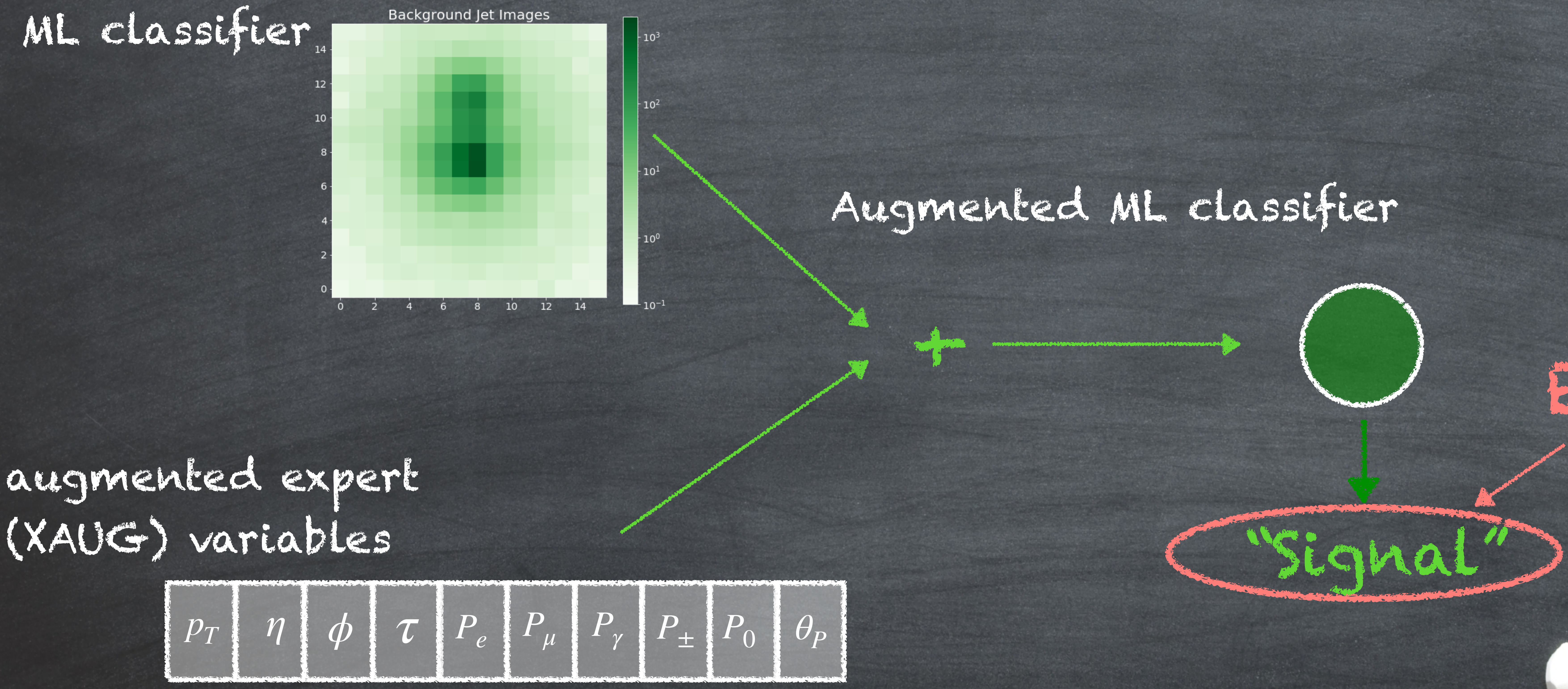
XAI for ML Jet Taggers

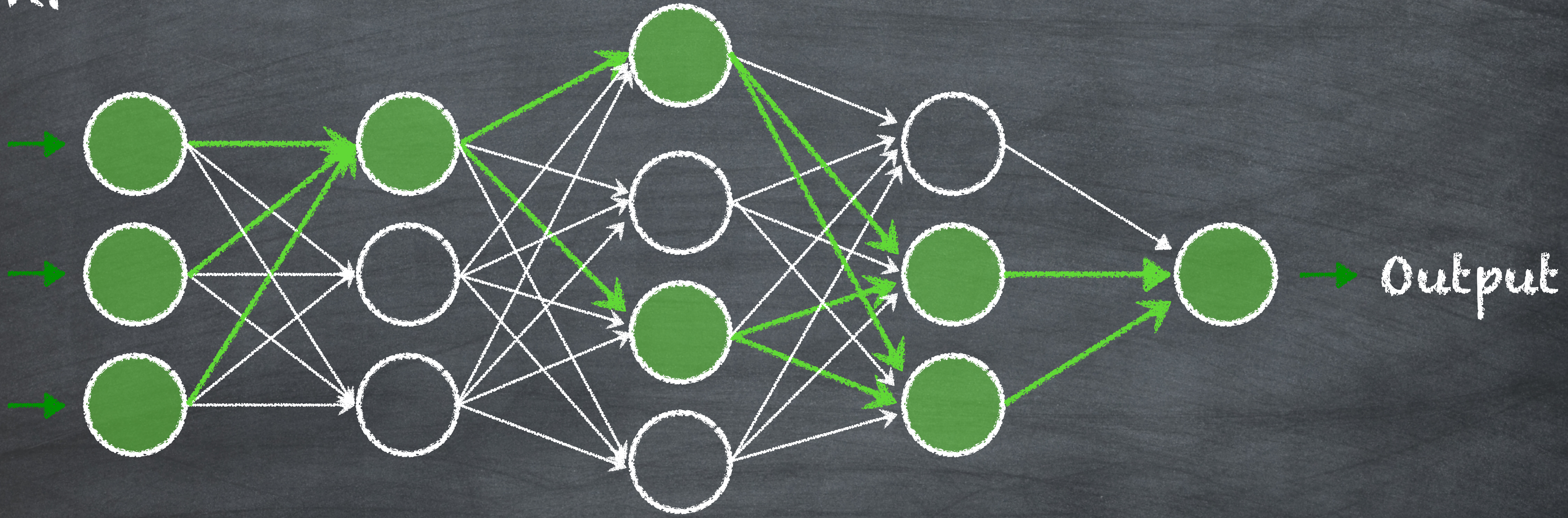
Motivation:

Explain ML decisions of identification of jets Using augmented expert variables (XAUG).



Deep Neural Network

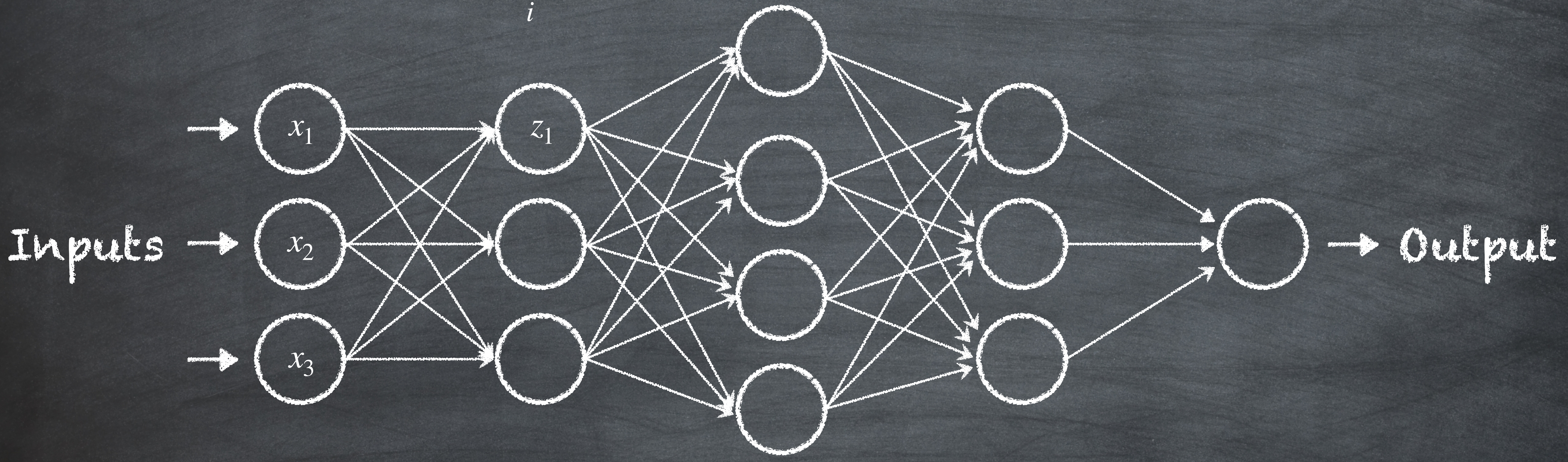
Information propagates forward through the network as a function of the inputs, weights, and biases to make a decision.



