al. 2018, https://doi.org/10.1051/0004-6361/2018327

How to study diffuse cluster radio emission

X-ray and SZ preselection

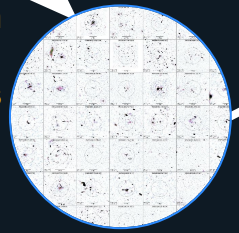
Radio observation of galaxy cluster

Slow



Data “manipulation”

- Point source subtraction
- Tapering
- Multi-frequency analysis
- etc.



Statistical sample

Specific explanation

PSZ2 G122.30+54.52

- $M = 4.5 \times 10^{14} M_{\odot}$
- r_{500}=1.1 Mpc
- No diffuse emission

General explanation

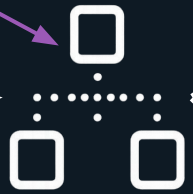
- Mini-halos more common in CCC, through sloshing.
- Duty cycle?
- B-field(M, z)?

How to study diffuse cluster radio emission

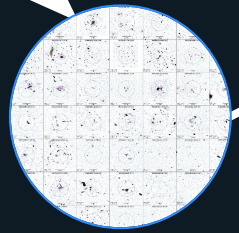
X-ray and SZ preselection

Radio observation of galaxy cluster

Fast



ML processing
- Image-blurring
- Classification on noisy images



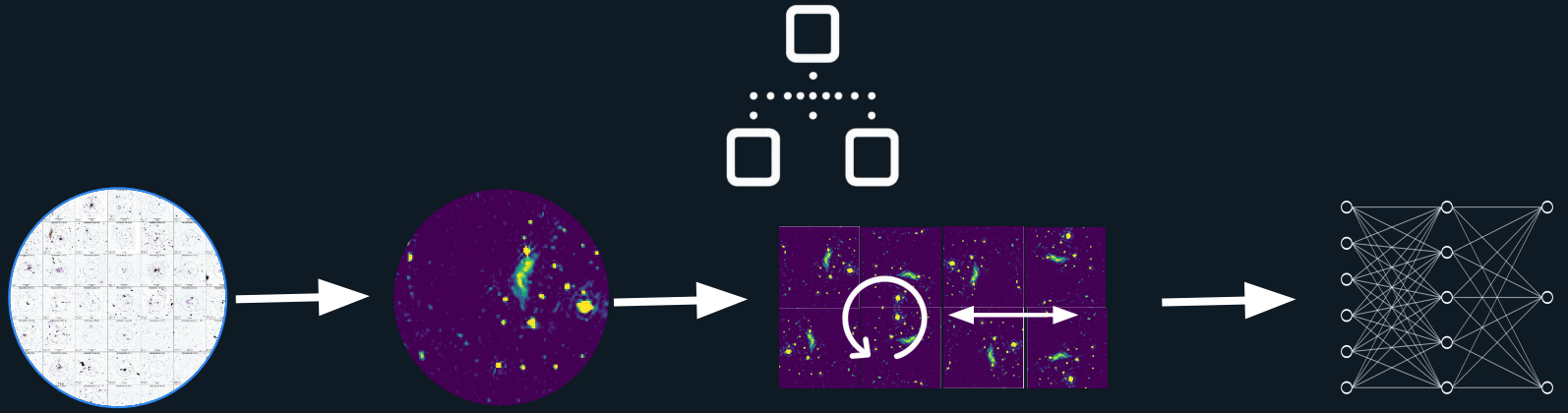
Statistical sample

Specific explanation

PSZ2 G122.30+54.52
- $M = 4.5 \times 10^{14} M_{\odot}$
- r_{500}=1.1 Mpc
- No diffuse emission

General explanation
- Mini-halos more common in CCC, through sloshing.
- Duty cycle?
- B-field(M, z)?

Machine learning processing



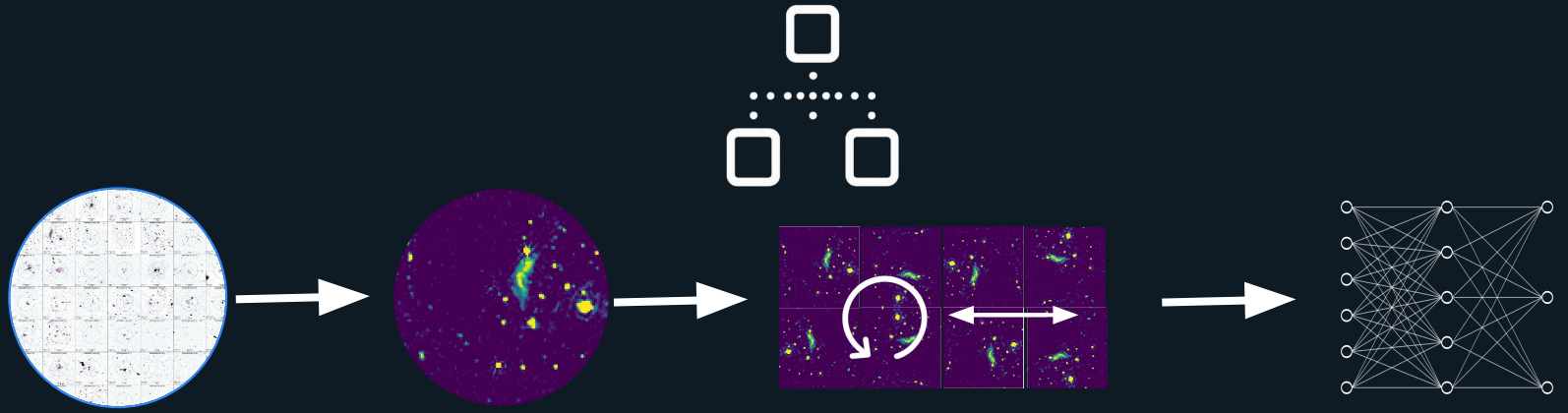
Acquired data

Preprocess
- Cropping
- Downsampling
- Normalisation
- Tapering

Classical augmentation
- Flipping
- Rotations

Classification
- Feature extraction
  - Scattering transform
  - Squeeze-excitation
- Feature correlations

Machine learning processing



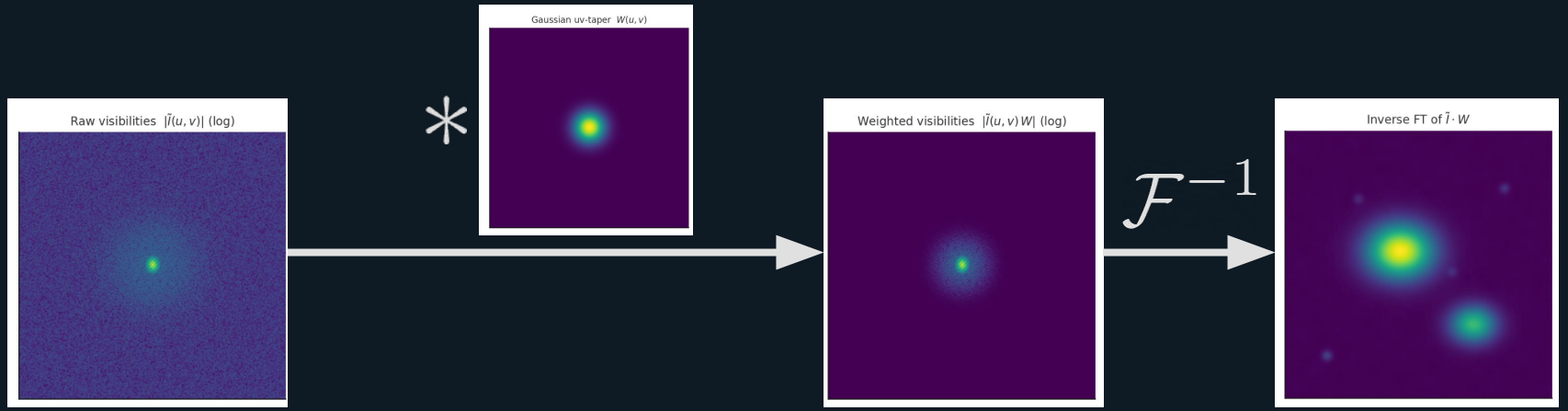
Acquired data

Preprocess
- Cropping
- Downsampling
- Normalisation
- Tapering

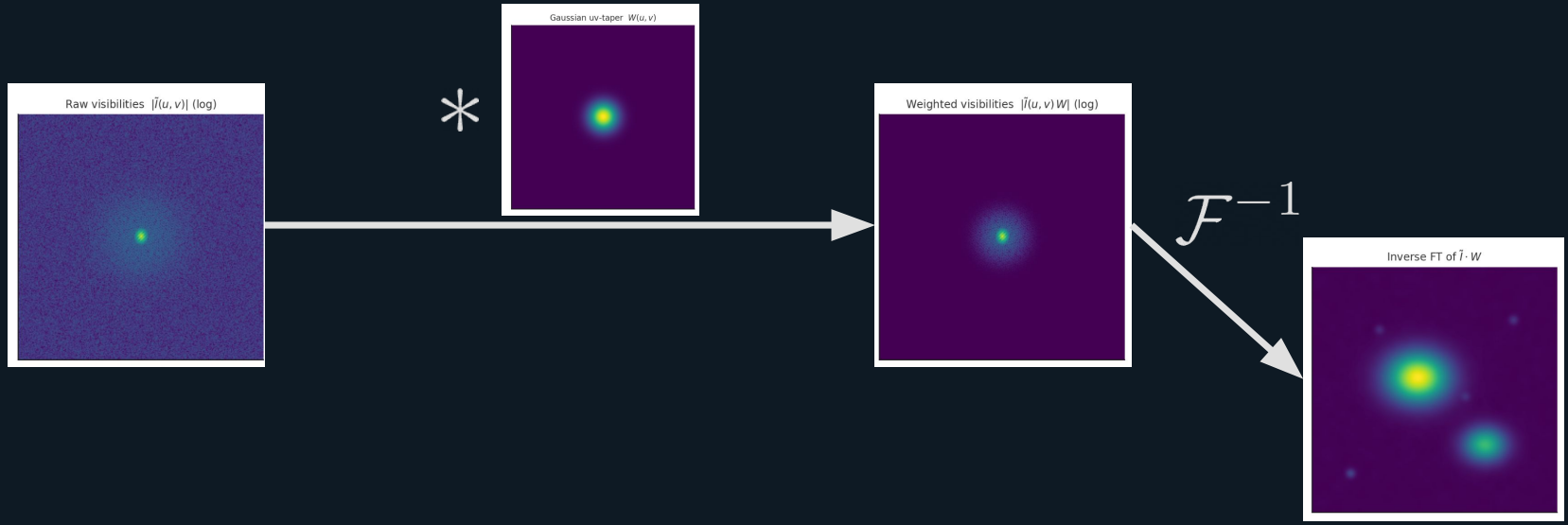
Classical augmentation
- Flipping
- Rotations

Classification
- Feature extraction
  - Scattering transform
  - Squeeze-excitation
- Feature correlations