$$z_j = \sum_i x_i w_i + b$$

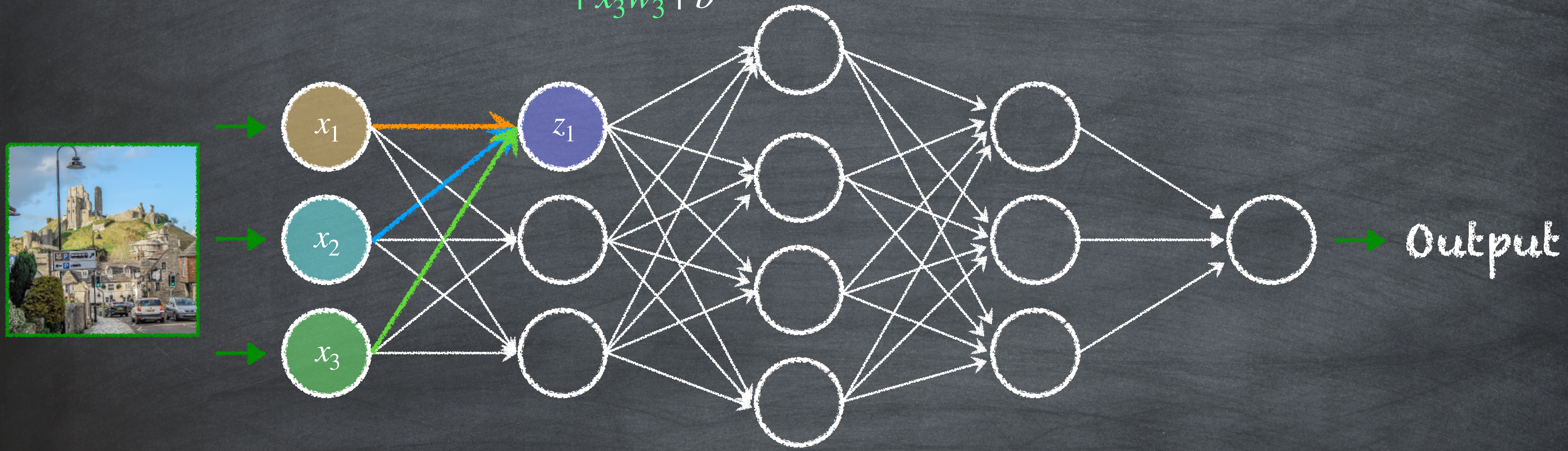
Deep Neural Network

$$z_j = \sum_i x_i w_i + b$$



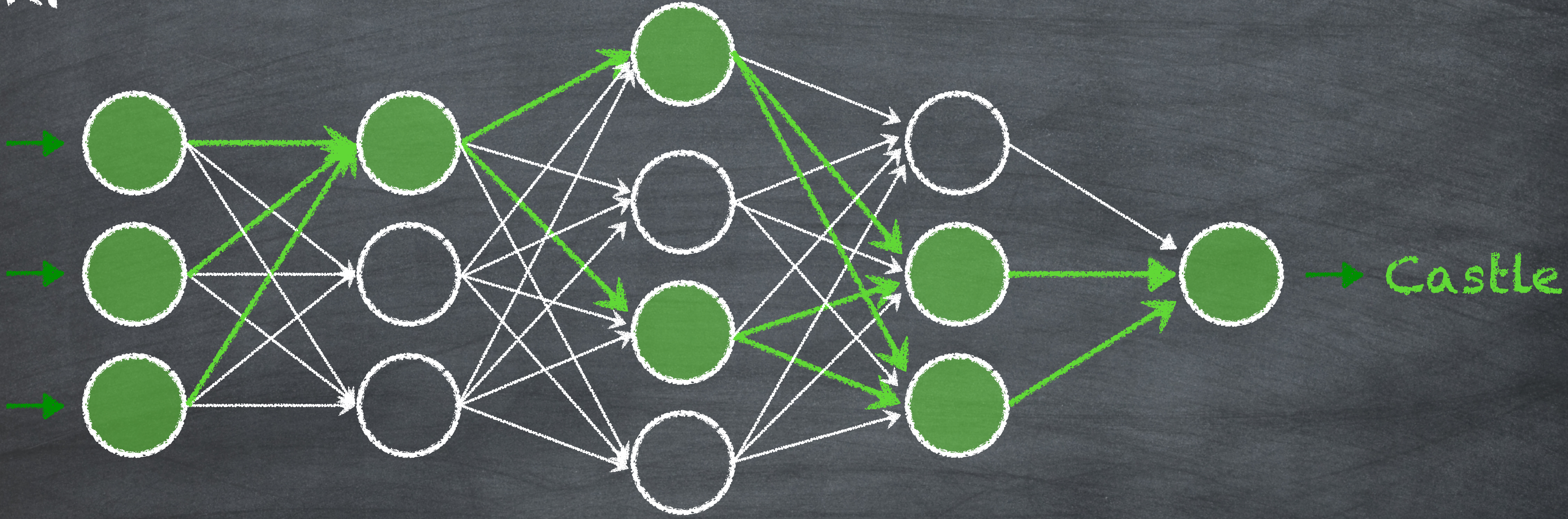
Deep Neural Network

$$z_1 = x_1 w_1 + x_2 w_2 + x_3 w_3 + b$$



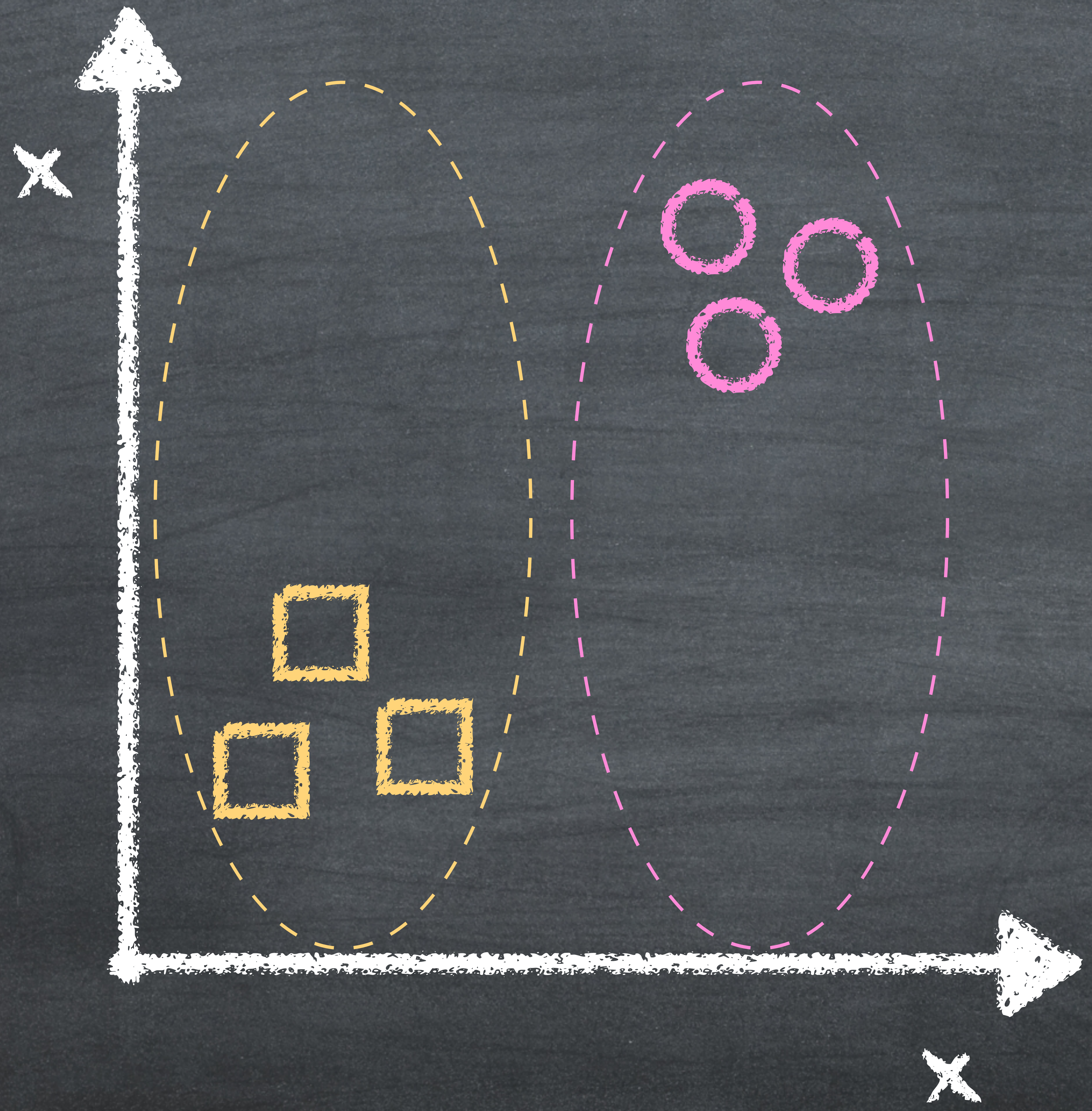
Deep Neural Network

Information propagates forward through the network as a function of the inputs, weights, and biases to make a decision.



$$z_j = \sum_i x_i w_i + b$$

Consider a trivial DNN which has been trained on only the circles and square shown in the plot as inputs.
The ovals represent where the full dataset lives.



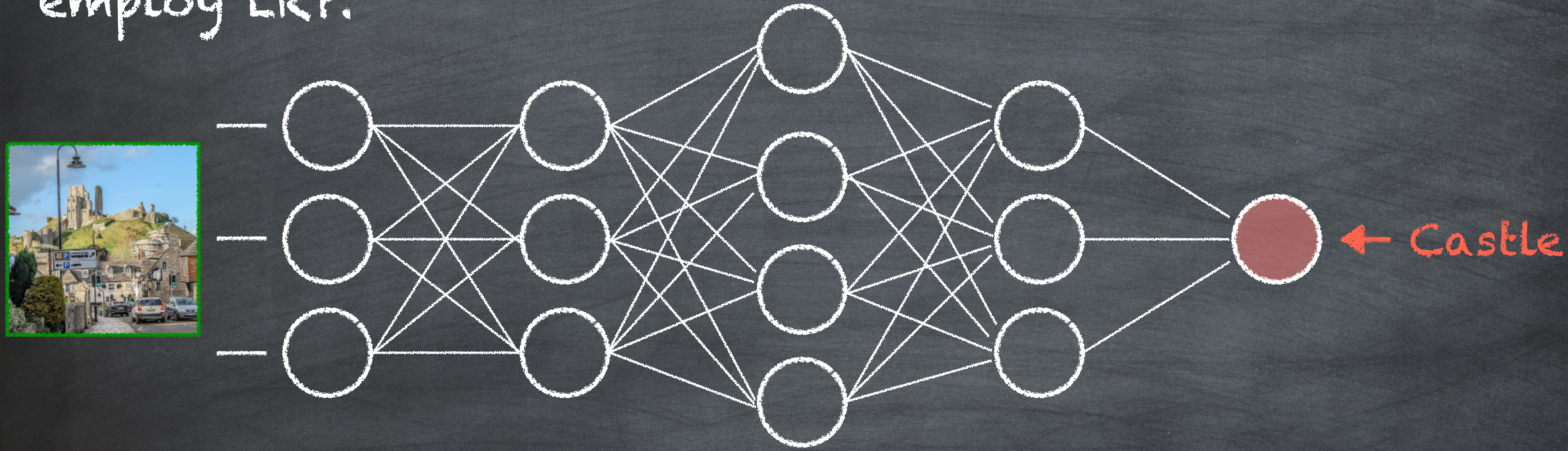
Trained on only the shapes in the plot, the network could fall into various local minimum, where the dashed lines represents the decision boundary of the minima.

However, not all of these minima accurately categorize the full dataset. We want to ensure that a network has fallen into a minima representative of the full dataset.

Want predictions supported by meaningful patterns in data.

Layerwise Relevance Propagation

To discover whether or not a network's decision-making is meaningful, we employ LRP.

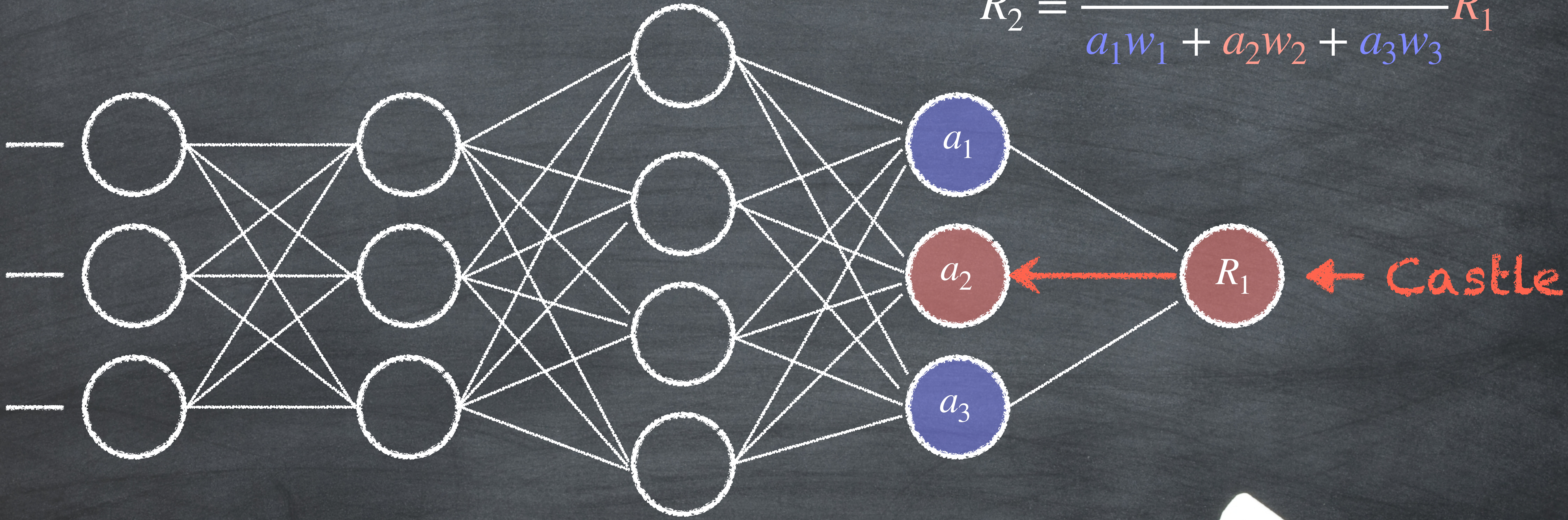


Layerwise Relevance Propagation

LRP propagates the output back to the input while conserving the value of the output as the total relevance score, R.

$$R_j = \sum_k \frac{a_j w_{jk}}{\sum_{0,j} a_j w_{jk}} R_k$$

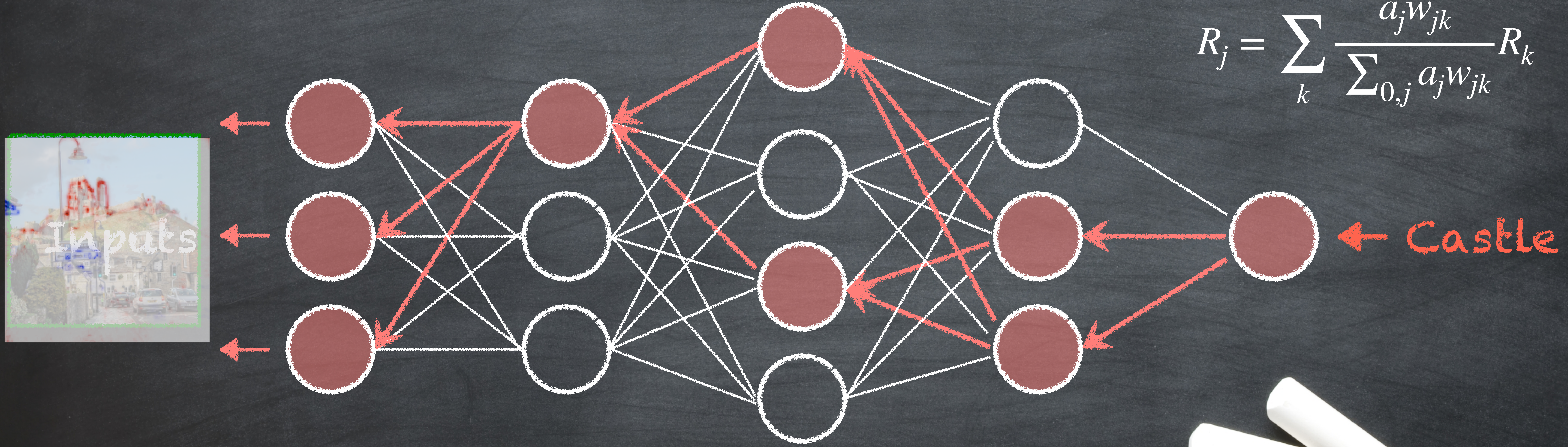
$$R_2 = \frac{a_2 w_2}{a_1 w_1 + a_2 w_2 + a_3 w_3} R_1$$



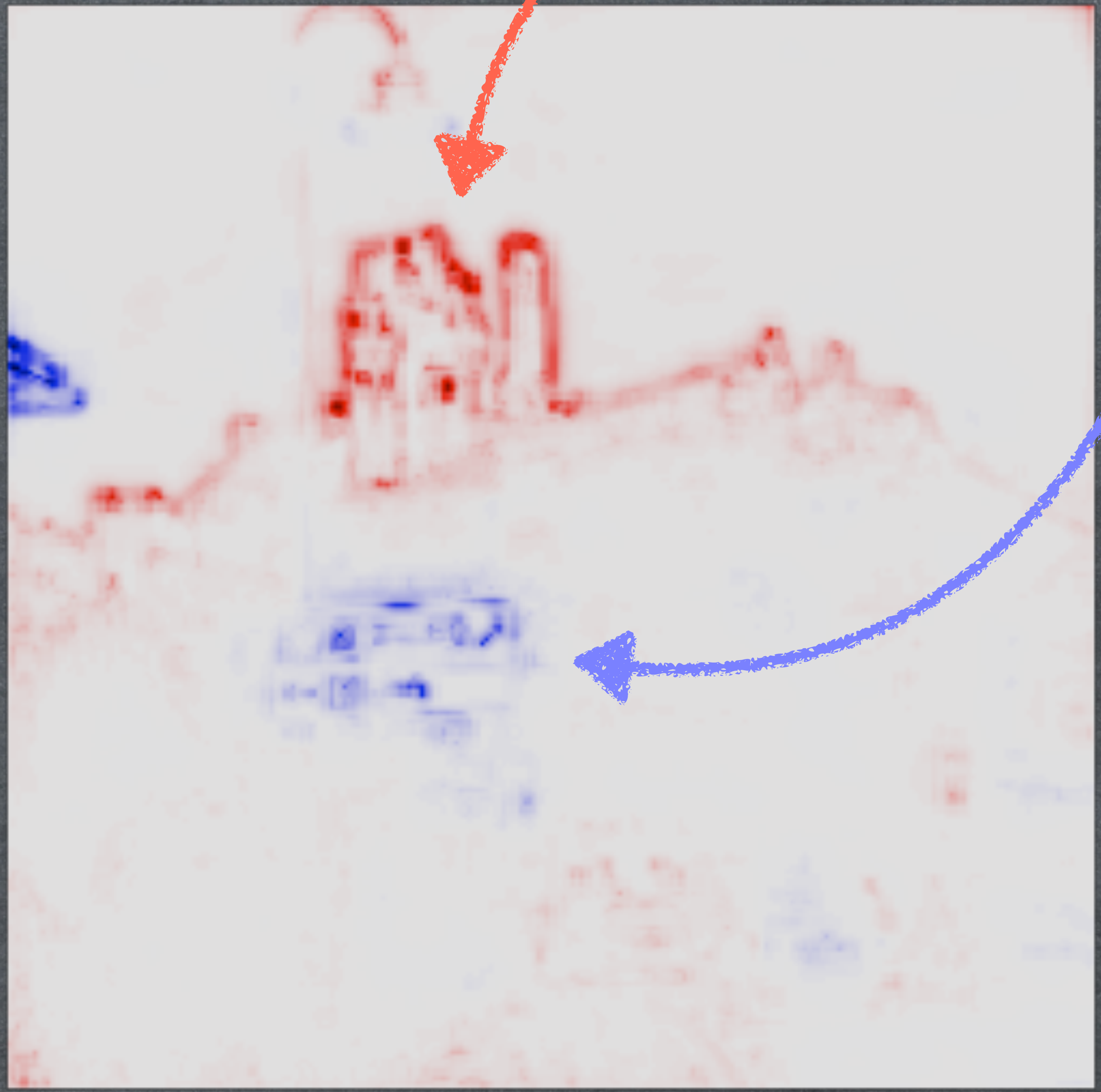
Layerwise Relevance Propagation

The total relevance is distributed among the weights, until reaching the input, where it is further distributed among the relevant features.