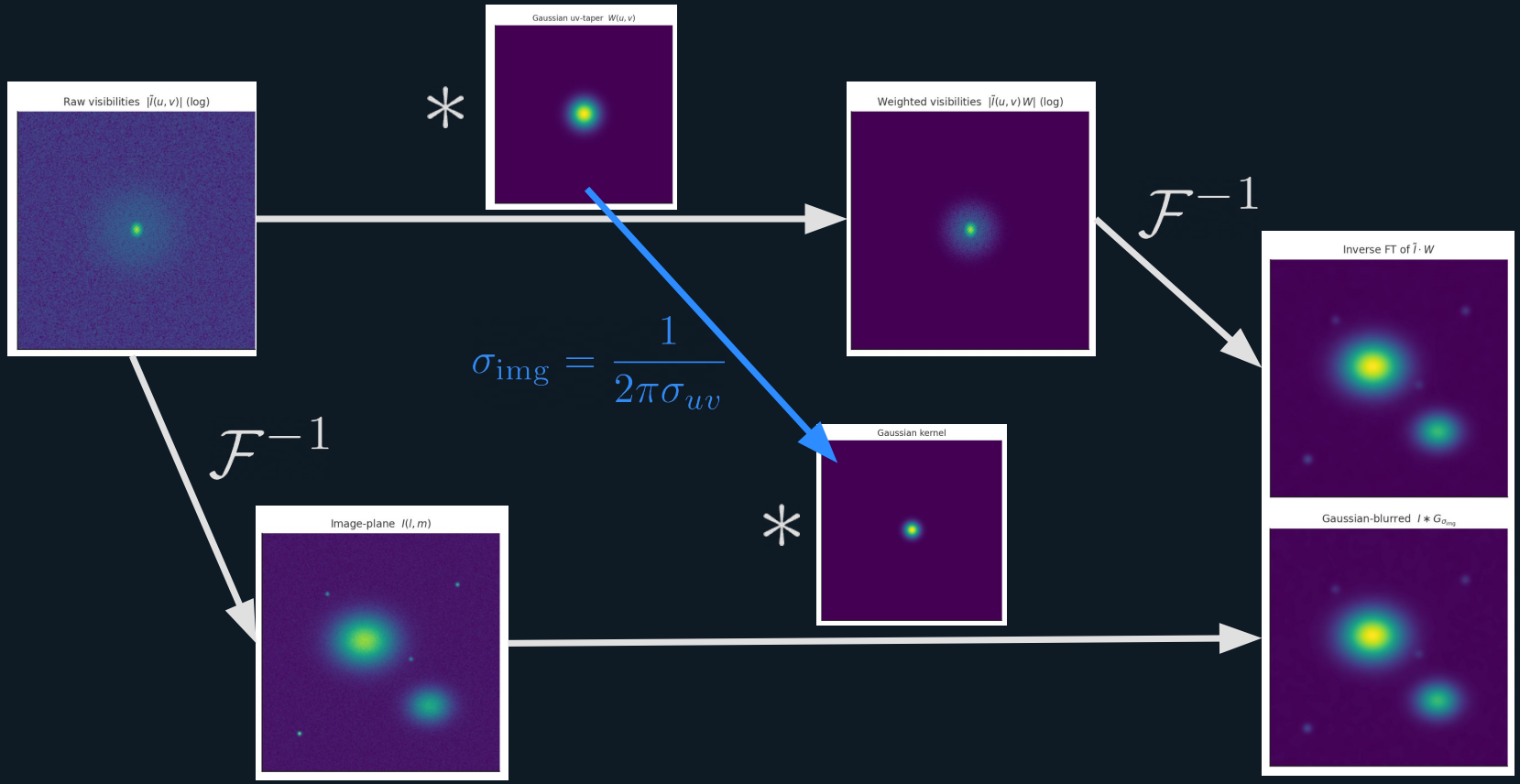
Tapering downweights longer baselines



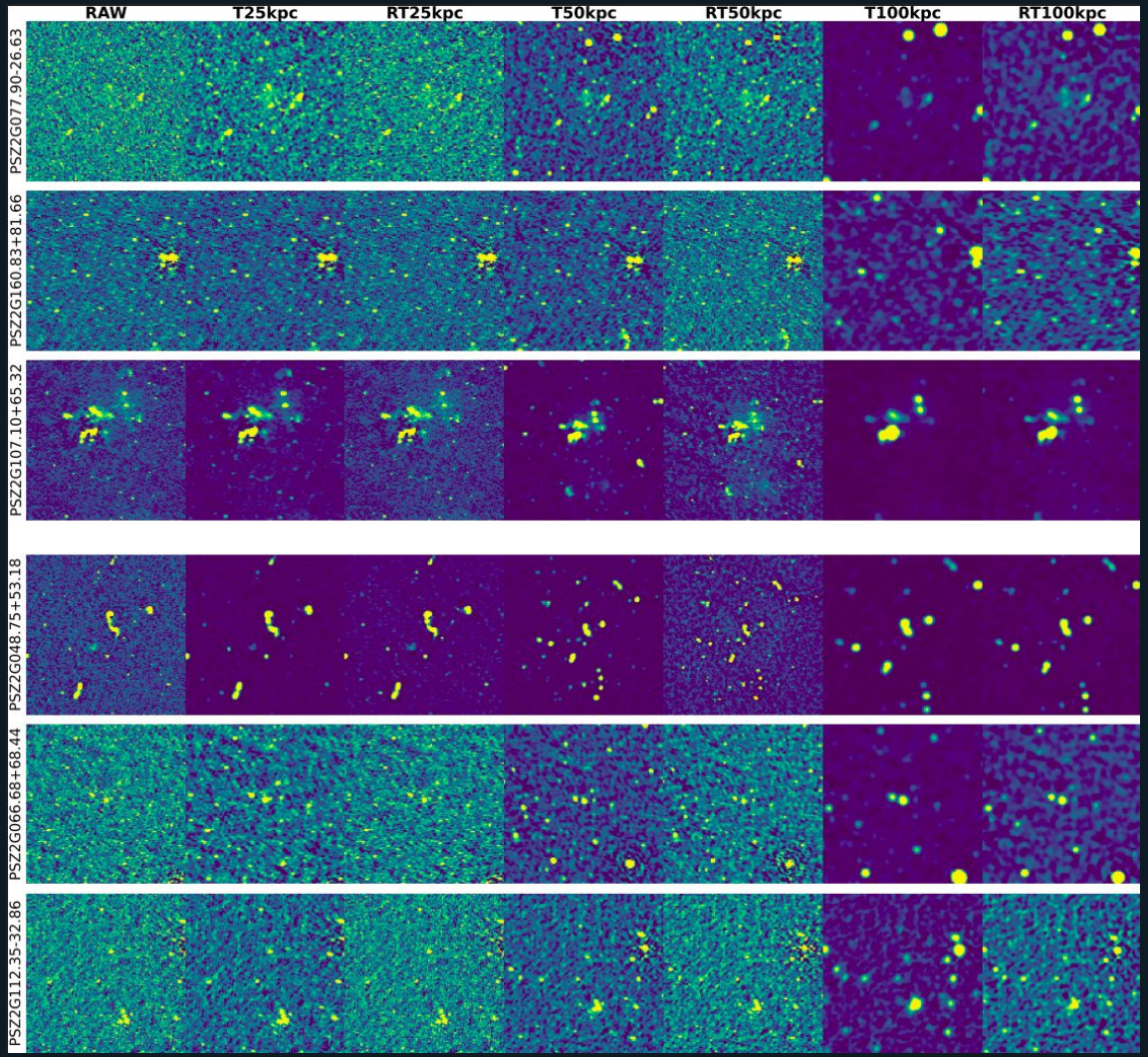
Tapering downweights smaller scales



Tapering downweights smaller scales

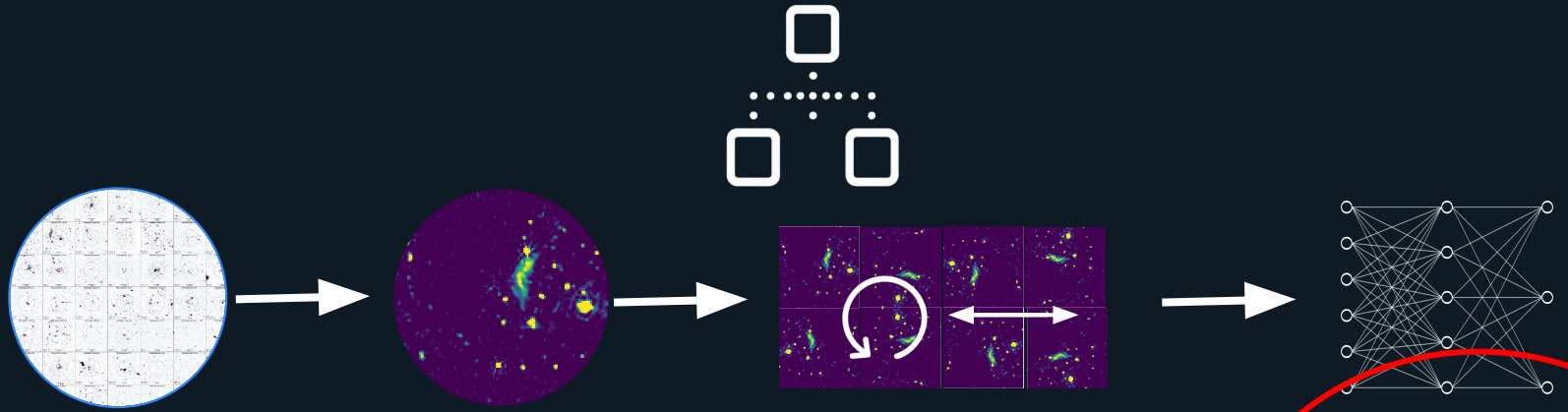


Blurring almost reproduces tapering



RT = blurring of image
T = tapering of visibilities

Machine learning processing



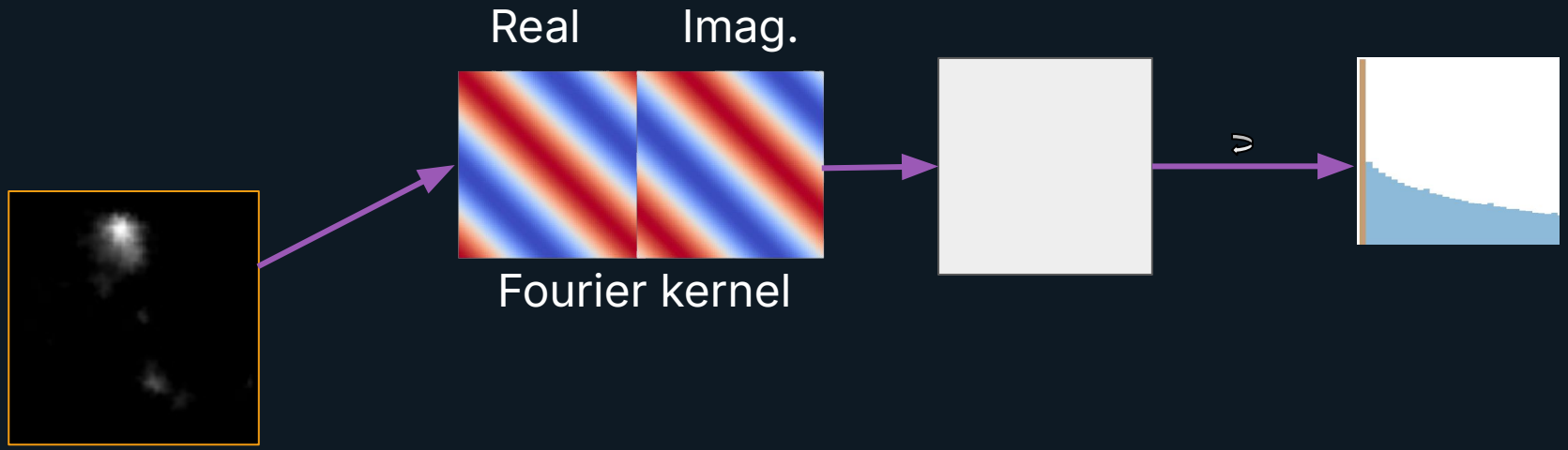
Acquired data

Preprocess
- Cropping
- Downsampling
- Normalisation
- Tapering

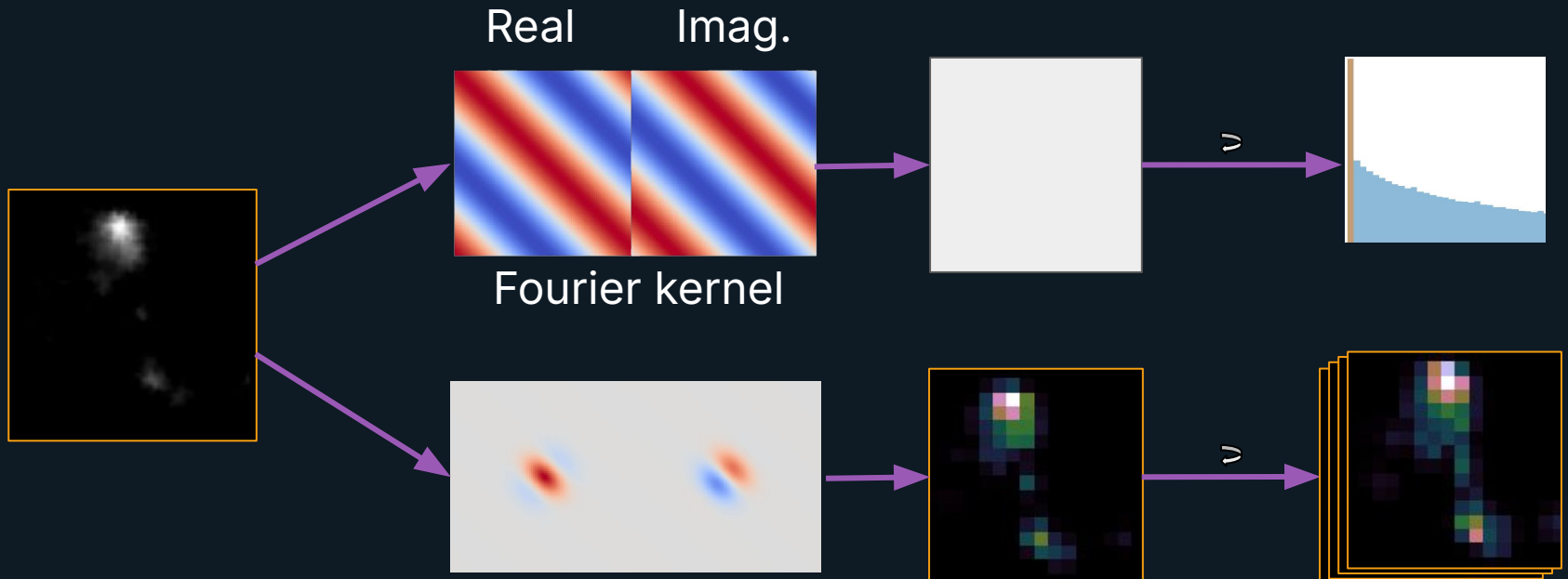
Classical augmentation
- Flipping
- Rotations

Classification
- Feature extraction
  - Scattering transform
  - Squeeze-excitation
- Feature correlations

Kernels are localised and extract spatial information

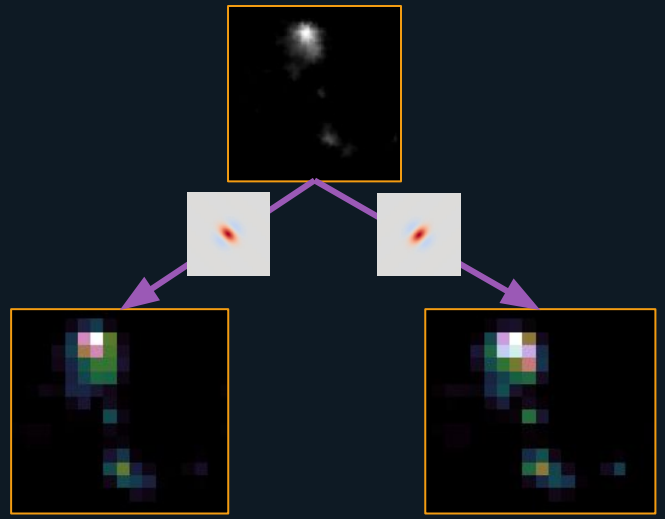


Kernels are localised and extract spatial information

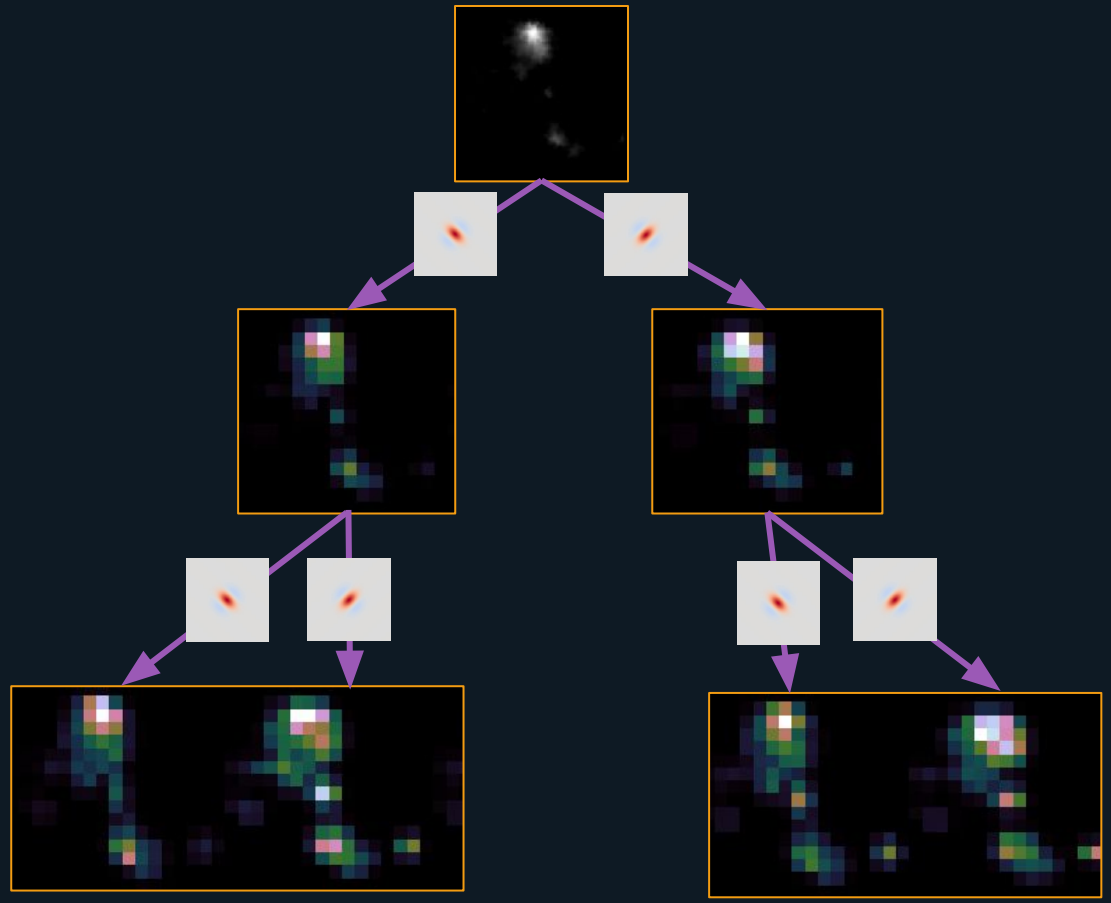


The scattering transform iteratively applies a wavelet transform

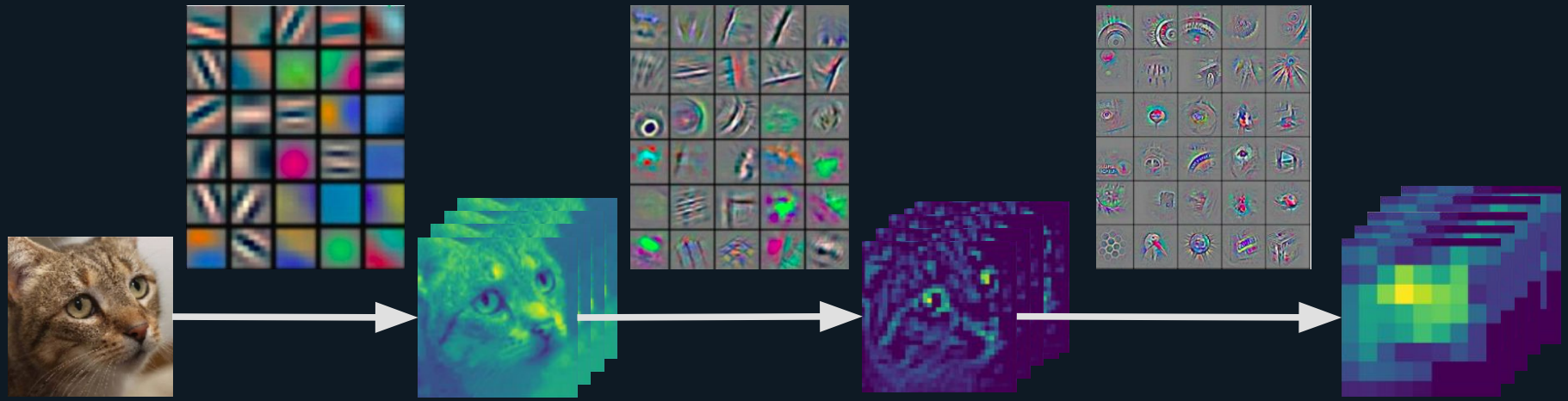
The scattering transform iteratively applies a wavelet transform



The scattering transform iteratively applies a wavelet transform

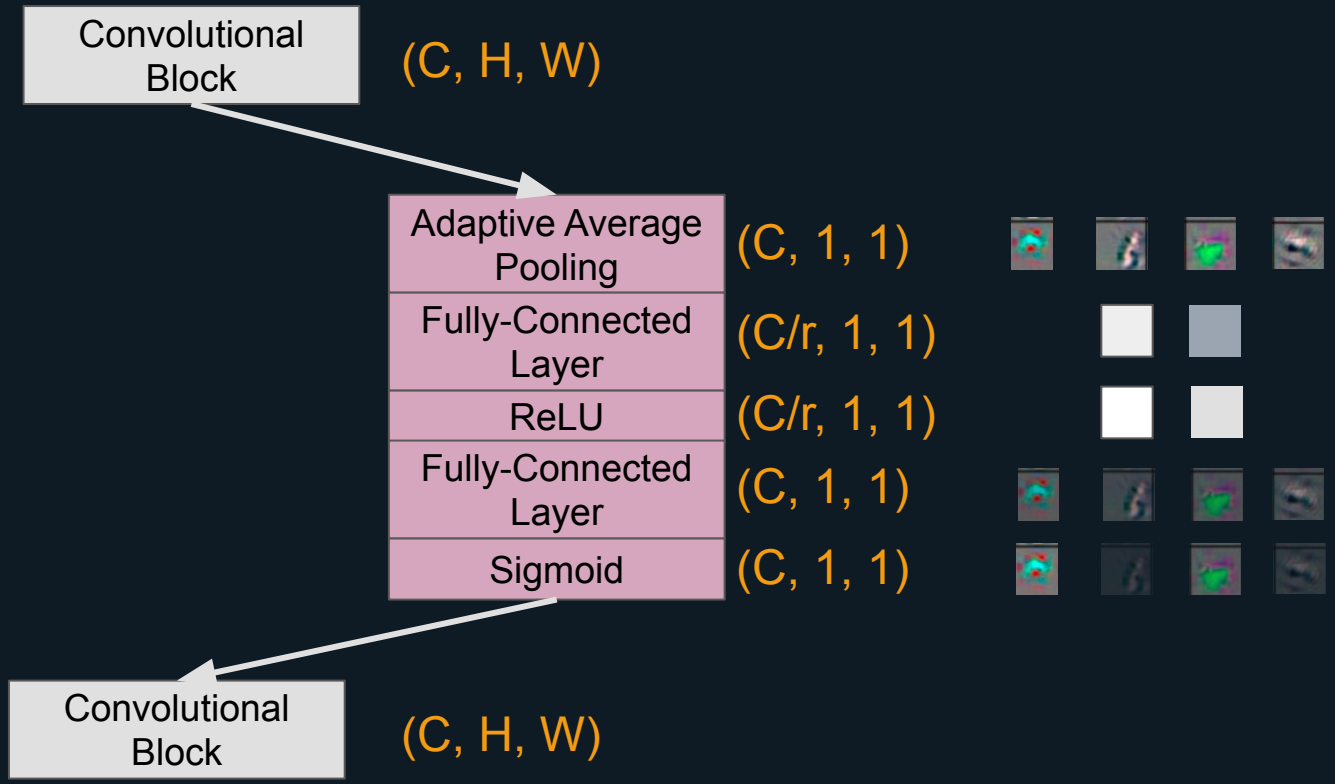


A convolutional neural network (CNN) extracts features

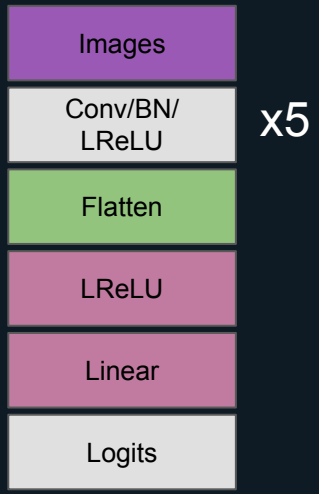


Credit: https://stackoverflow.com/questions/38450104/how-to-visualize-filters-after-the-1st-layer-trained-by-cnns [edited]

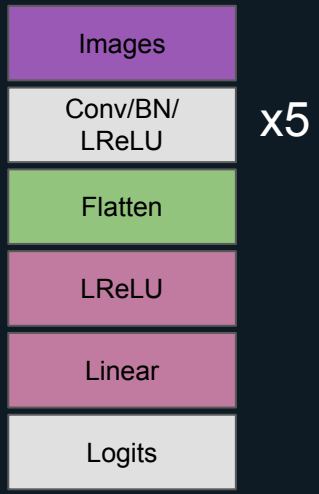
Squeeze-excitation attention weighs the channels



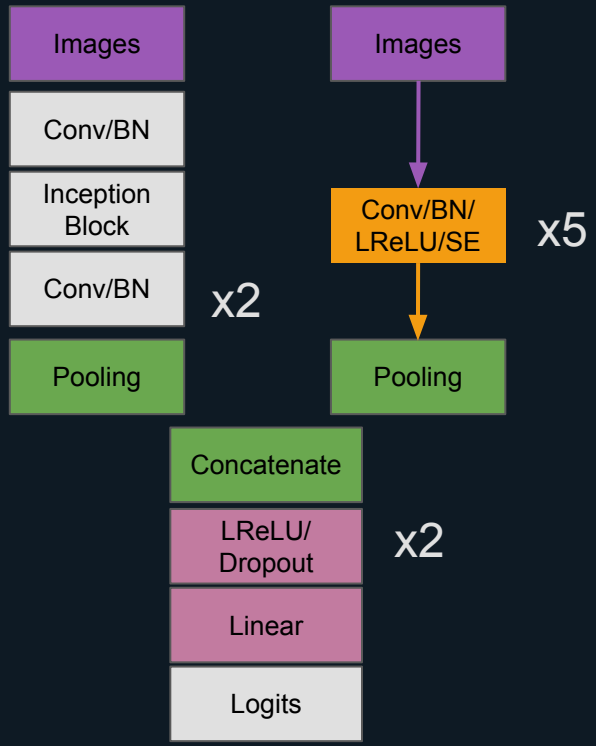
Convolutional neural network (CNN)



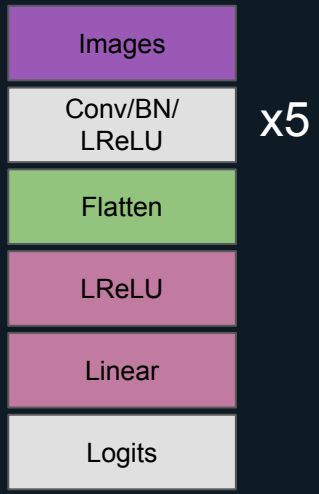
Convolutional neural network (CNN)



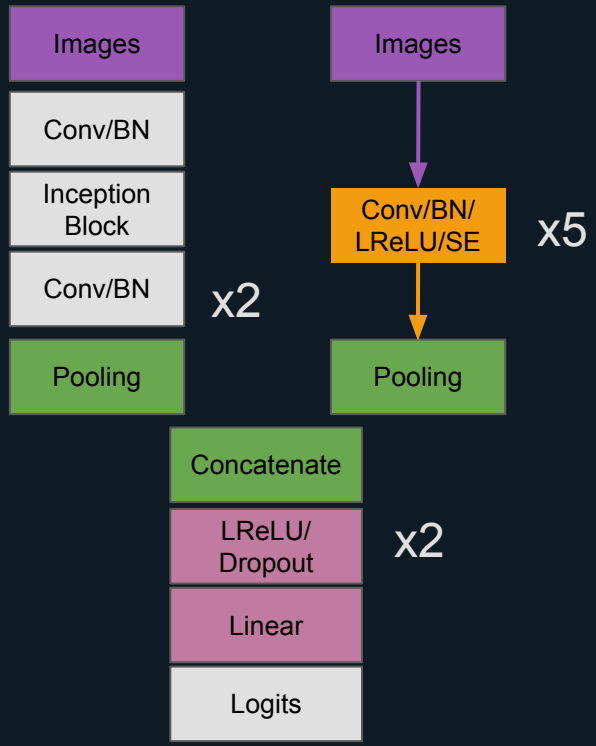
CNN-squeeze dual net (DualCSN)



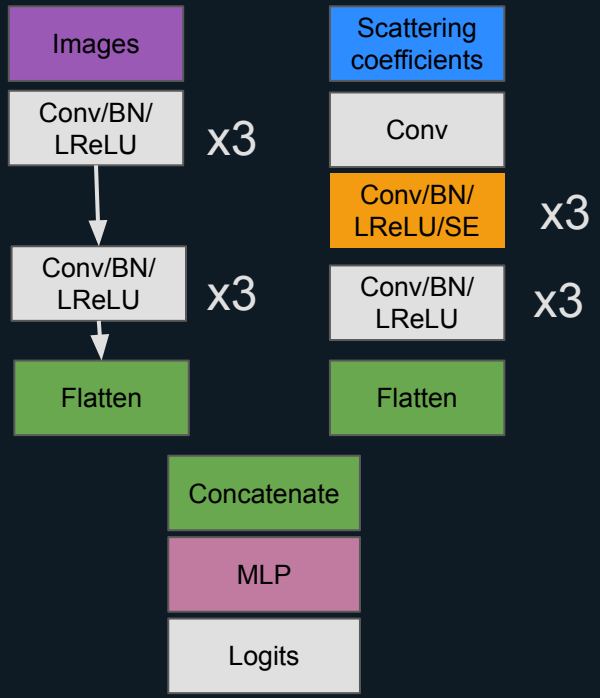
Convolutional neural network (CNN)



CNN-squeeze dual net (DualCSN)



Scatter-squeeze dual net (DualSSN)



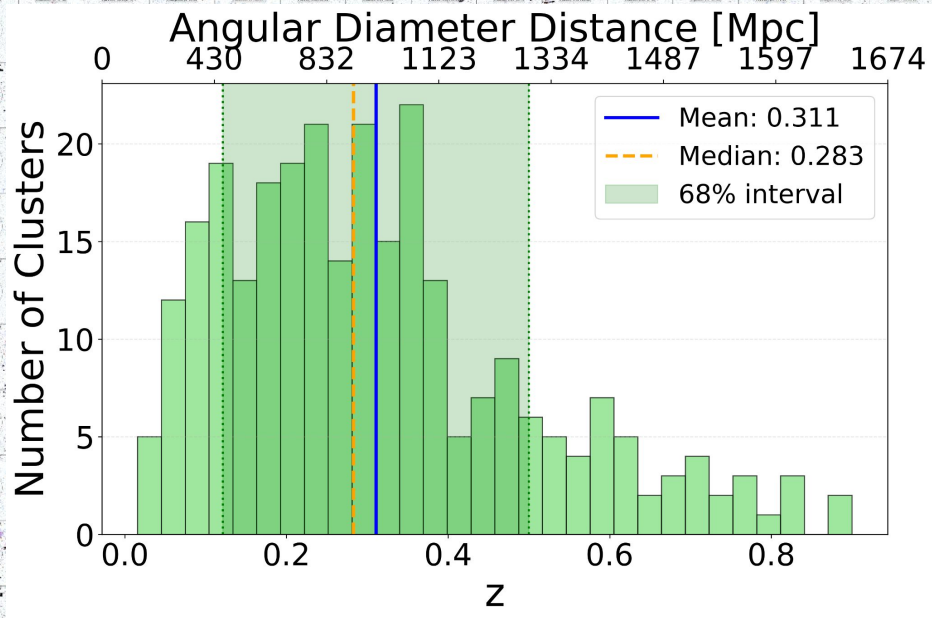
Classification of diffuse radio emission from LoTSSDR2

LOTSSDR2/PSZ2

LOFAR Two-metre Sky Survey Data Release (LoTSS-DR2) belonging to the second Planck catalogue of Sunyaev-Zel'dovich sources (PSZ2)

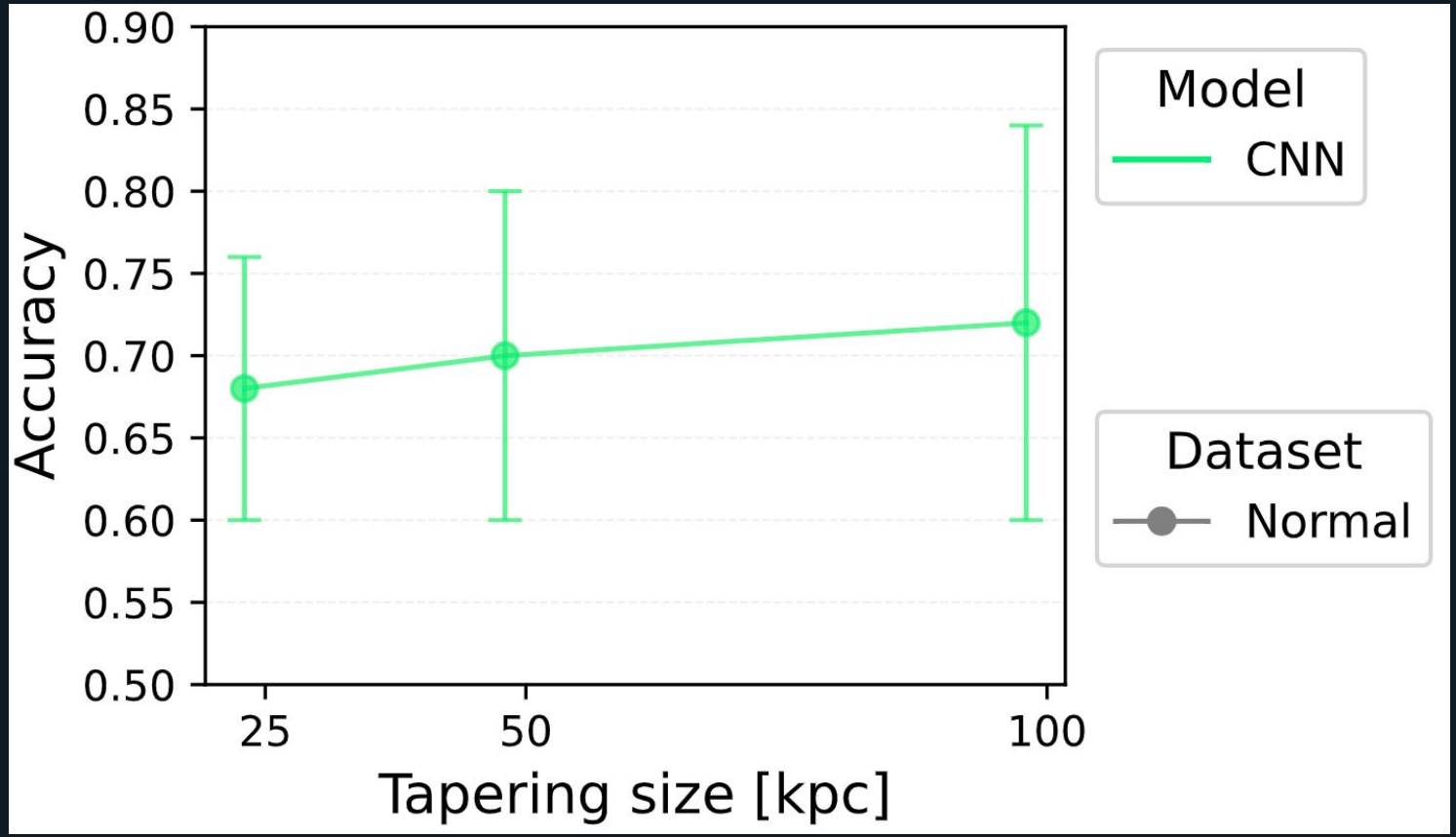
62 Diffuse Emission
114 No Diffuse Emission