$$R_j = \sum_k \frac{a_j w_{jk}}{\sum_{0,j} a_j w_{jk}} R_k$$



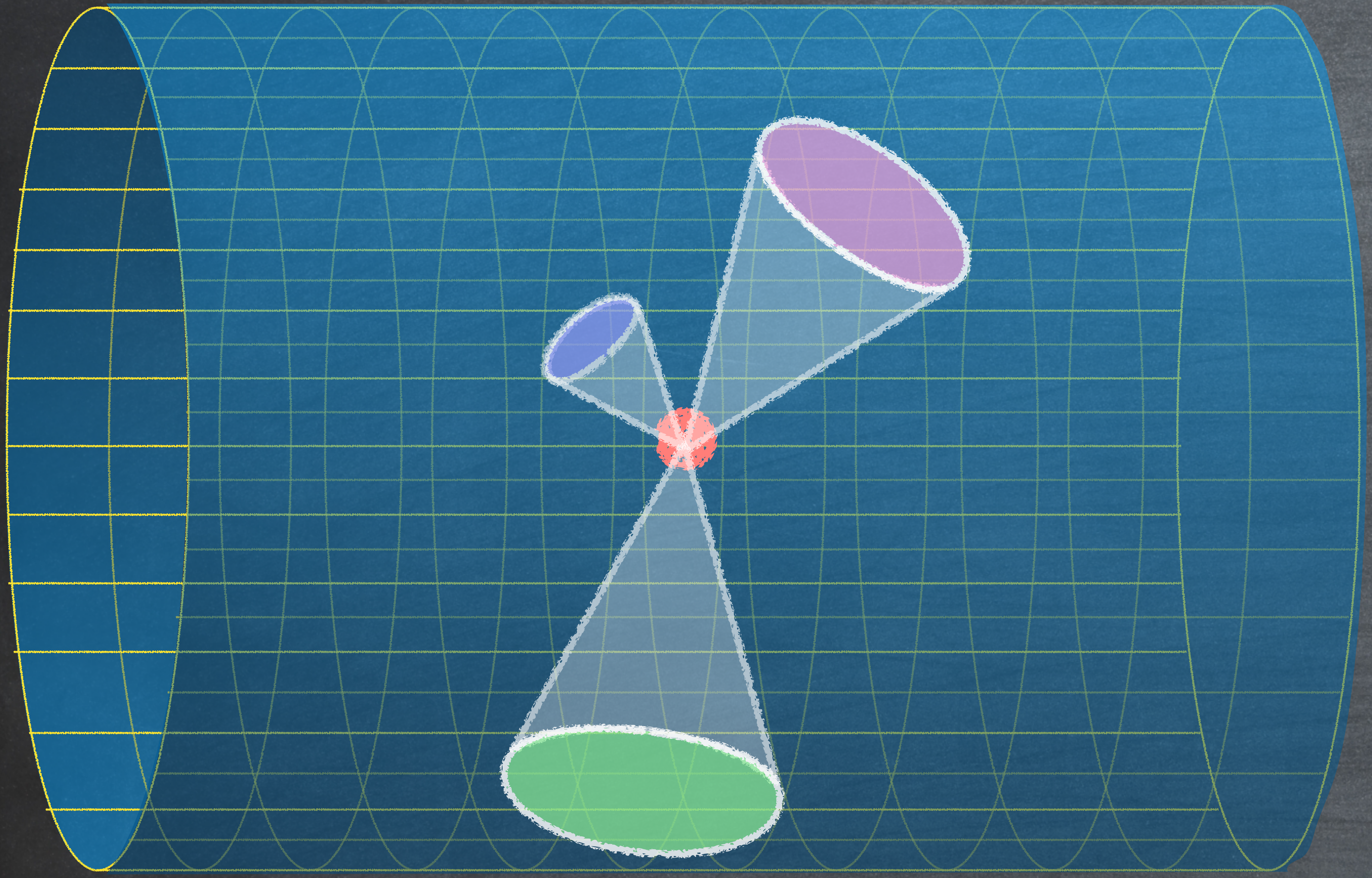
Layerwise Relevance Propagation

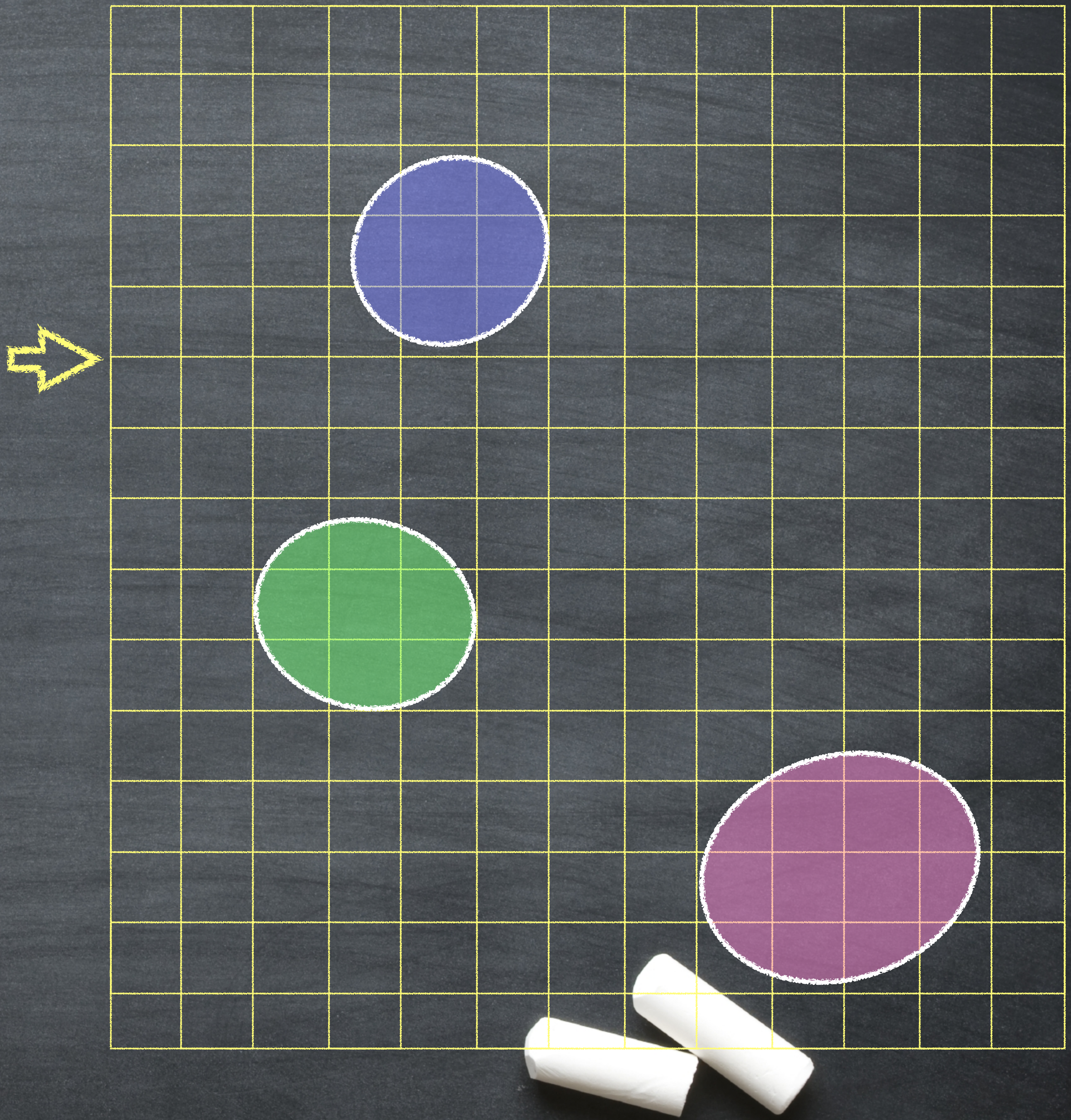


Pixels supporting the prediction

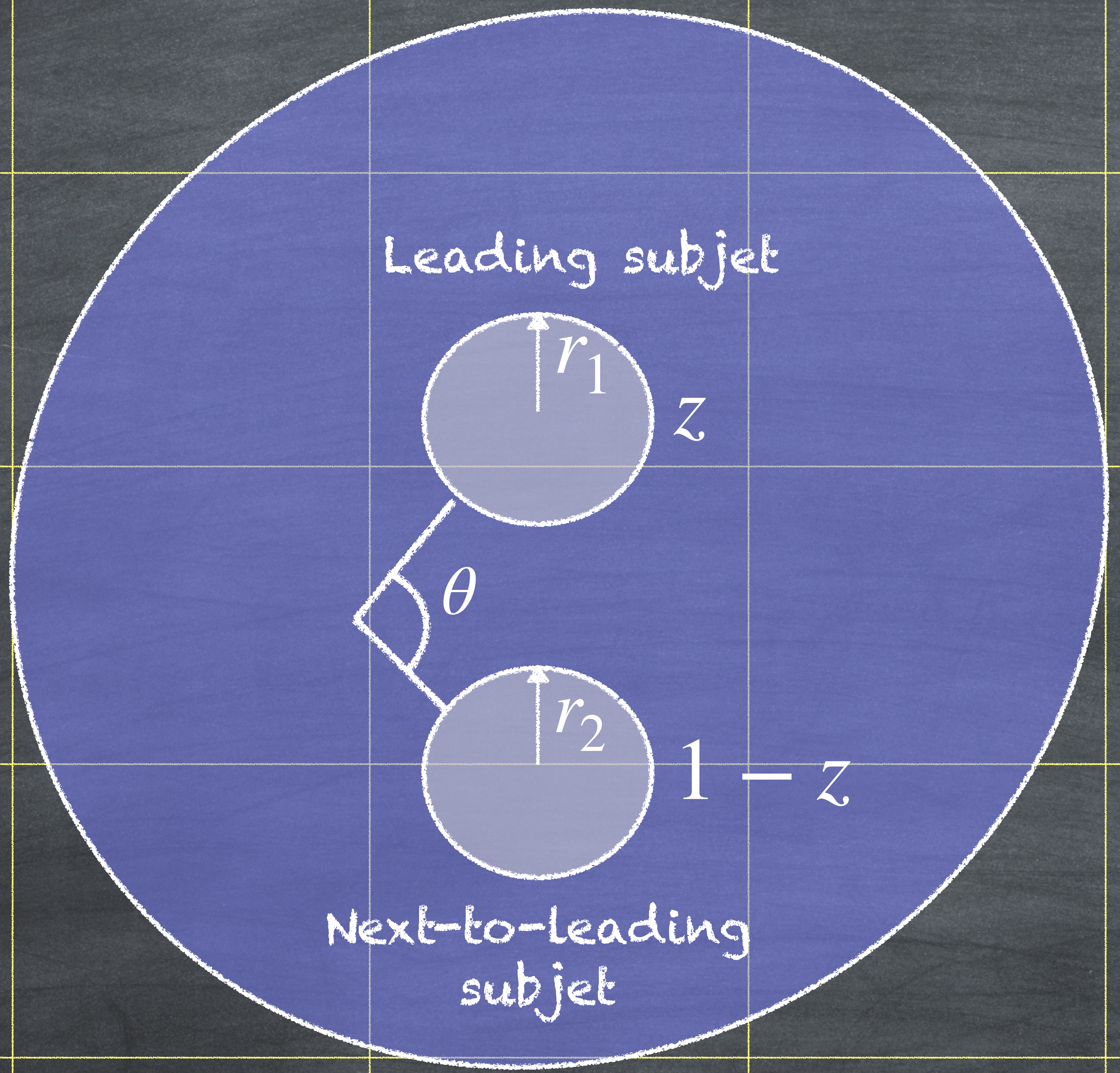
Pixels opposing the prediction

Consider a Toy Model...