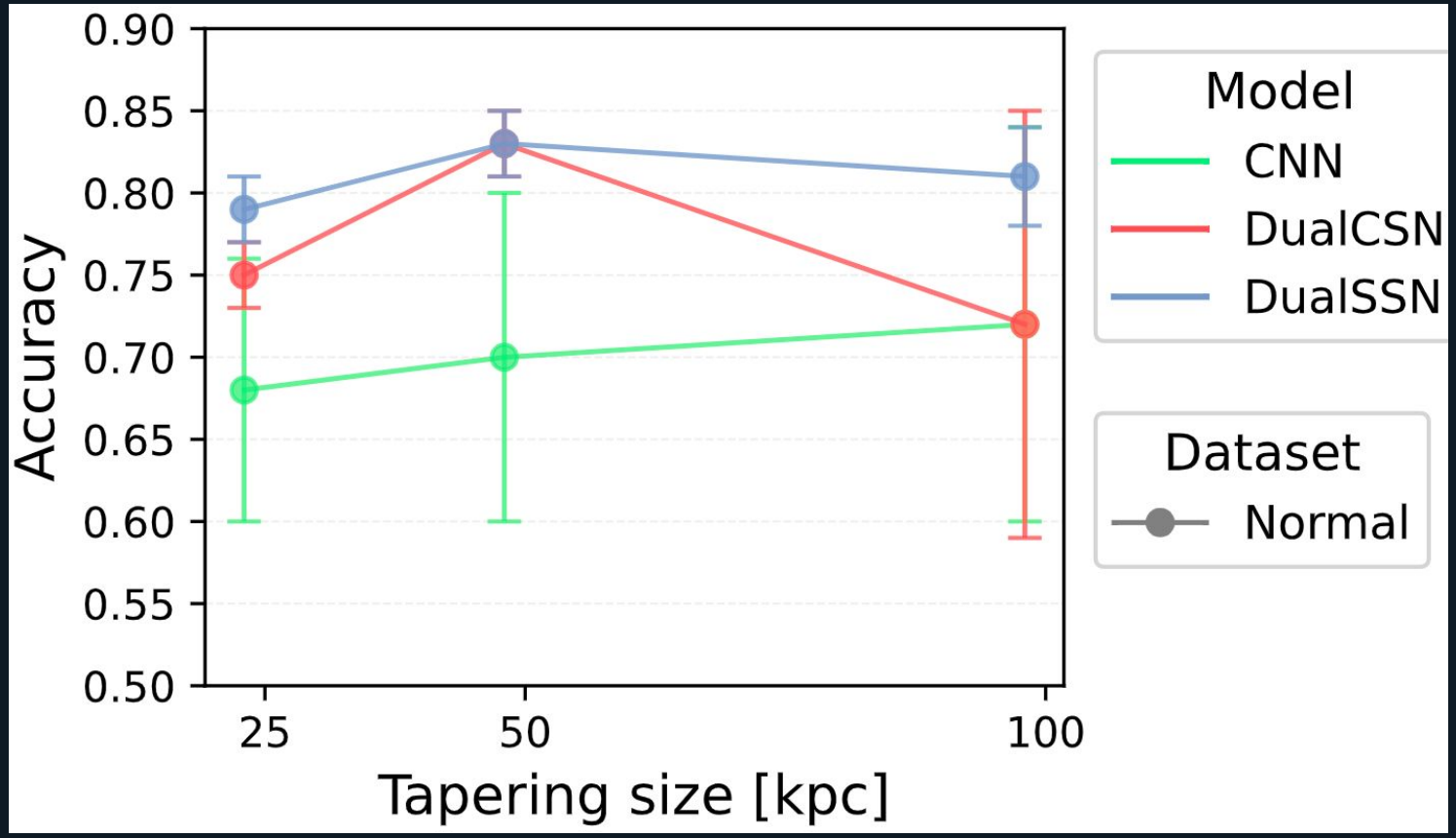
Processed to 1x128x128 pixels

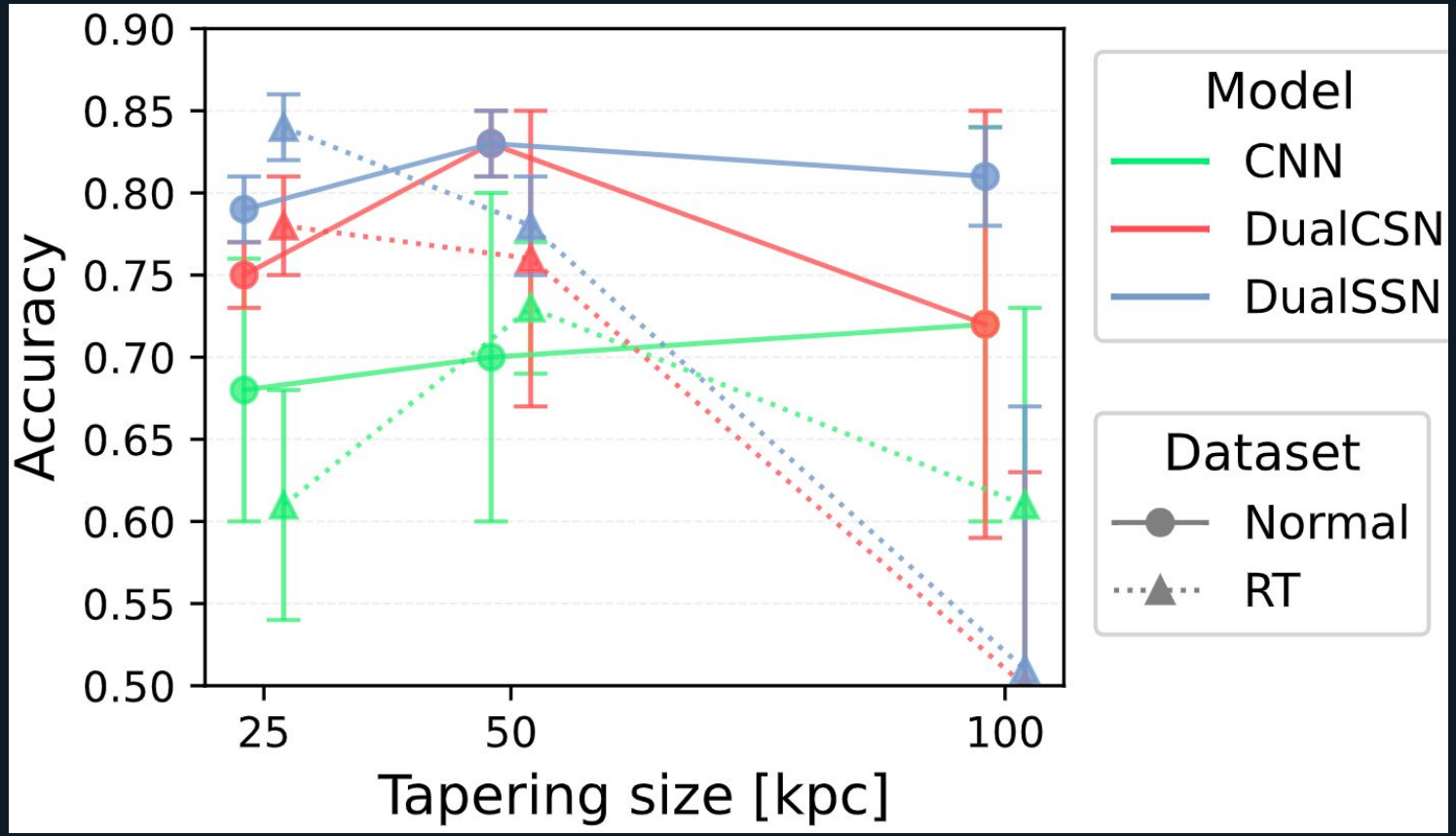


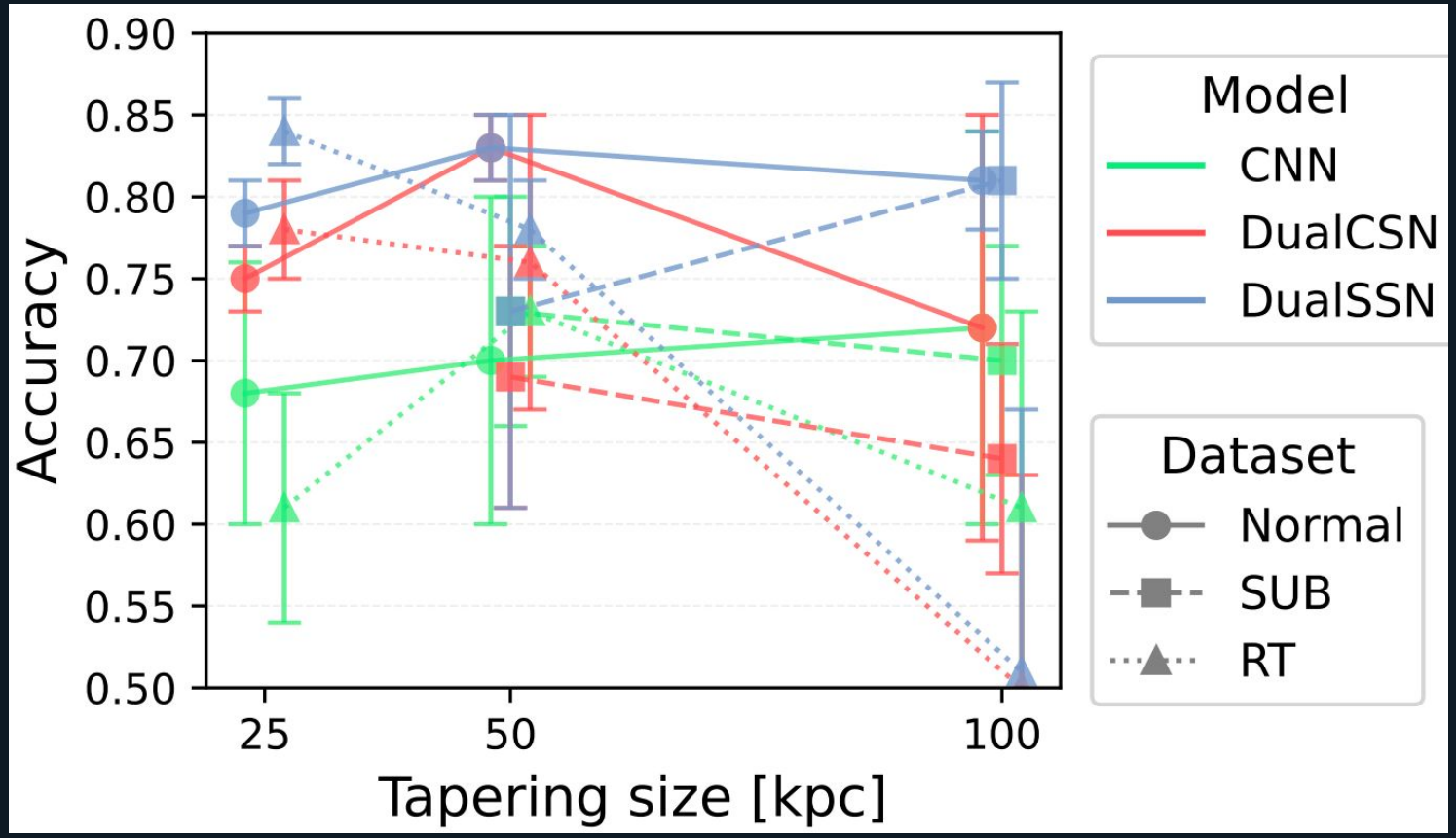
Data credit: Botteon et al. 2022, arXiv:2202.11720

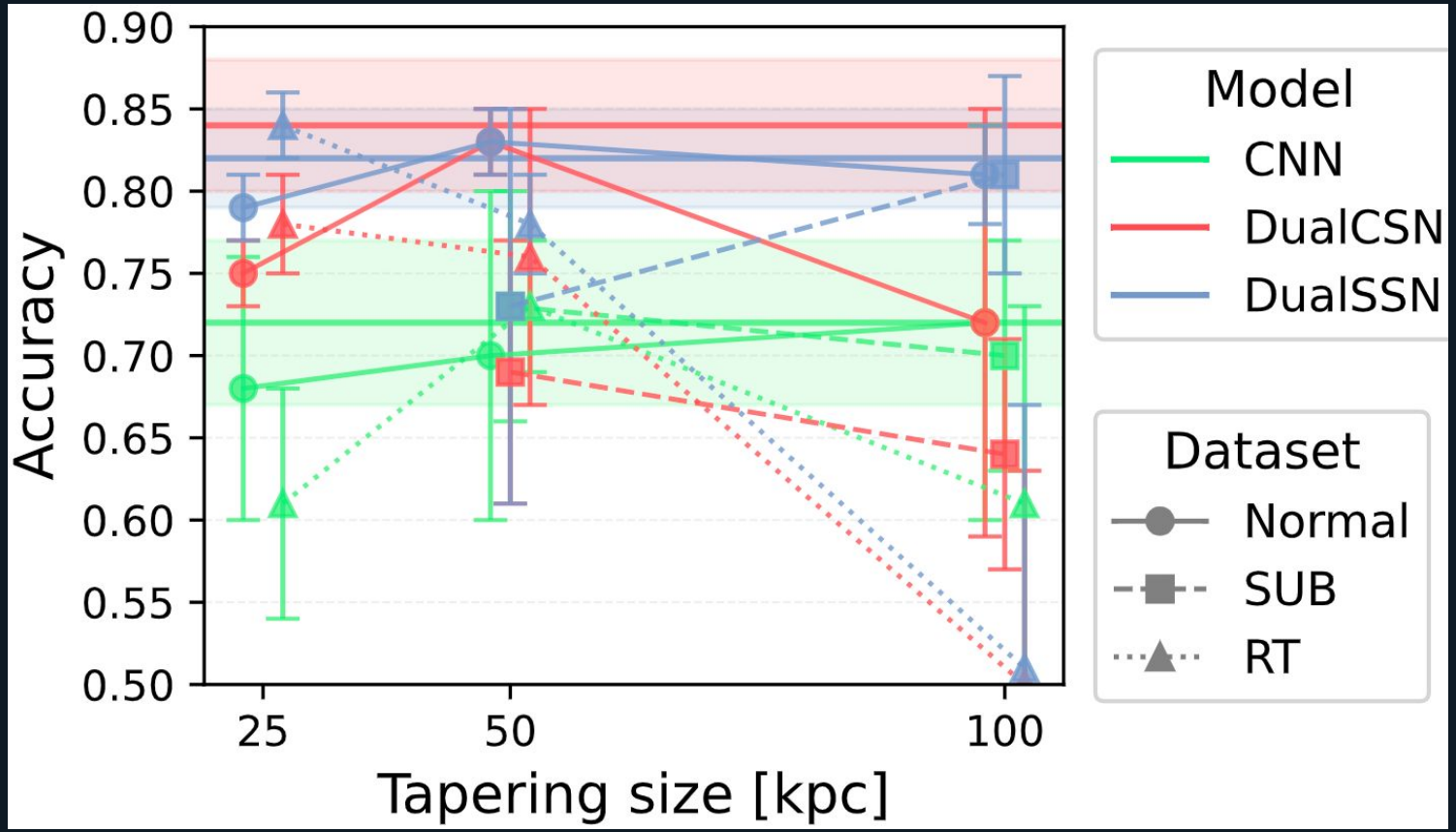
Classification of diffuse radio emission from LoTSSDR2

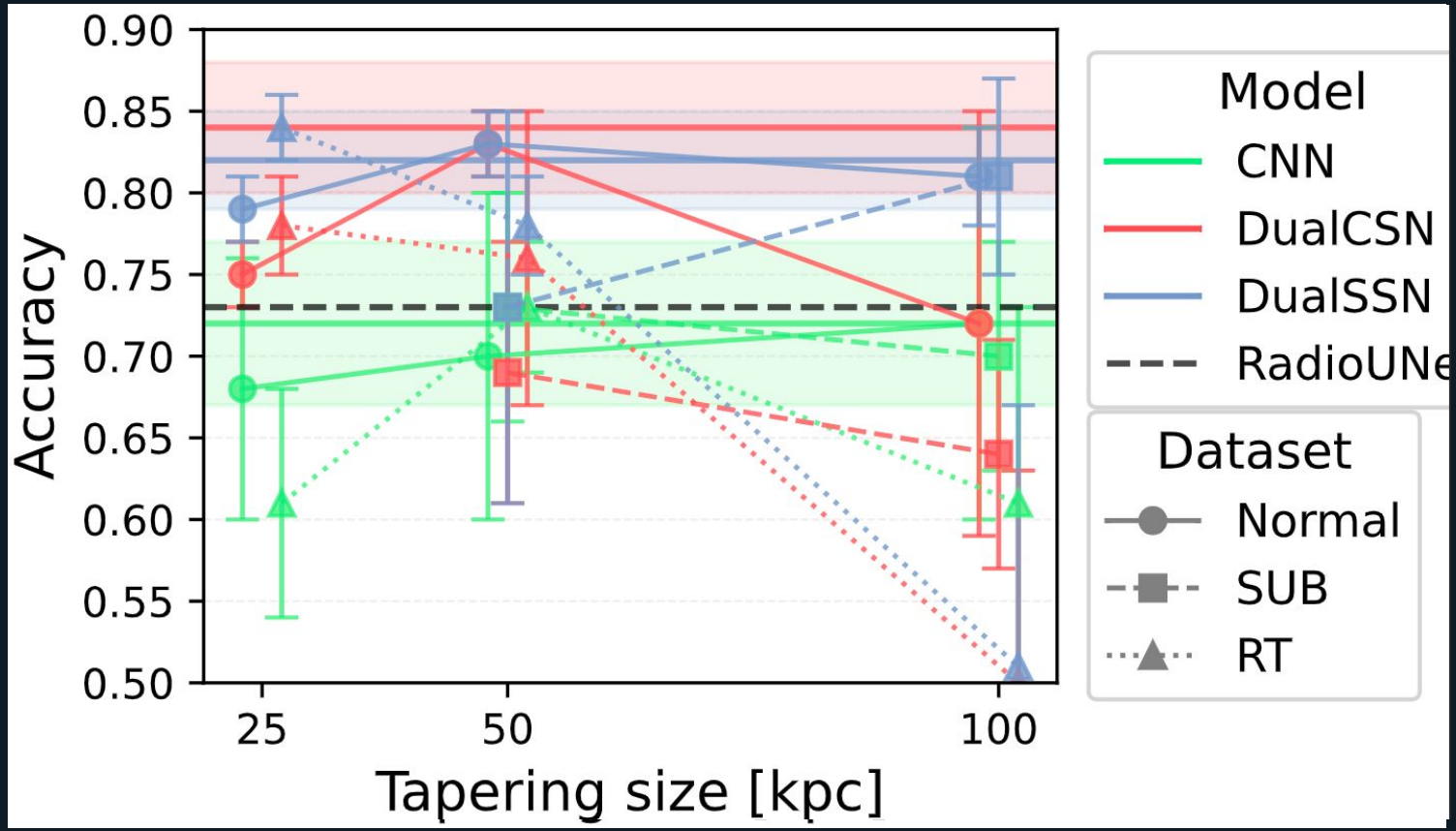

0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
Accuracy
25
50
100
Tapering size [kpc]
Model
CNN
Dataset
Normal


0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
Accuracy
25
50
100
Tapering size [kpc]
Model
CNN
DualCSN
DualSSN
Dataset
Normal


0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
Accuracy
25
50
100
Tapering size [kpc]
Model
CNN
DualCSN
DualSSN
Dataset
Normal
RT

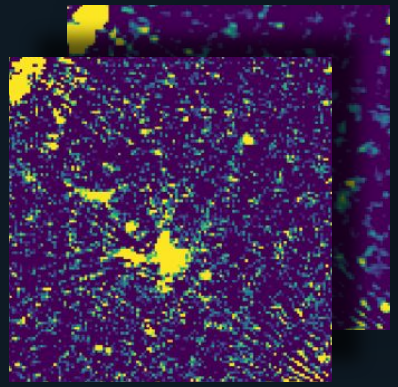

0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
Accuracy
25
50
100
Tapering size [kpc]
Model
CNN
DualCSN
DualSSN
Dataset
Normal
SUB
RT


0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
Accuracy
25
50
100
Tapering size [kpc]
Model
CNN
DualCSN
DualSSN
Dataset
Normal
SUB
RT

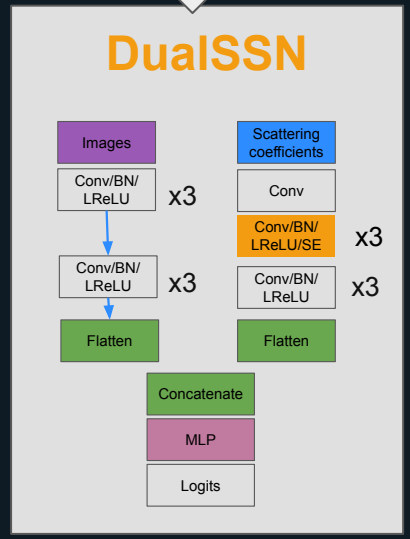

0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
Accuracy
25
50
100
Tapering size [kpc]
Model
CNN
DualCSN
DualSSN
RadioUNe
Dataset
Normal
SUB
RT

Stacking images could render more information salient

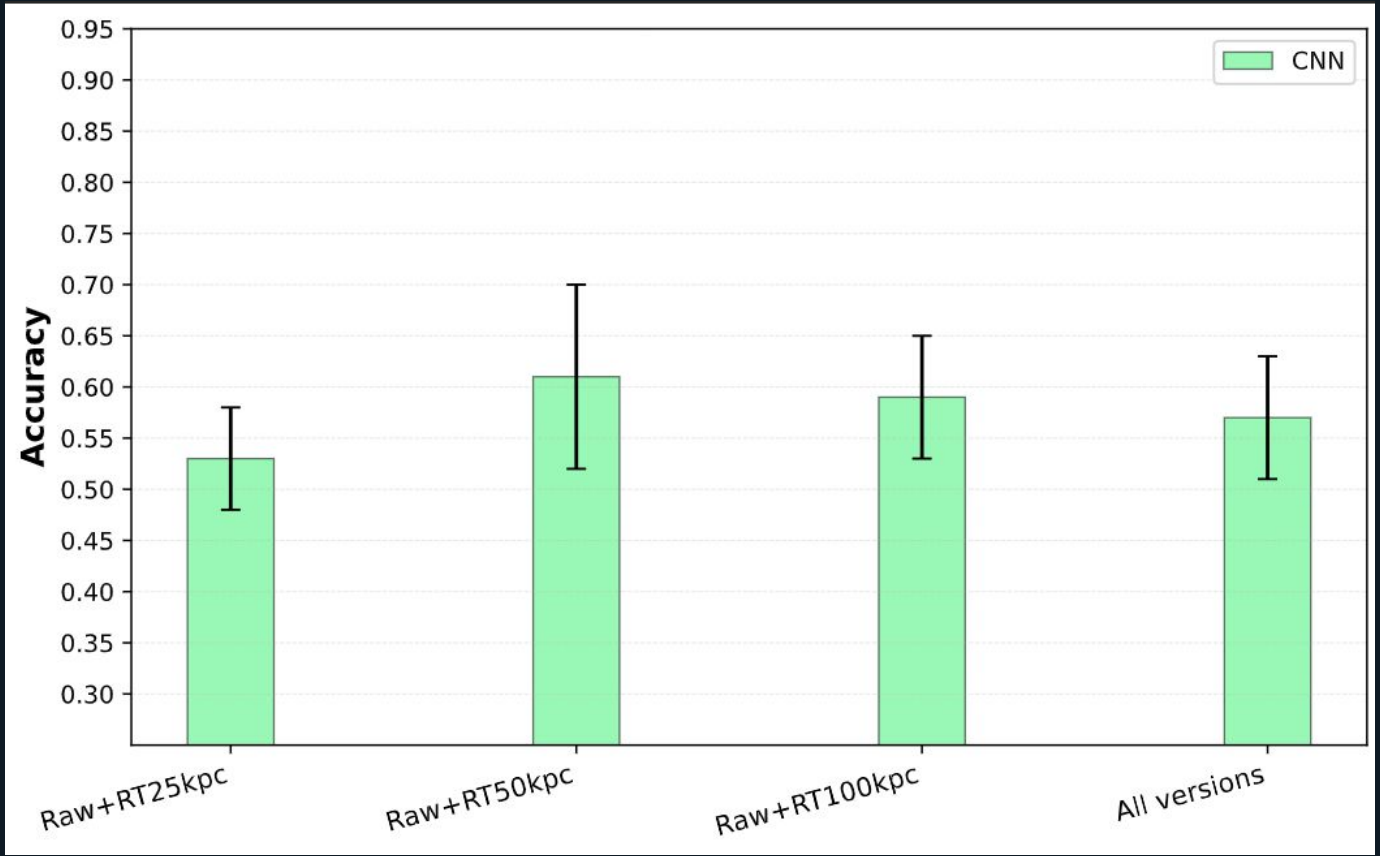
Raw image



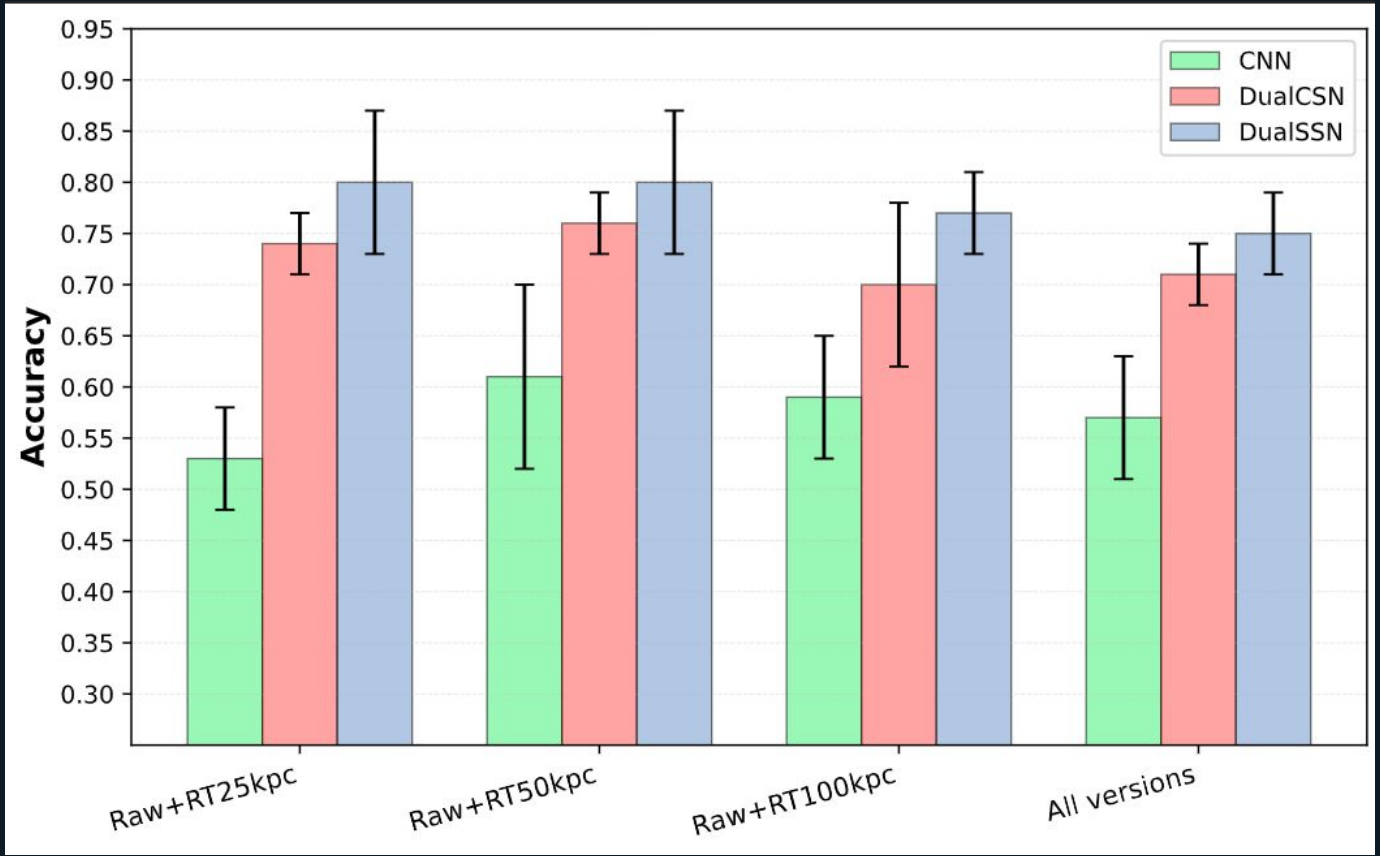
Tapered image



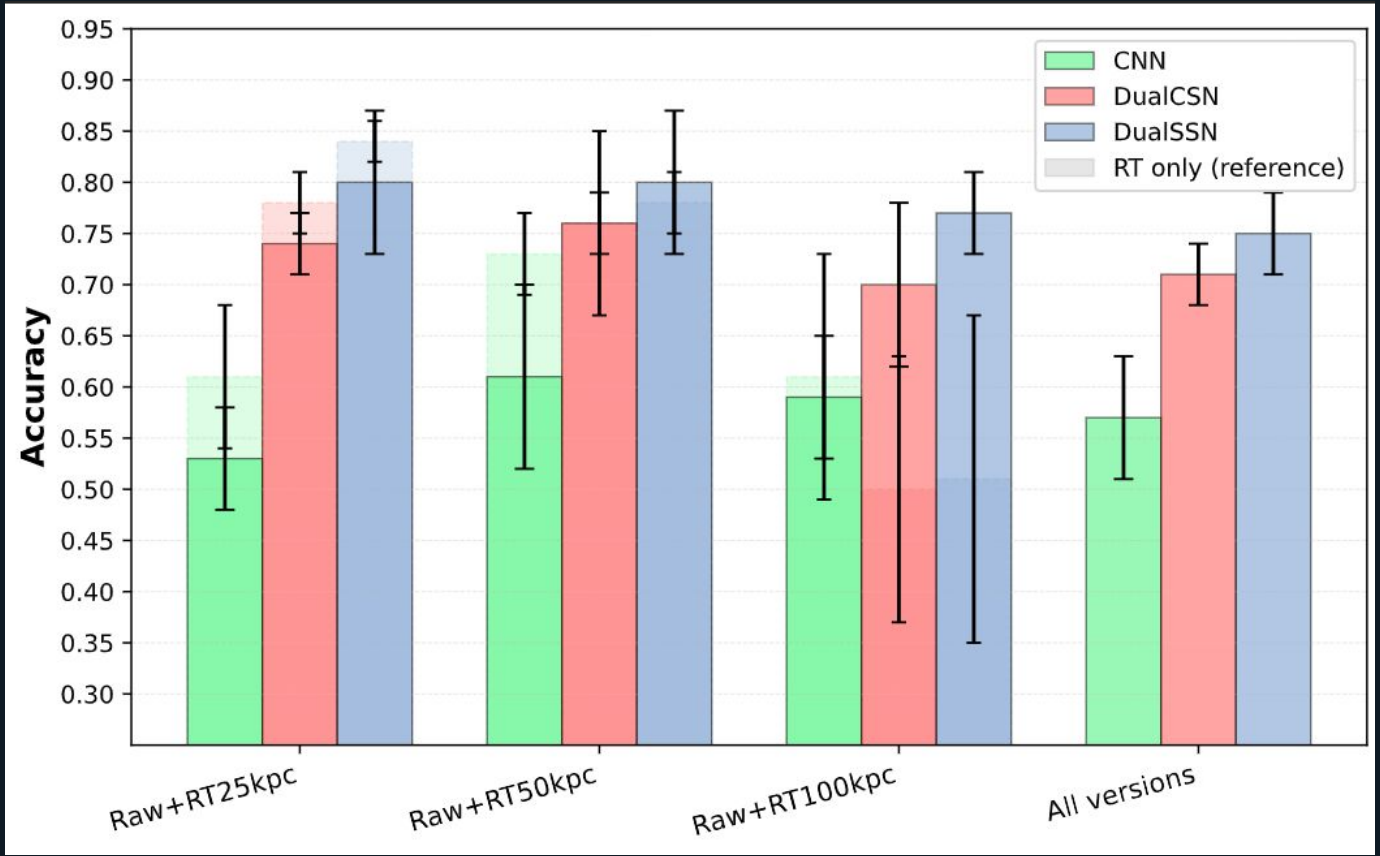
Stacking images could render more information salient