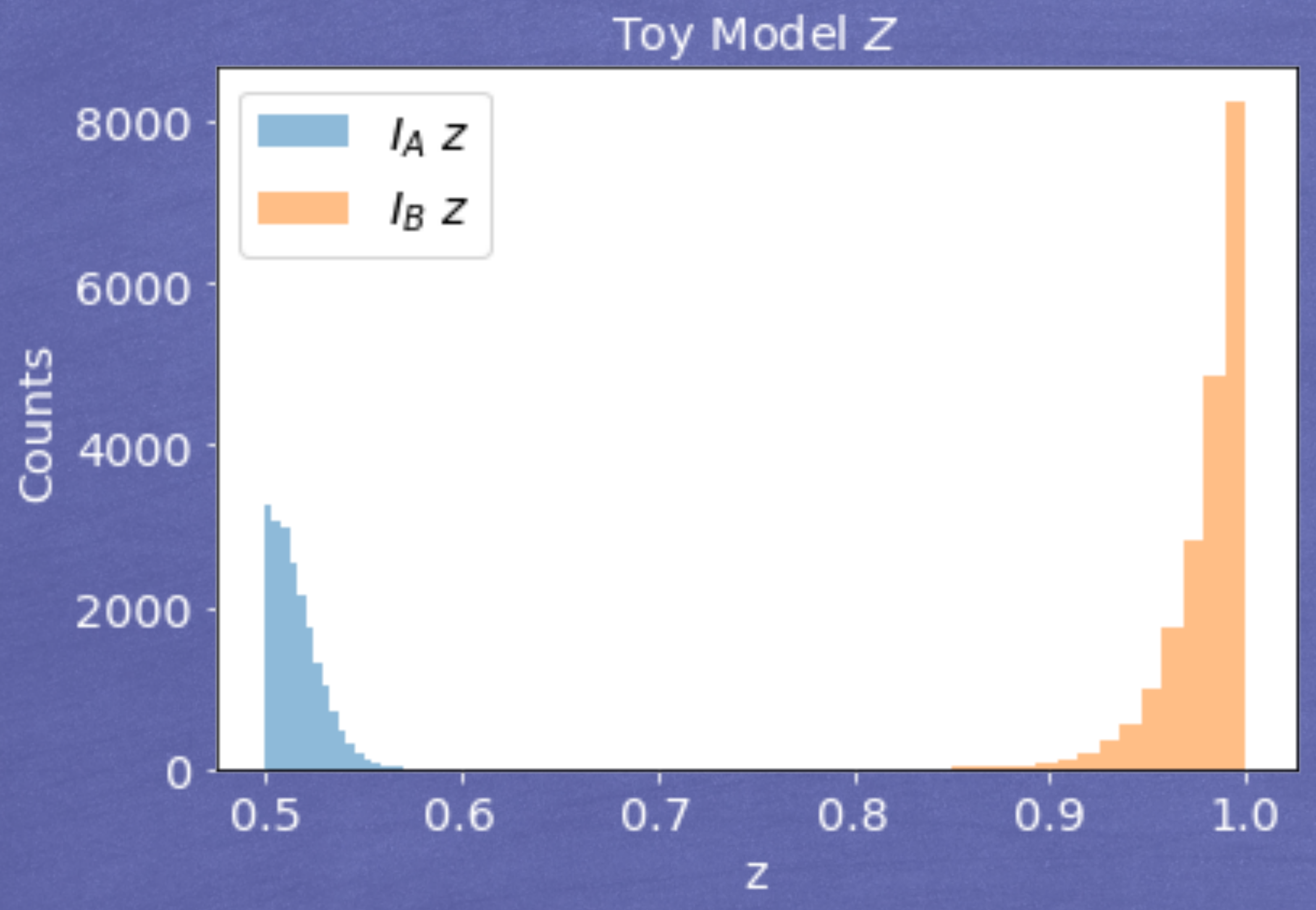


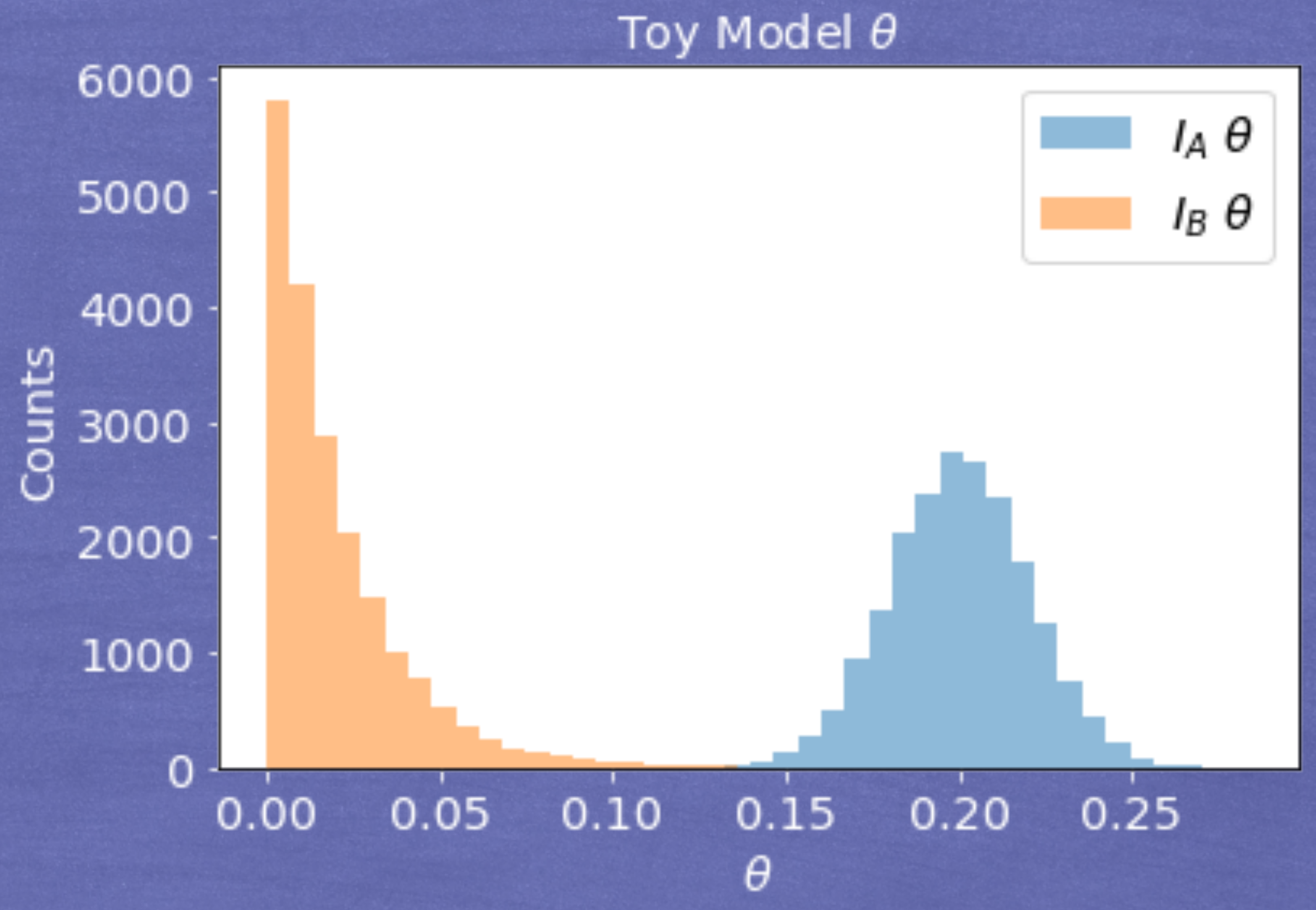


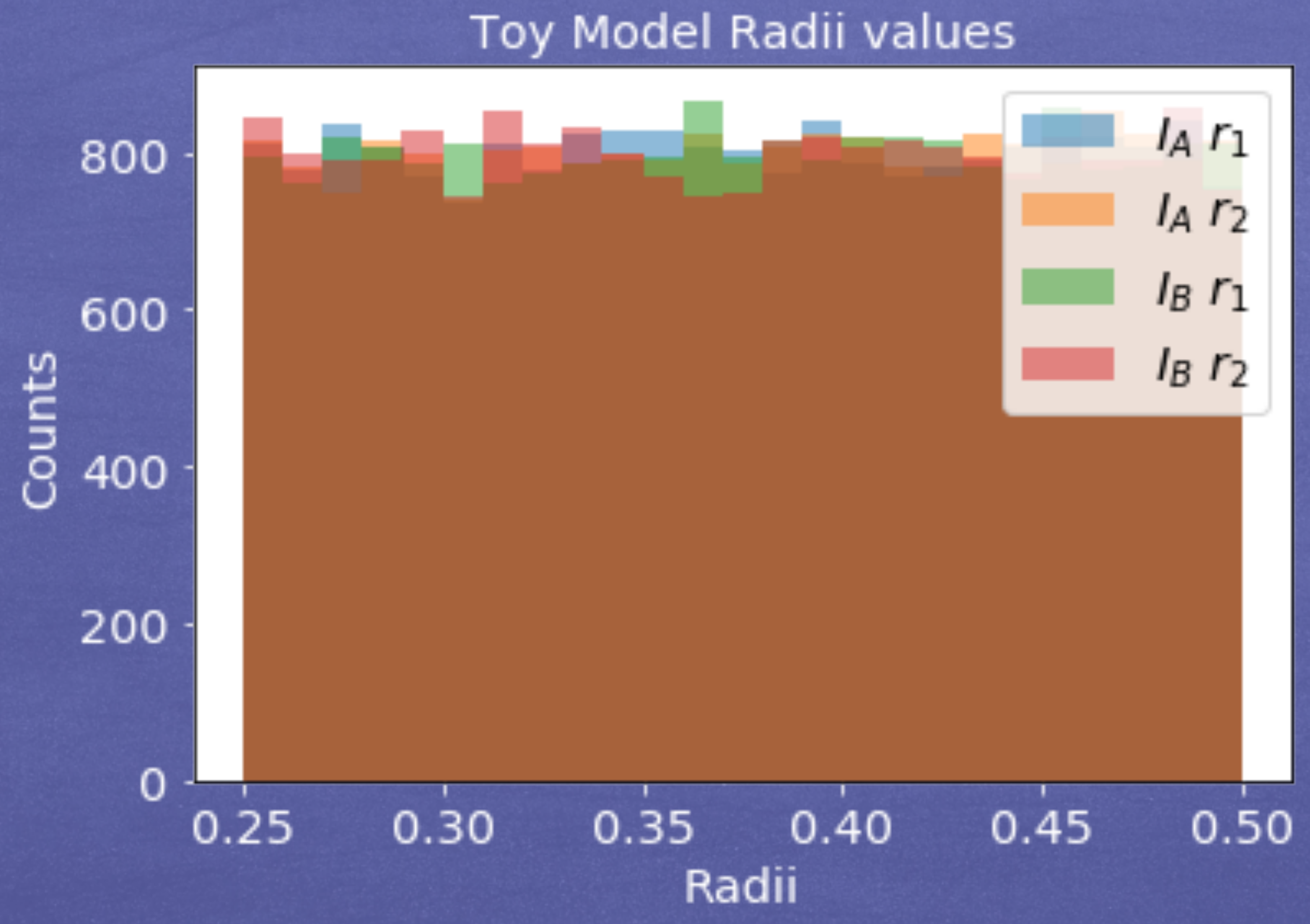
… where an event's jets are represented as images.