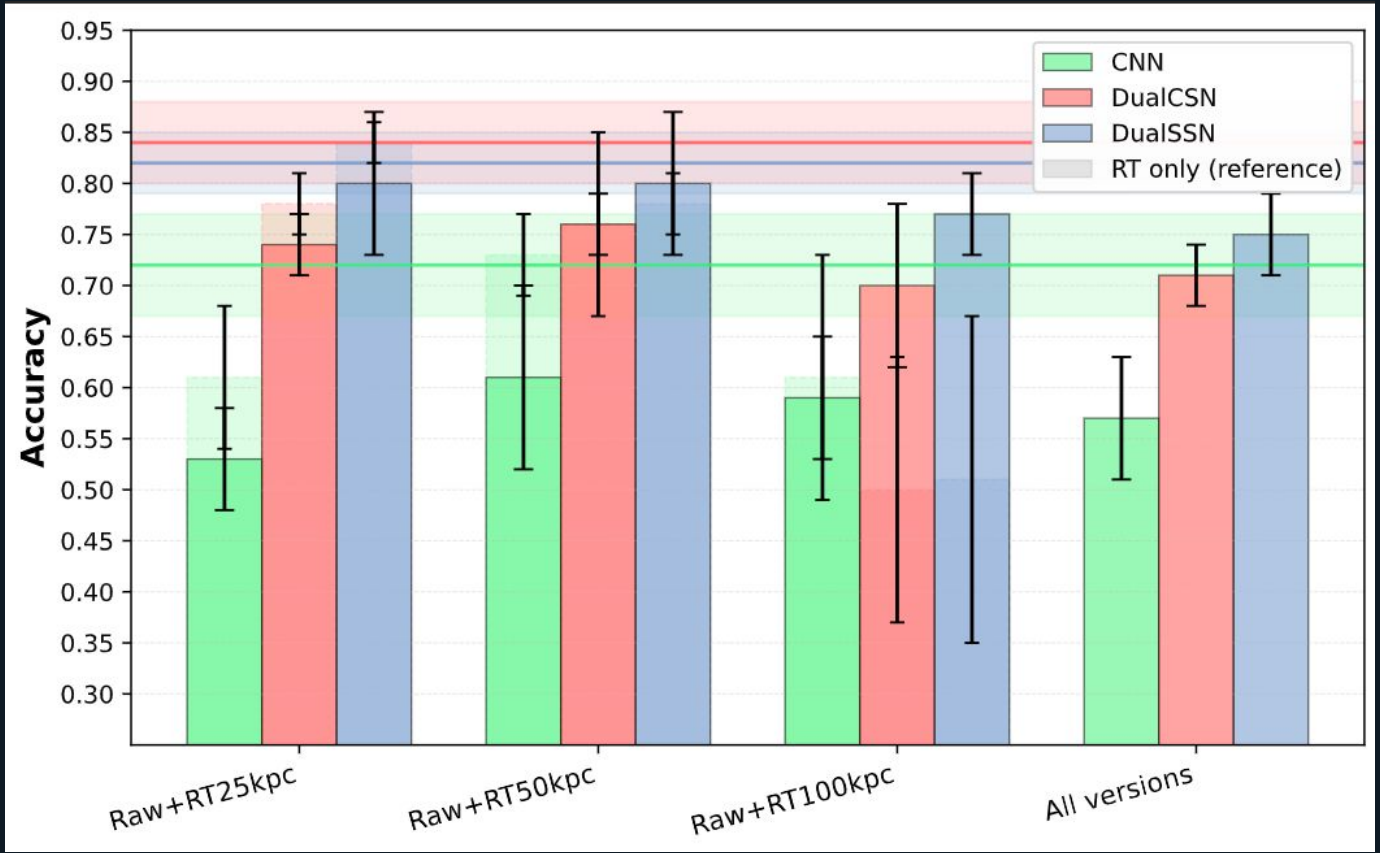
Stacking images could render more information salient



Stacking images could render more information salient



Stacking images could render more information salient



Conclusions and thank you

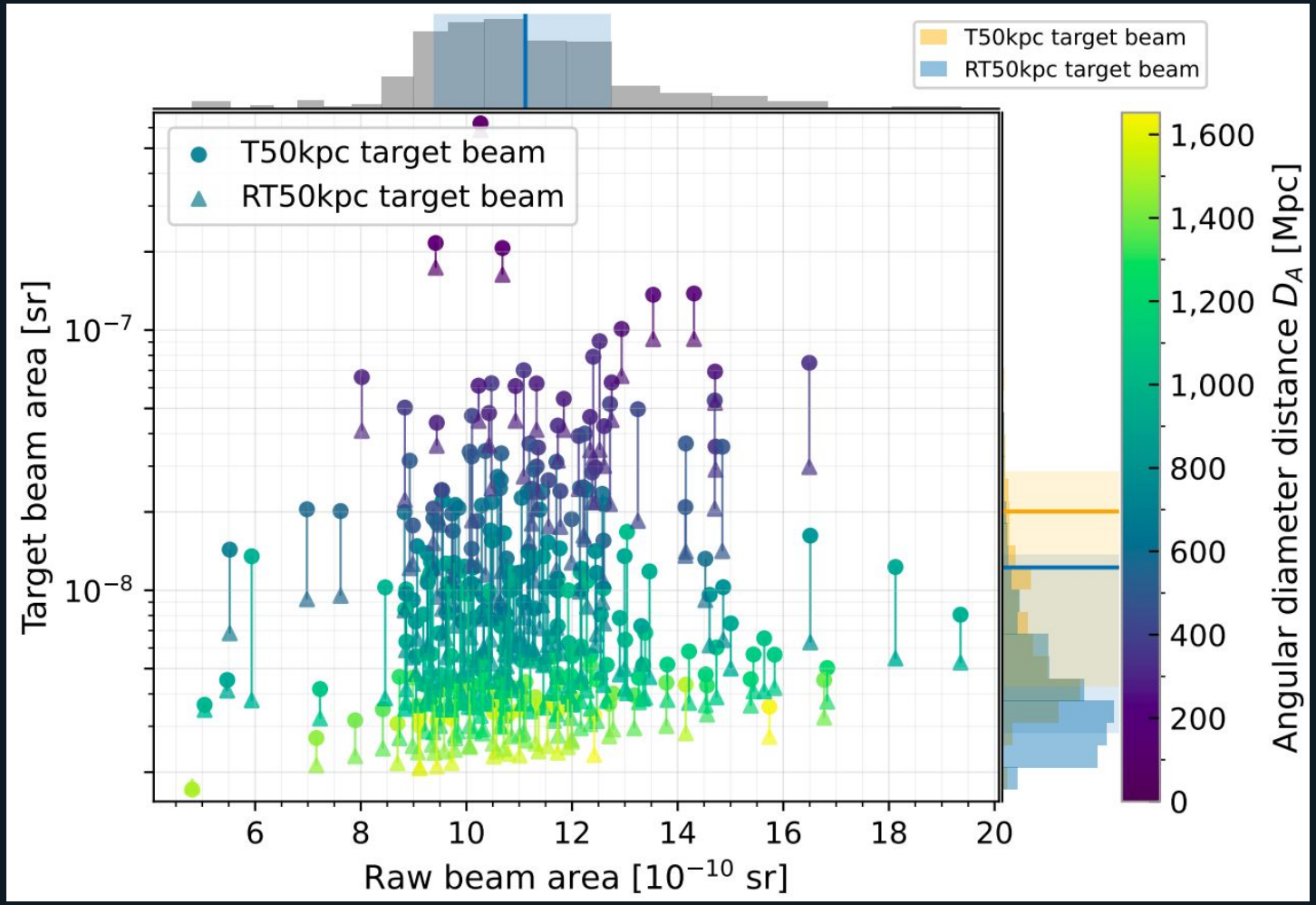
- The upcoming decennium can expect thousands of new discoveries of diffuse cluster radio emission
- Scattering transform, squeeze-excitation attention, and multi-branch architecture may all improve classification performance
- Tapering or blurring and stacking did not prove useful in this study

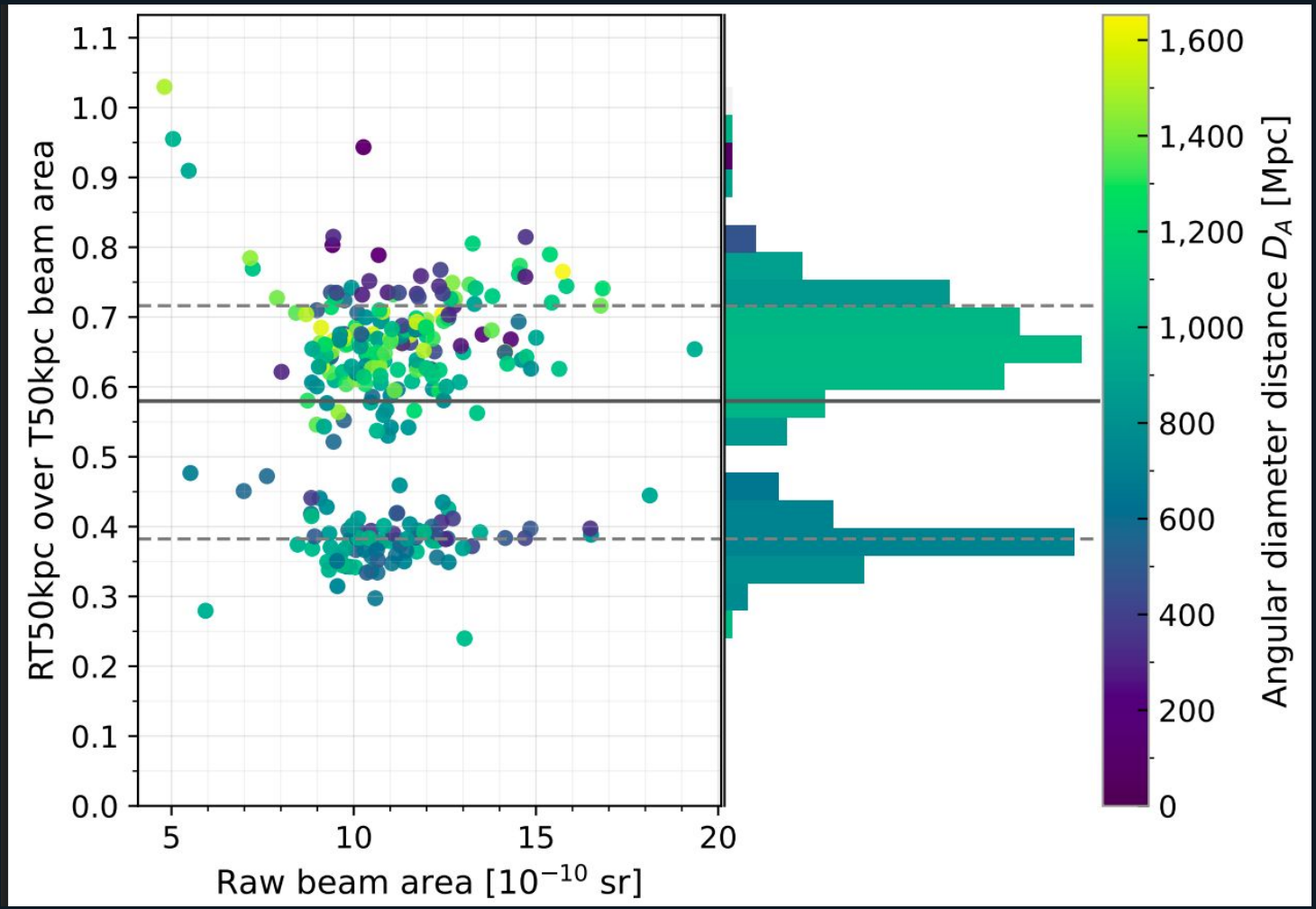
Thank you for your attention!

Bonus slides 1 - What more?

Number of sources per class and data split for the two processing strategies

Class	Split	Constant FoV	Constant n_{beams}
DE	Train	46-51	41-46
	Validation	3-8	3-8
	Test	13	15
	TOTAL	67	64
NDE	Train	98-103	82-87
	Validation	9-14	6-11
	Test	28	21
	TOTAL	140	114
Total	–	207	178


T50kpc target beam
RT50kpc target beam
T50kpc target beam
RT50kpc target beam
Target beam area [sr]
10^{-7}
10^{-8}
6
8
10
12
14
16
18
20
Raw beam area [10^{-10} sr]
1,600
1,400
1,200
1,000
800
600
400
200
0
Angular diameter distance D_A [Mpc]


RT50kpc over T50kpc beam area
Raw beam area [10^{-10} sr]
Angular diameter distance D_A [Mpc]
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
5
10
15
20
0
200
400
600
800
1,000
1,200
1,400
1,600

CNN classifier

Branch	Layer	Component	Depth	Activation	Regularizer	Parameters
Feature Extractor	1-2	5×5 Conv (s=2), BN	16	Leaky ReLU	Dropout2d (0.15)	448
	3-4	5×5 Conv (s=2), BN	32	Leaky ReLU	Dropout2d (0.20)	12,896
	5-6	5×5 Conv (s=2), BN	32	Leaky ReLU	Dropout2d (0.25)	25,696
	7-8	3×3 Conv (s=2), BN	48	Leaky ReLU	Dropout2d (0.30)	13,968
	9-10	3×3 Conv (s=2), BN	48	Leaky ReLU	Dropout2d (0.35)	20,880
	11-12	3×3 Conv (s=2), BN	64	Leaky ReLU	Dropout2d (0.40)	27,840
	13	AdaptiveAvgPool2d	64	-	-	0
Classifier	1-2	Flatten, Linear, BN	32	Leaky ReLU	Dropout (0.5)	2,144
	3	Linear	2	-	-	66
					Total:	103,938

DualCSN classifier

Branch	Layer	Component	Depth	Activation	Regularizer	Parameters
Image Encoder 1	1-2	5×5 Conv, BN	8	Leaky ReLU	Dropout2d (0.3)	208
	3-4	3×3 Conv, BN	16	Leaky ReLU	Dropout2d (0.3)	1,168
	5	2×2 MaxPool	16	-	-	0
	6-7	5×5 Conv, BN	32	Leaky ReLU	Dropout2d (0.4)	12,832
	8-9	3×3 Conv (s=2), BN	32	Leaky ReLU	Dropout2d (0.4)	9,248
	10-11	3×3 Conv (s=2), BN	32	Leaky ReLU	Dropout2d (0.4)	9,248
	12-15	2× [3×3 Conv (s=2), BN]	32	Leaky ReLU	Dropout2d (0.4, 0.5)	18,496
Image Encoder 2	1-2	3×3 Conv, BN, SE	32	Leaky ReLU	Dropout2d (0.3)	448
	3-4	3×3 Conv, BN, SE	32	Leaky ReLU	Dropout2d (0.3)	9,440
	5-6	3×3 Conv (s=2), BN, SE	32	Leaky ReLU	Dropout2d (0.3)	9,440
	7-8	3×3 Conv, BN, SE	32	Leaky ReLU	Dropout2d (0.4)	9,440
	9-10	3×3 Conv (s=2), BN, SE	32	Leaky ReLU	Dropout2d (0.4)	9,440
	11-12	3×3 Conv, BN, SE	32	Leaky ReLU	Dropout2d (0.5)	9,440
	13-14	3×3 Conv (s=2), BN, SE	32	Leaky ReLU	Dropout2d (0.5)	9,440
Classifier	1-2	Linear, BN	32	Leaky ReLU	Dropout (0.5)	278,624
	3-4	Linear, BN	32	Leaky ReLU	Dropout (0.5)	1,120
	5	Linear	2	-	-	66
					Total:	388,530

DualSSN classifier

Branch	Layer	Component	Depth	Activation	Regularizer	Parameters
Image Encoder	1-2	5×5 Conv, BN	8	Leaky ReLU	Dropout2d (0.2)	208
	3-4	3×3 Conv, BN	16	Leaky ReLU	Dropout2d (0.2)	1,168
	5	2×2 MaxPool	16	-	-	0
	6-7	5×5 Conv, BN	32	Leaky ReLU	Dropout2d (0.3)	12,832
	8-9	3×3 Conv (s=2), BN	32	Leaky ReLU	Dropout2d (0.3)	9,248
	10-11	3×3 Conv (s=2), BN	32	Leaky ReLU	Dropout2d (0.3)	9,248
	12-15	2× [3×3 Conv (s=2), BN]	32	Leaky ReLU	Dropout2d (0.3, 0.4)	18,496
Scattering Encoder	1-2	3×3 Conv, BN, SE	16	Leaky ReLU	Dropout2d (0.2)	24,384
	3-4	3×3 Conv, BN, SE	16	Leaky ReLU	Dropout2d (0.2)	2,352
	5-6	3×3 Conv (s=2), BN, SE	16	Leaky ReLU	Dropout2d (0.2)	2,352
	7-8	3×3 Conv, BN, SE	16	Leaky ReLU	Dropout2d (0.3)	2,352
	9-10	3×3 Conv (s=2), BN, SE	16	Leaky ReLU	Dropout2d (0.3)	2,352
	11-12	3×3 Conv, BN, SE	16	Leaky ReLU	Dropout2d (0.4)	2,352
	13-14	3×3 Conv (s=2), BN, SE	16	Leaky ReLU	Dropout2d (0.4)	2,352
Classifier	1-2	Linear, BN	32	Leaky ReLU	Dropout (0.5)	24,672
	3-4	Linear, BN	32	Leaky ReLU	Dropout (0.5)	1,120
	5	Linear	2	-	-	66
					Total:	116,146

Table B1. Classification performance across datasets of CNN

Dataset	Accuracy	Precision	Recall	F1-score
RT25kpc	0.61 ± 0.07	0.54 ± 0.19	0.59 ± 0.23	0.55 ± 0.19
RT50kpc	**0.73 ± 0.04**	0.72 ± 0.08	0.61 ± 0.07	0.65 ± 0.03
RT100kpc	0.61 ± 0.12	0.56 ± 0.13	0.79 ± 0.14	0.63 ± 0.08
T25kpc	0.68 ± 0.08	0.60 ± 0.10	0.67 ± 0.12	0.62 ± 0.07
T50kpc	0.70 ± 0.10	**0.74 ± 0.14**	0.57 ± 0.15	0.61 ± 0.03
T100kpc	0.72 ± 0.12	0.67 ± 0.13	0.74 ± 0.13	0.68 ± 0.05
T50kpcSUB	0.73 ± 0.07	0.65 ± 0.08	**0.80 ± 0.25**	**0.70 ± 0.20**
T100kpcSUB	0.70 ± 0.07	0.58 ± 0.20	0.80 ± 0.28	0.67 ± 0.22

Table B2. Classification performance across datasets of DualCSN

Dataset	Accuracy	Precision	Recall	F1-score
RT25kpc	0.78 ± 0.03	0.77 ± 0.07	0.77 ± 0.07	0.77 ± 0.02
RT50kpc	0.76 ± 0.09	0.78 ± 0.12	0.67 ± 0.08	0.70 ± 0.06
RT100kpc	0.50 ± 0.13	0.47 ± 0.09	0.92 ± 0.04	0.61 ± 0.06
T25kpc	0.75 ± 0.02	0.67 ± 0.03	0.71 ± 0.06	0.69 ± 0.03
T50kpc	**0.83 ± 0.02**	**0.89 ± 0.08**	0.69 ± 0.08	0.77 ± 0.04
T100kpc	0.72 ± 0.13	0.61 ± 0.16	0.74 ± 0.26	0.64 ± 0.22
T50kpcSUB	0.69 ± 0.08	0.62 ± 0.06	**0.92 ± 0.04**	**0.74 ± 0.04**
T100kpcSUB	0.64 ± 0.07	0.59 ± 0.06	0.82 ± 0.10	0.68 ± 0.03

Table B3. Classification performance across datasets of DualSSN

Dataset	Accuracy	Precision	Recall	F1-score
RT25kpc	**0.84 ± 0.02**	**0.84 ± 0.04**	0.80 ± 0.04	0.82 ± 0.02
RT50kpc	0.78 ± 0.03	0.77 ± 0.06	0.68 ± 0.06	0.72 ± 0.04
RT100kpc	0.51 ± 0.16	0.50 ± 0.17	**0.95 ± 0.10**	0.63 ± 0.08
T25kpc	0.79 ± 0.02	0.73 ± 0.06	0.76 ± 0.11	0.74 ± 0.04
T50kpc	0.83 ± 0.02	0.85 ± 0.07	0.72 ± 0.08	0.77 ± 0.03
T100kpc	0.81 ± 0.03	0.76 ± 0.05	0.75 ± 0.14	0.74 ± 0.07
T50kpcSUB	0.73 ± 0.12	0.66 ± 0.10	0.94 ± 0.04	**0.77 ± 0.07**
T100kpcSUB	0.81 ± 0.06	0.77 ± 0.10	0.88 ± 0.07	**0.81 ± 0.04**

Table C1. Classification performance across datasets of CNN

Dataset	Accuracy	Precision	Recall	F1-score
Raw+RT25kpc	0.53 ± 0.05	0.51 ± 0.06	0.56 ± 0.23	0.49 ± 0.17
Raw+RT50kpc	**0.61 ± 0.09**	**0.60 ± 0.11**	**0.65 ± 0.13**	**0.61 ± 0.04**
Raw+RT100kpc	0.59 ± 0.06	0.55 ± 0.07	0.63 ± 0.12	0.58 ± 0.07
All versions	0.57 ± 0.06	0.54 ± 0.08	0.65 ± 0.17	0.57 ± 0.04

Table C2. Classification performance across datasets of DualCSN

Dataset	Accuracy	Precision	Recall	F1-score
Raw+RT25kpc	0.74 ± 0.03	**0.74 ± 0.08**	0.72 ± 0.08	0.72 ± 0.02
Raw+RT50kpc	**0.76 ± 0.03**	0.72 ± 0.04	**0.79 ± 0.05**	**0.76 ± 0.02**
Raw+RT100kpc	0.70 ± 0.08	0.67 ± 0.09	0.74 ± 0.06	0.70 ± 0.05
All versions	0.71 ± 0.03	0.70 ± 0.04	0.66 ± 0.10	0.67 ± 0.05

Table C3. Classification performance across datasets of DualSSN

Dataset	Accuracy	Precision	Recall	F1-score
Raw+RT25kpc	**0.80 ± 0.07**	**0.82 ± 0.14**	0.80 ± 0.10	**0.79 ± 0.05**
Raw+RT50kpc	0.80 ± 0.07	0.77 ± 0.10	**0.83 ± 0.07**	0.79 ± 0.06
Raw+RT100kpc	0.77 ± 0.04	0.76 ± 0.11	0.78 ± 0.09	0.76 ± 0.02
All versions	0.75 ± 0.04	0.76 ± 0.08	0.71 ± 0.15	0.72 ± 0.07

Beam cropping proved important

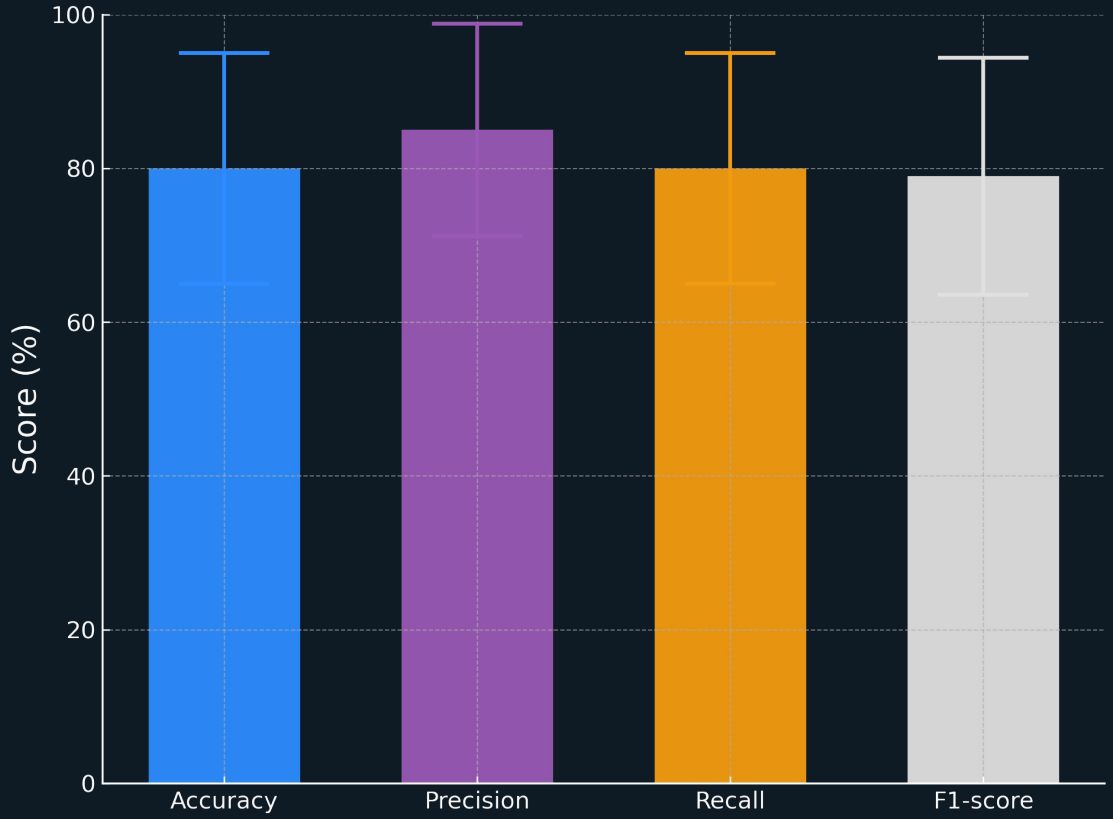
Table 2. Classification performance across four classifiers on raw data with percentile clipping and constant field of view.

Classifier	Accuracy	Precision	Recall	F1-score
CNN	0.56 ± 0.09	0.41 ± 0.06	**0.76 ± 0.23**	0.51 ± 0.10
DualCSN	0.77 ± 0.09	0.64 ± 0.12	0.76 ± 0.08	0.68 ± 0.07
DualSSN	**0.77 ± 0.13**	**0.66 ± 0.12**	0.75 ± 0.09	**0.69 ± 0.07**

Table 3. Classification performance across three classifiers on raw data with percentile clipping and beam-cropping.

Classifier	Accuracy	Precision	Recall	F1-score
CNN	0.72 ± 0.05	0.69 ± 0.07	0.64 ± 0.20	0.63 ± 0.17
DualCSN	**0.84 ± 0.04**	0.86 ± 0.06	**0.76 ± 0.07**	**0.80 ± 0.05**
DualSSN	0.82 ± 0.03	**0.86 ± 0.06**	0.69 ± 0.10	0.76 ± 0.05

Classify Radio Relics against Radio Halos



This plot is for T50

The classification pipeline starts with processing of data

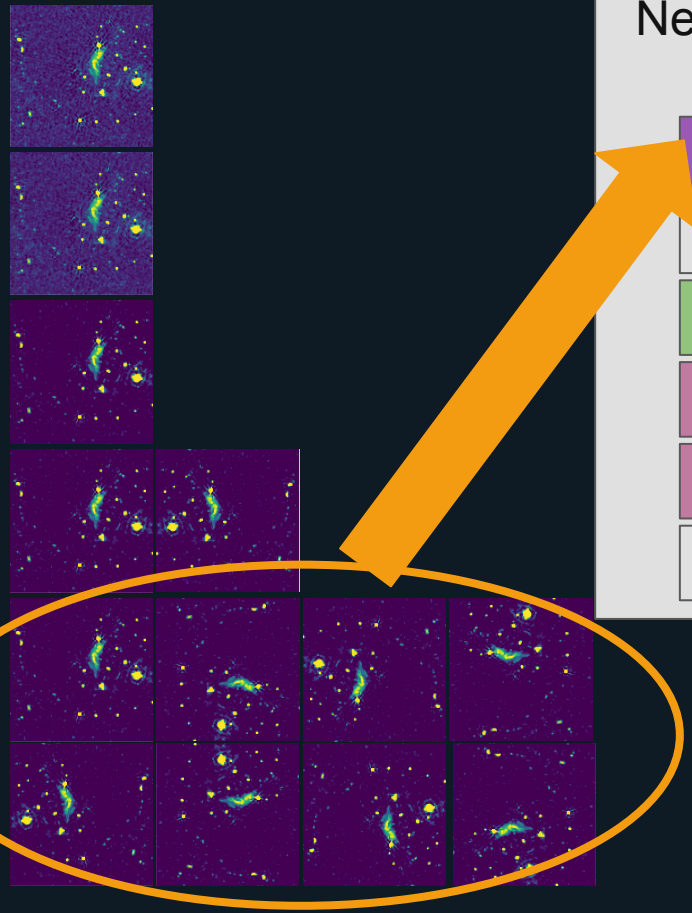
Cropping to (1, 512, 512)

Downsample to (1, 128, 128)

Normalisation

Horizontal flipping

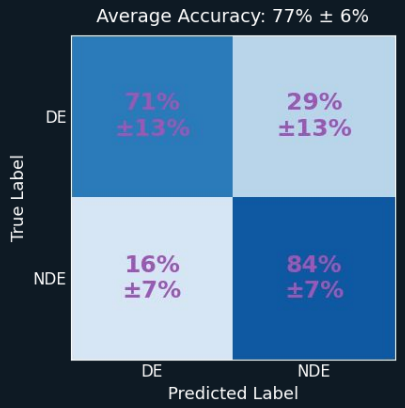
Quarter turn rotations



Convolutional Neural Network (CNN)