Leading subjet
r_1
z
θ
r_2
$1-z$
Next-to-leading subjet

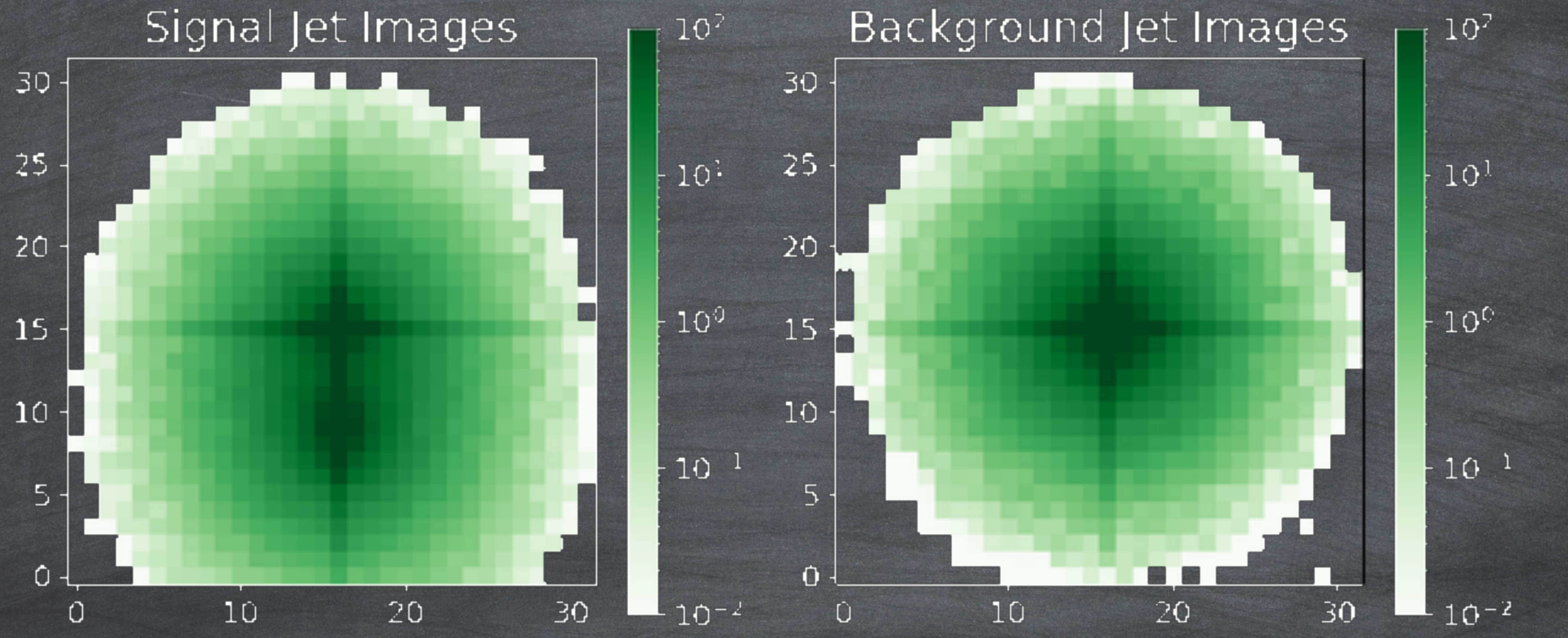
Toy Model – Input variables



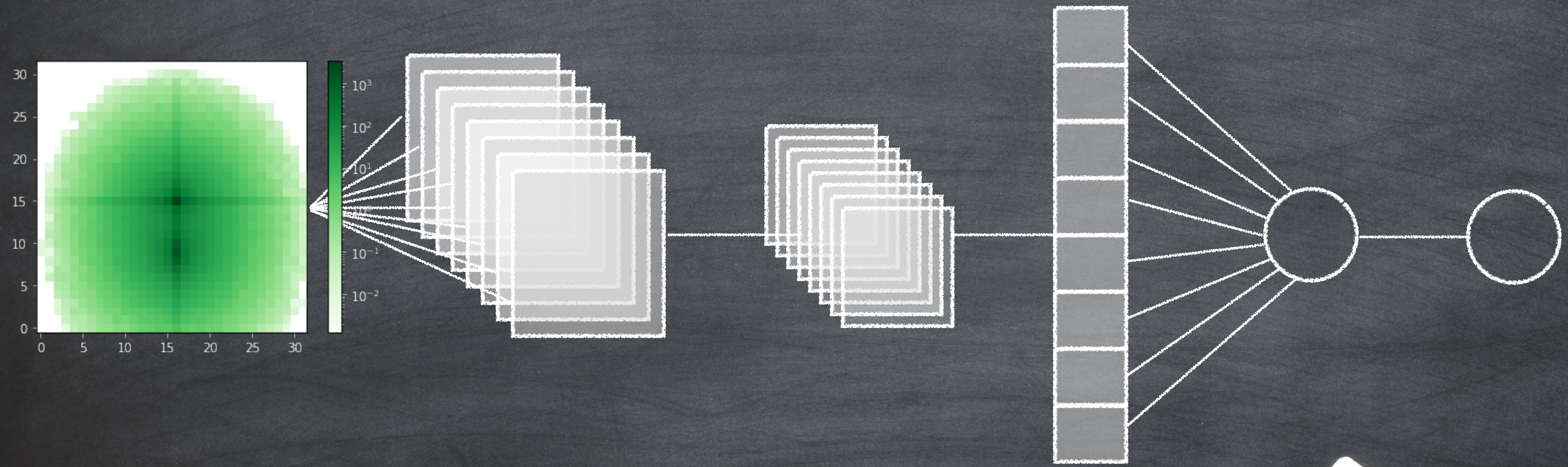




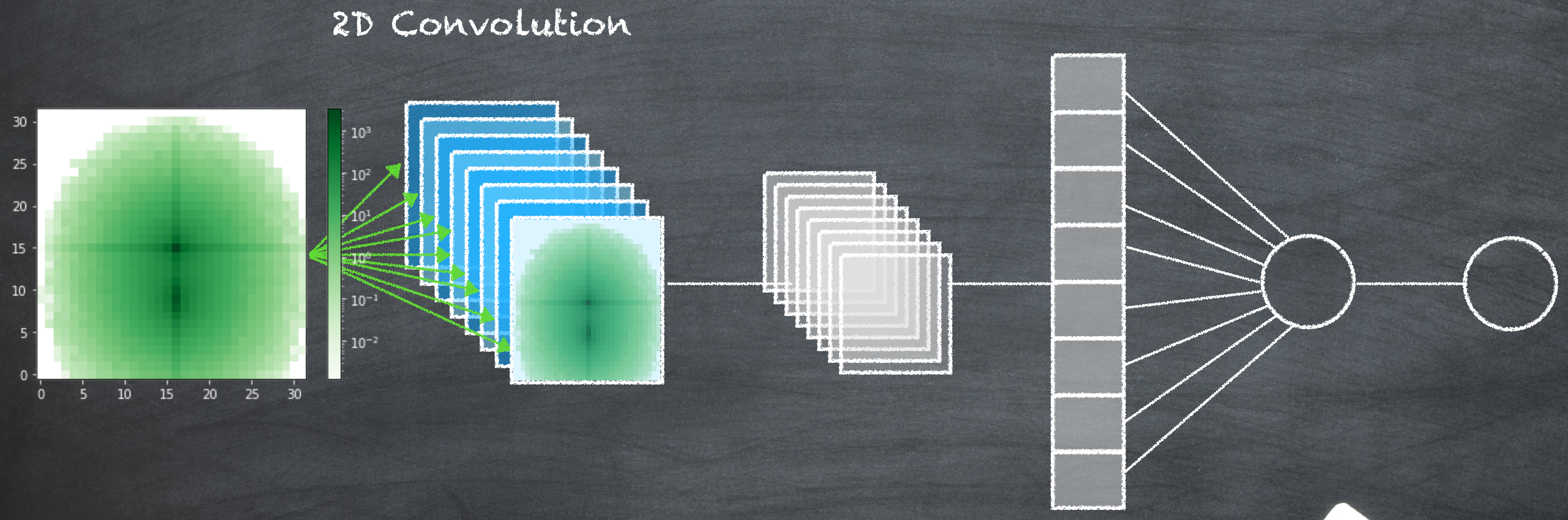
Toy Model – Images

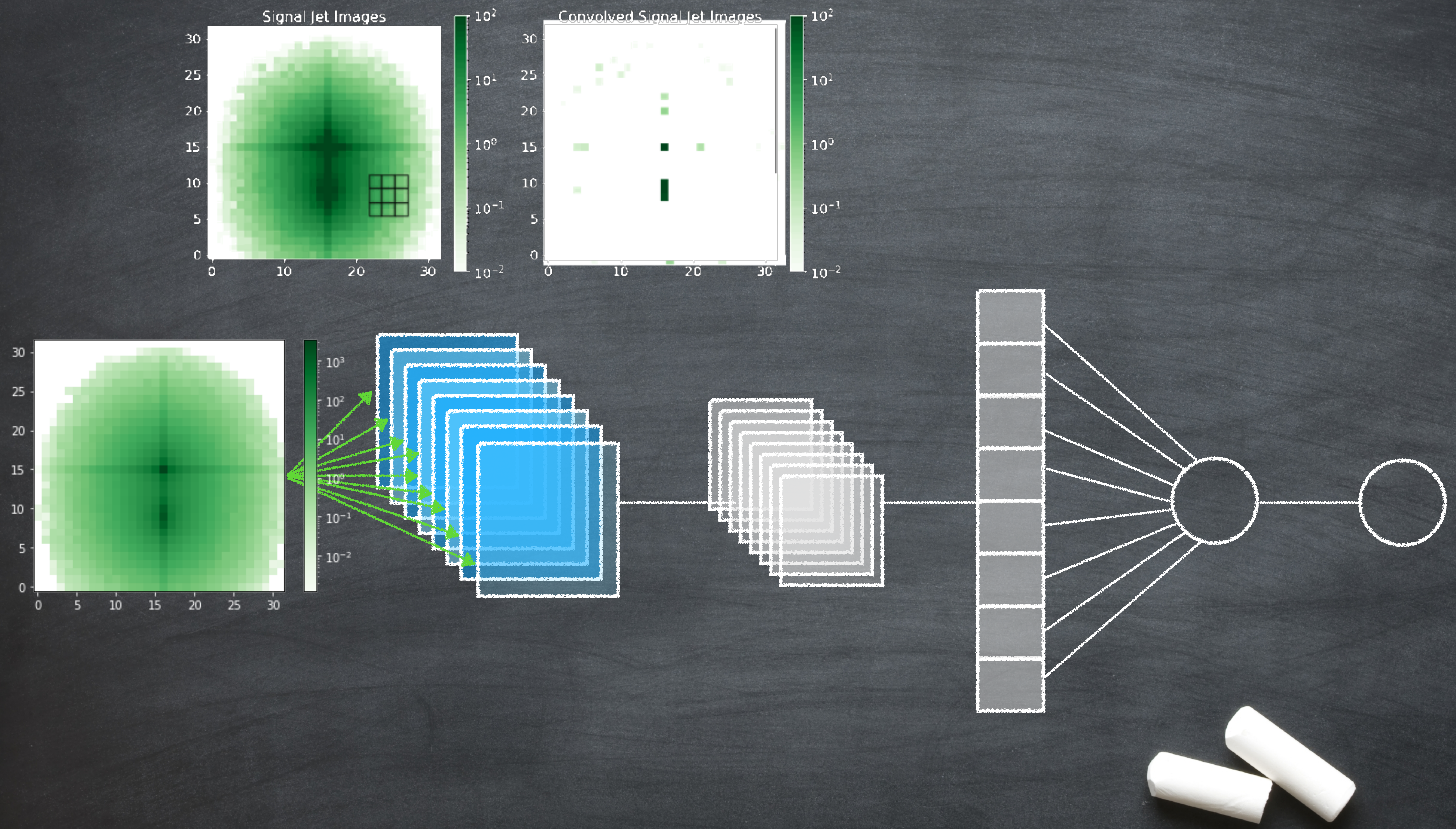


Toy Model - Forward Propagation

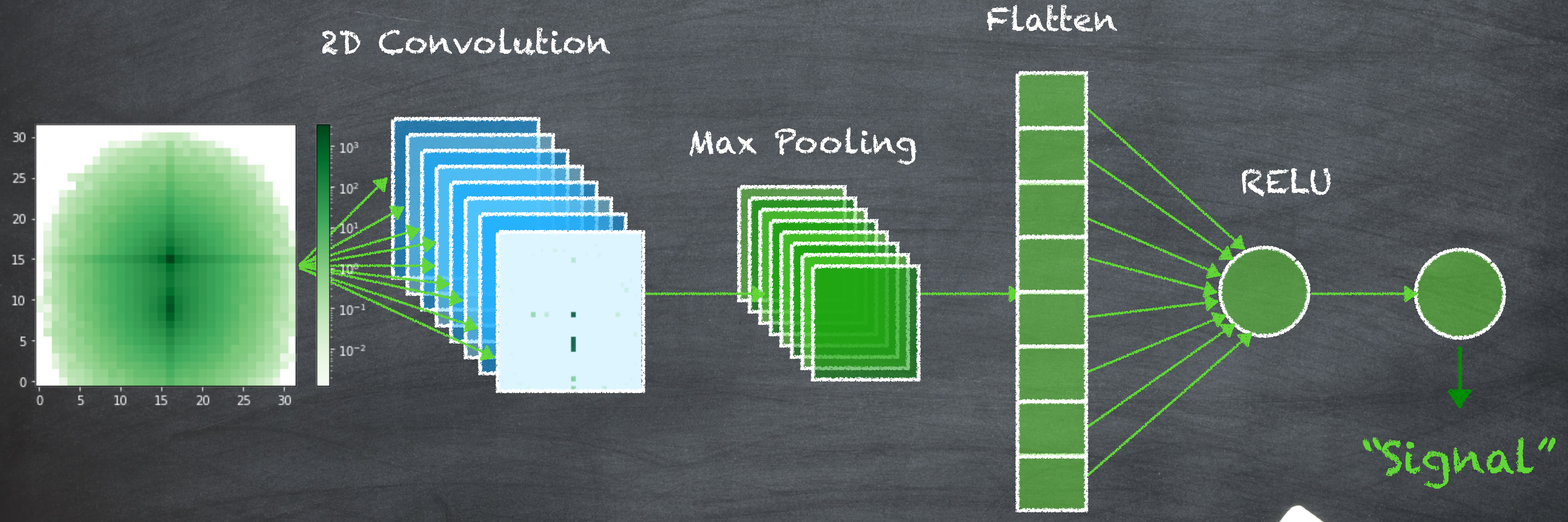


Toy Model - Forward Propagation

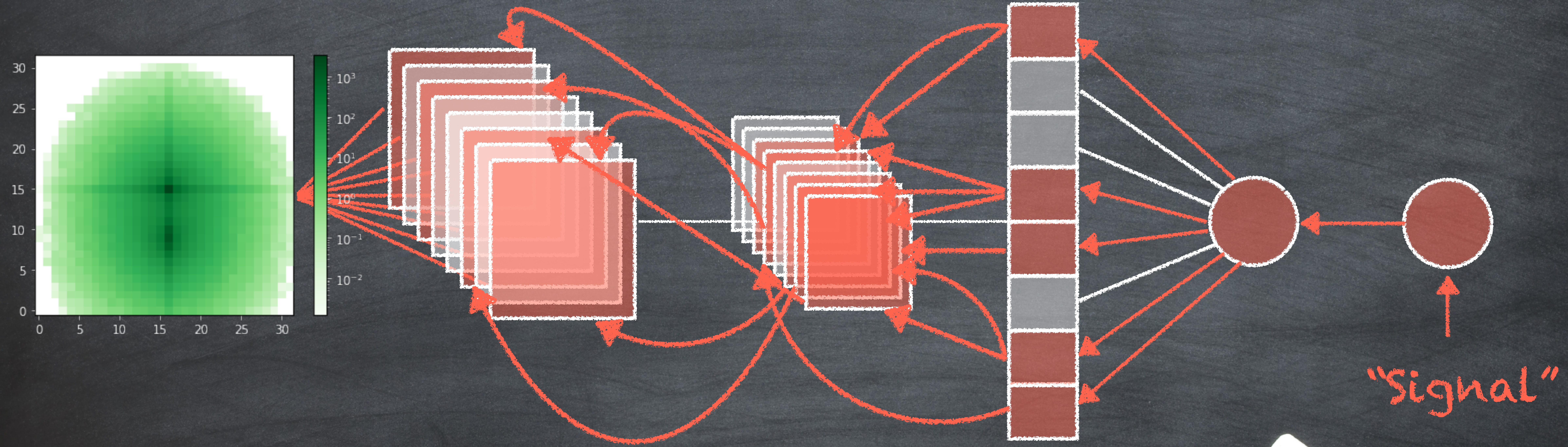