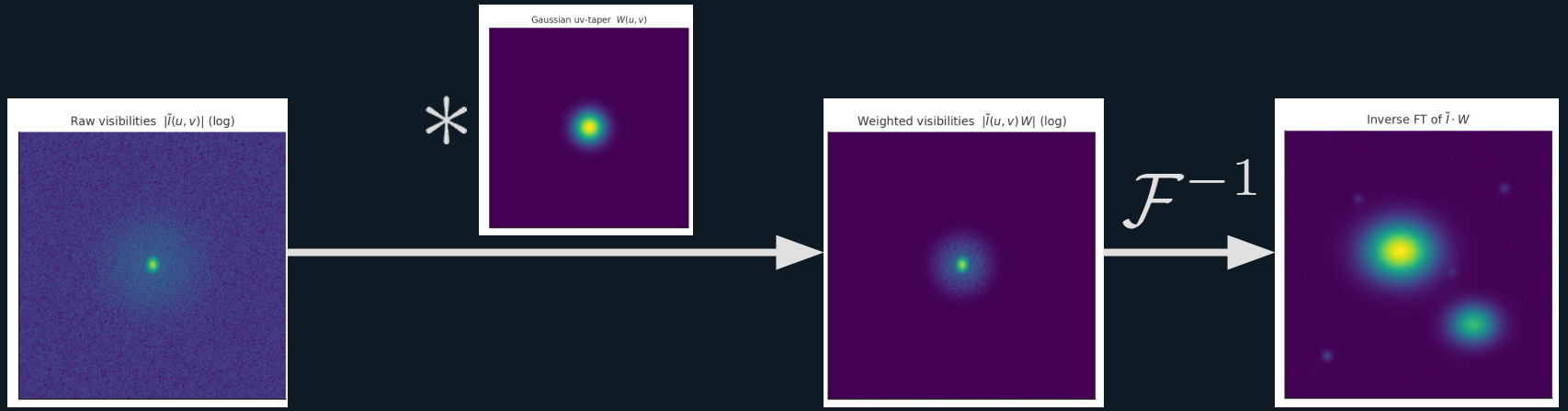
- Images
- Conv/BN/ LReLU
- Flatten
- LReLU
- Linear
- Logits

x5

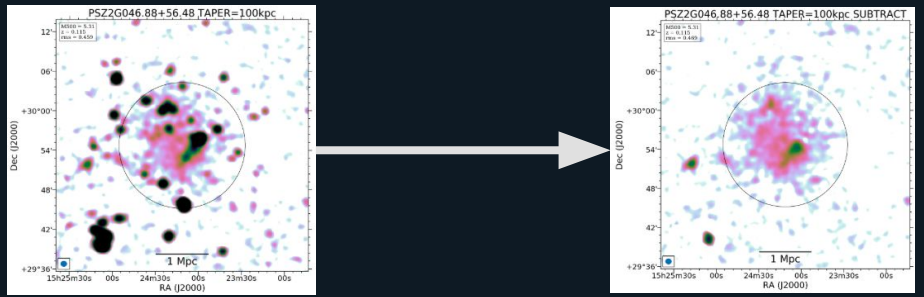


Trained and evaluated model

Tapering downweights longer baselines

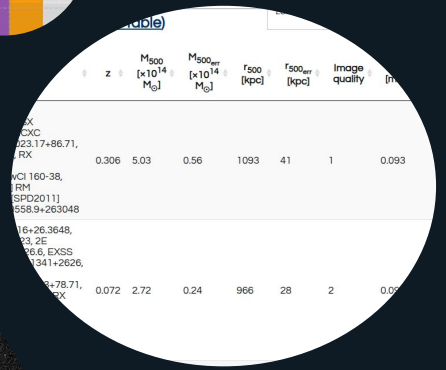


Point source subtraction

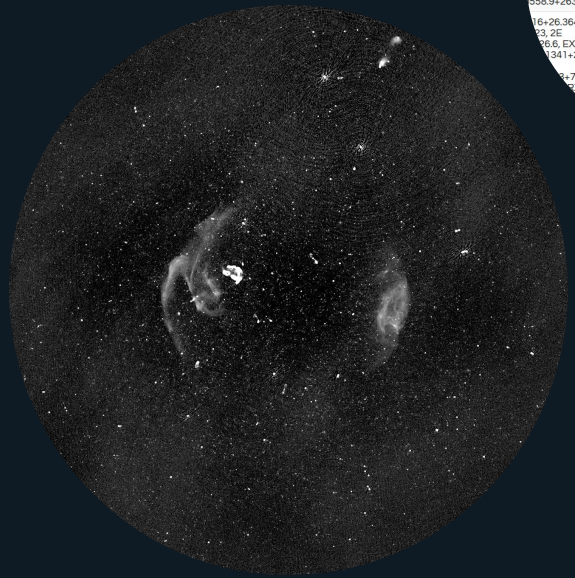


Future prospects

- Encode redshift, cluster mass, and cluster size as input
- Apply method on other datasets (e.g. MGCLS, MERGHERS)



z	M_{500} [$\times 10^{14}$ $M_{\odot}$]	$M_{500_{err}}$ [$\times 10^{14}$ $M_{\odot}$]	r_{500} [kpc]	$r_{500_{err}}$ [kpc]	Image quality	
0.306	5.03	0.56	1093	41	1	0.093
0.072	2.72	0.24	966	28	2	0.0

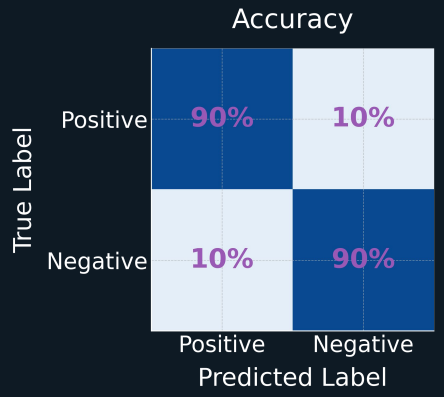


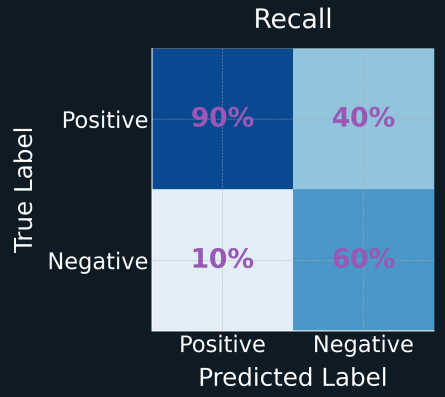
Courtesy: Konstantinos Kolokythas

Bonus slides 2 - Machine learning technicalities

The four horsemen of classification evaluation

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

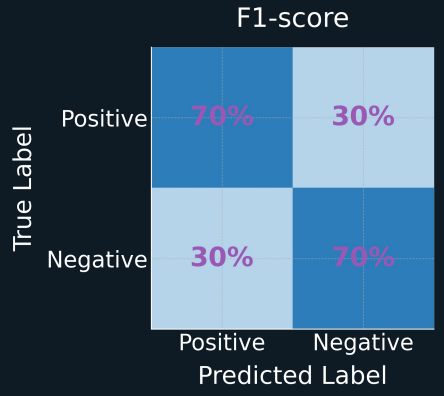


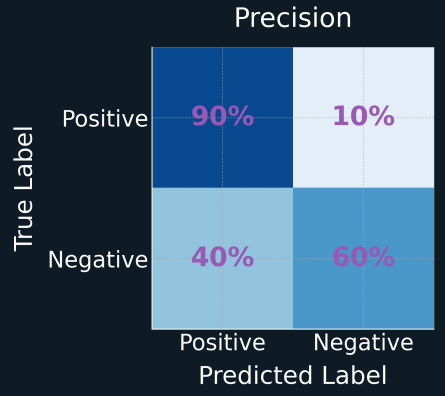


"Completeness"

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$





"Purity"

$$Precision = \frac{TP}{TP + FP}$$

Scattering transforms are iterative wavelet transforms

$$\Psi_{J,L}(x) = \left[X, |X \star \psi_{j,\ell}|, ||X \star \psi_{j,\ell}| \star \psi_{j',\ell'}|\right]_{1 \le \ell,\ell' \le L,\ 1 \le j < j' \le J}$$

$$\phi_J(u) = 2^{-2J}\phi(2^{-J}u)$$

$$\begin{aligned} S_{J,L}(X) &= \Psi_{J,L}(X) \star \phi_J \\ &= \left[X \star \phi_J, |X \star \psi_{j,\ell}| \star \phi_J, ||X \star \psi_{j,\ell}| \star \psi_{j',\ell'}| \star \phi_J\right]_{1 \le j < j' \le J, 1 \le \ell,\ell' \le L} \end{aligned}$$

$X \star \phi_2$ $|X \star \psi_{0,0}| \star \phi_2$ $|X \star \psi_{0,1}| \star \phi_2$ $|X \star \psi_{1,0}| \star \phi_2$ $|X \star \psi_{1,1}| \star \phi_2$ $||X \star \psi_{0,0}| \star \psi_{1,0}| \star \phi_2$ $||X \star \psi_{0,0}| \star \psi_{1,1}| \star \phi_2$ $||X \star \psi_{0,1}| \star \psi_{1,0}| \star \phi_2$ $||X \star \psi_{0,1}| \star \psi_{1,1}| \star \phi_2$

Low pass filter for scattering transform

Littlewood-Paley equality $\sum_{j=-\infty}^{+\infty} |\hat{\psi}(2^j\omega)|^2 = 1\ ,\ \forall\omega > 0$

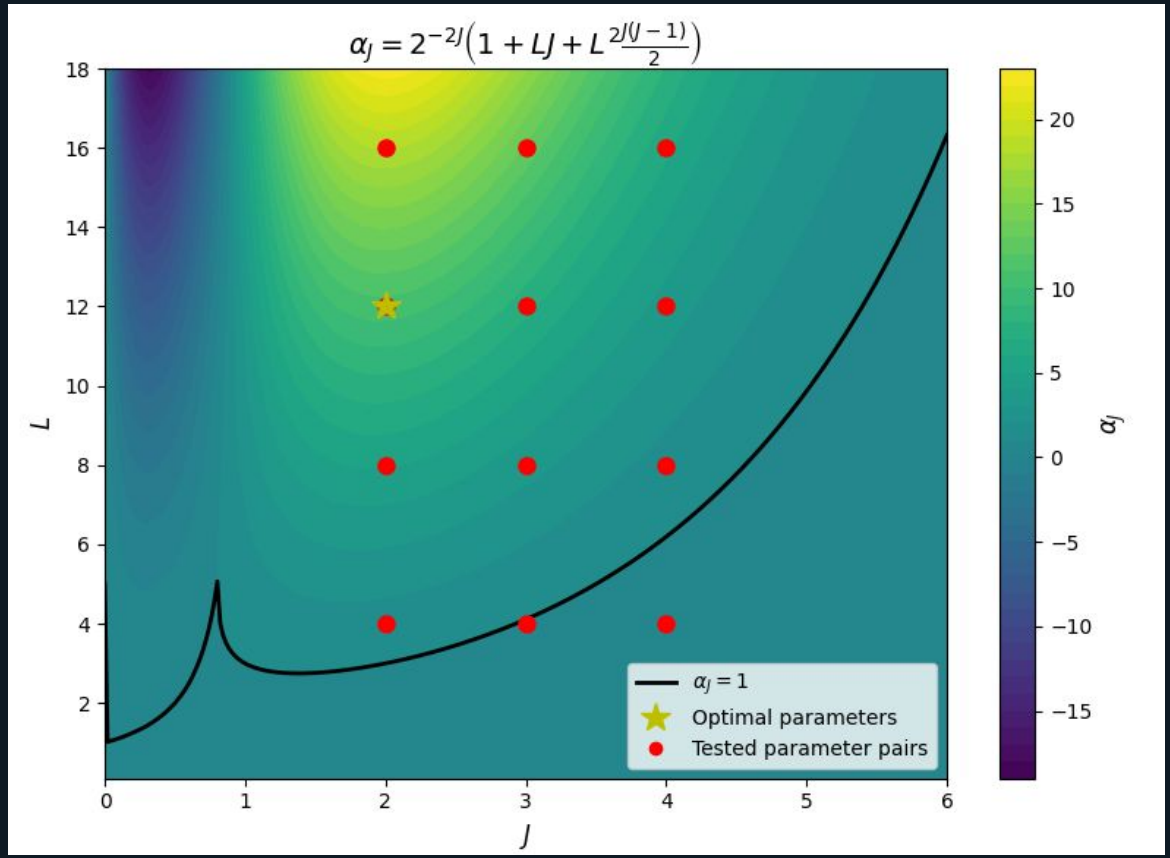
Conservation of energy

Capture lower frequencies of signal $[-2^{-J}\pi, 2^{-J}\pi]$

With low-pass filter $\hat{\Phi}_J(\omega) = \left(\sum_{j=J+1}^{+\infty} |\hat{\psi}(2^j\omega)|^2\right)^{1/2}$ **satisfying** $\int \Phi_J(t)dt = 1$

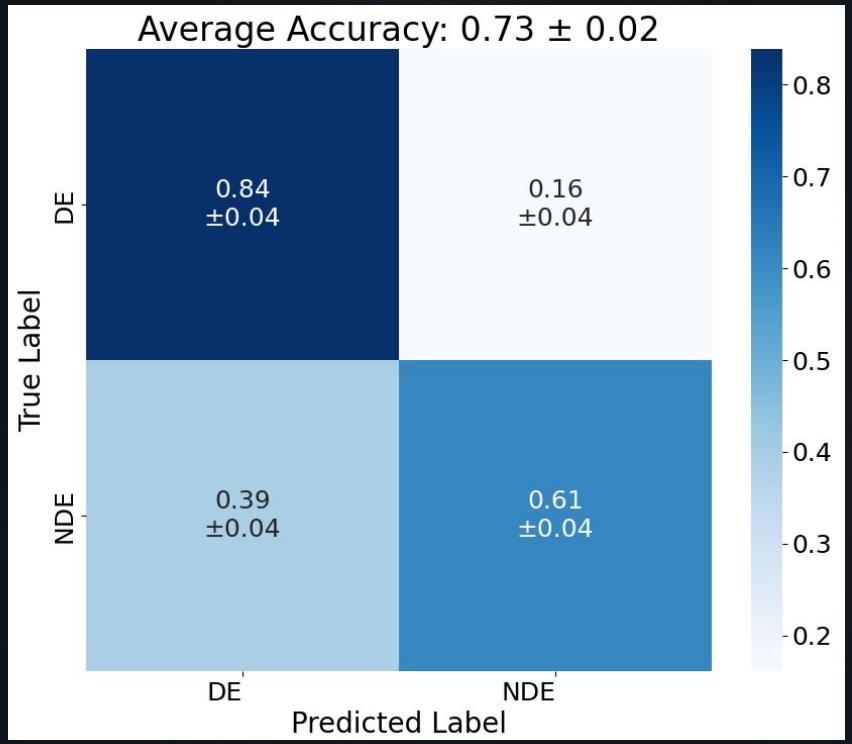
$\rightarrow$ preserves norm and is therefore invertible

Performance improves with number of scattering coefficients



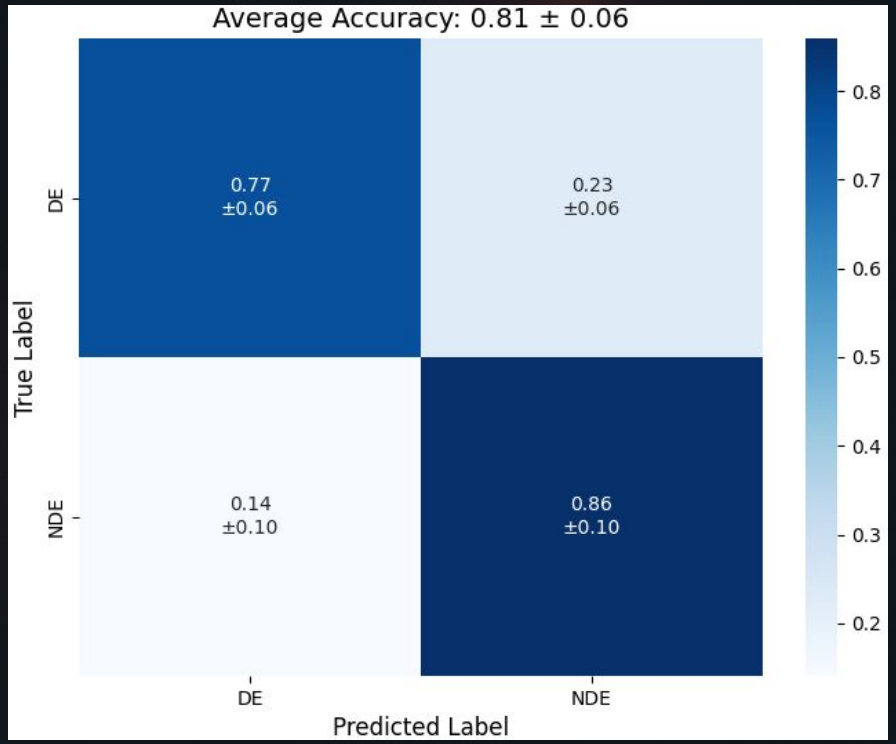
Confusion matrices for classification of radio clusters (June 2025)

MGCLS



CNN2: 8 layers, 2.5M parameters

PSZ2



ScatterSqueezeDualNet: 14 + 2 layers, 3M parameters