Signal Jet Images
Convolved Signal Jet Images

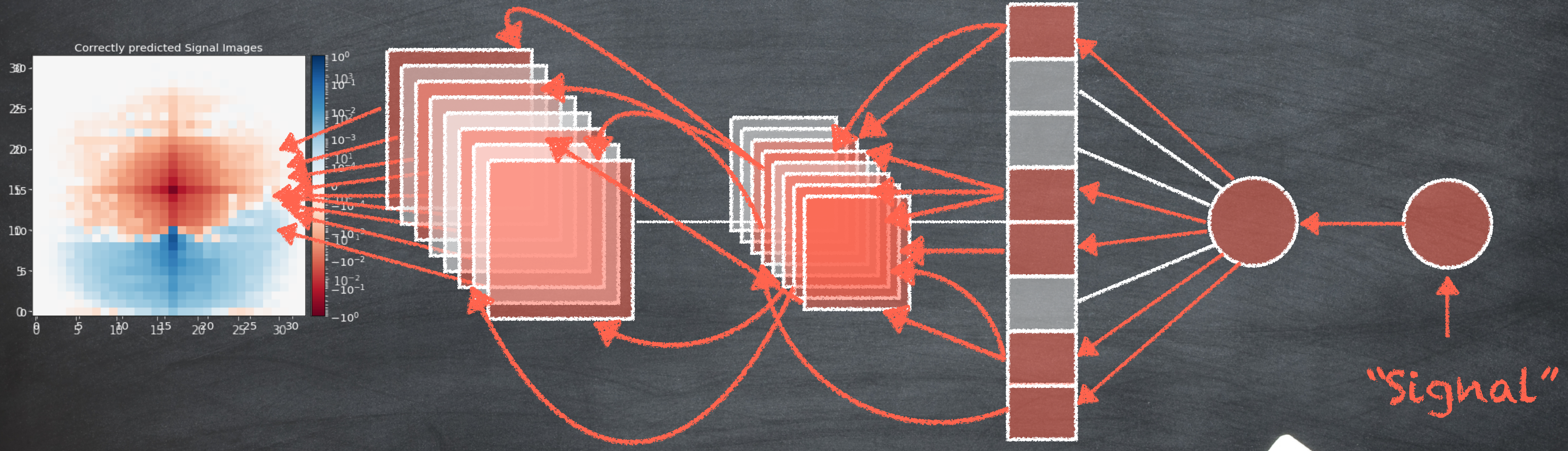
Toy Model – Forward Propagation

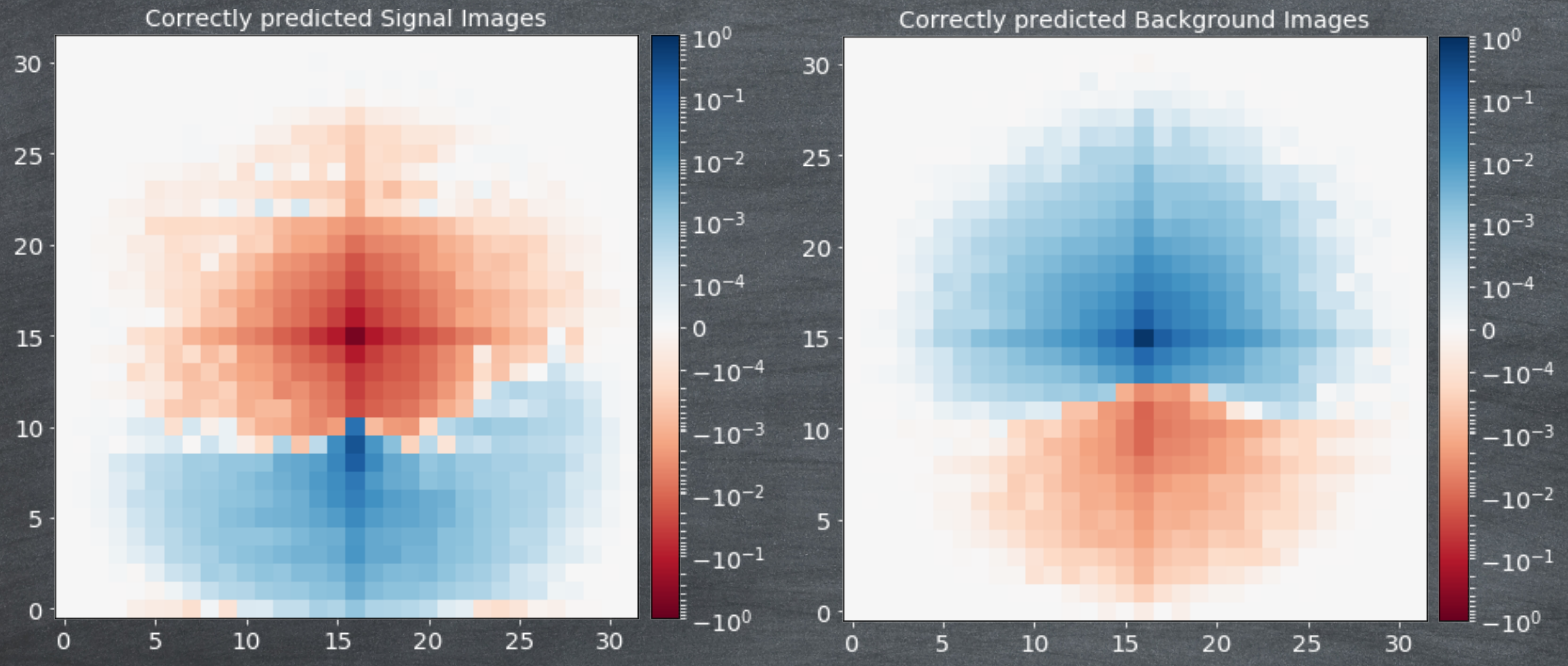


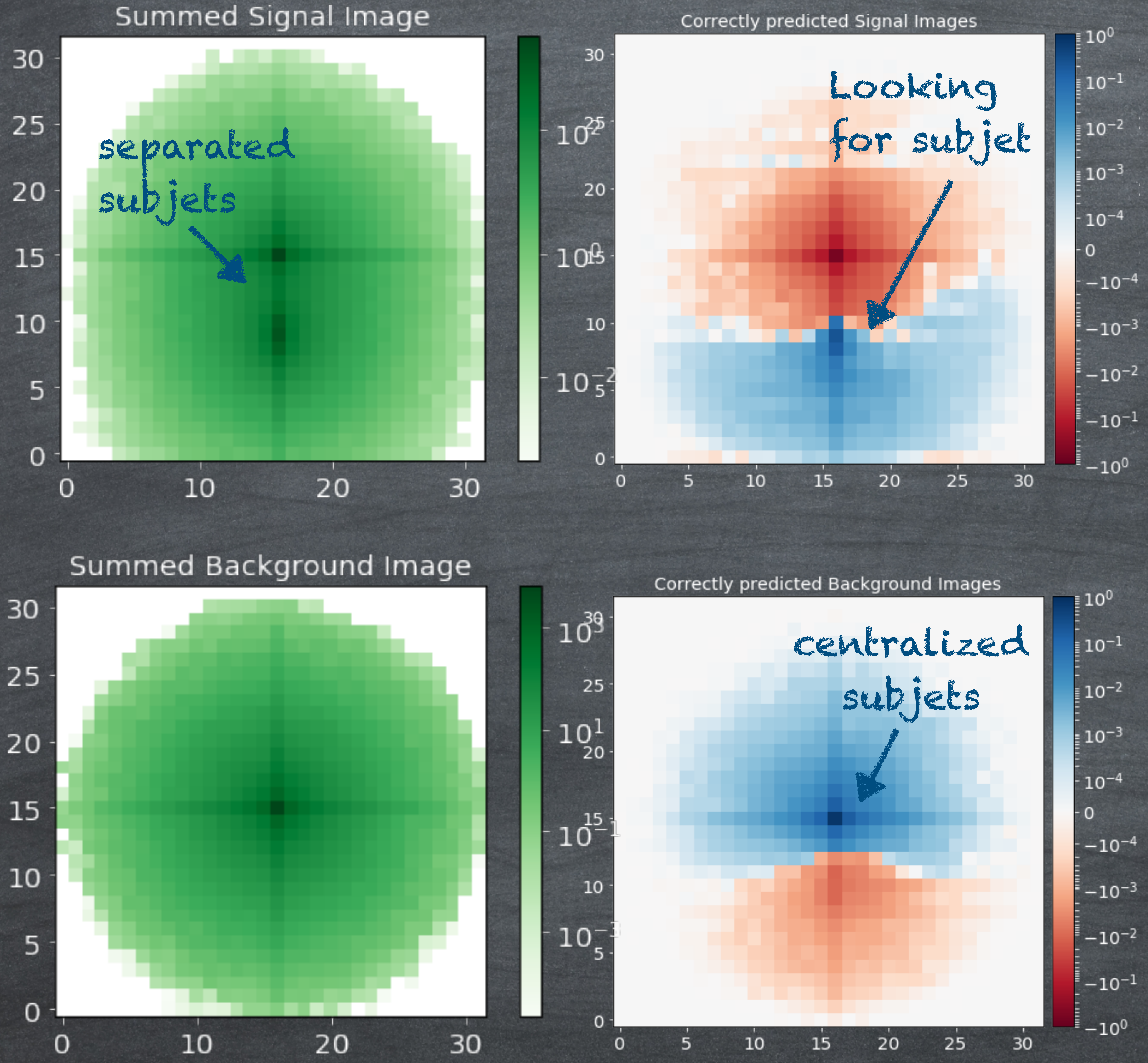
Toy Model – LRP

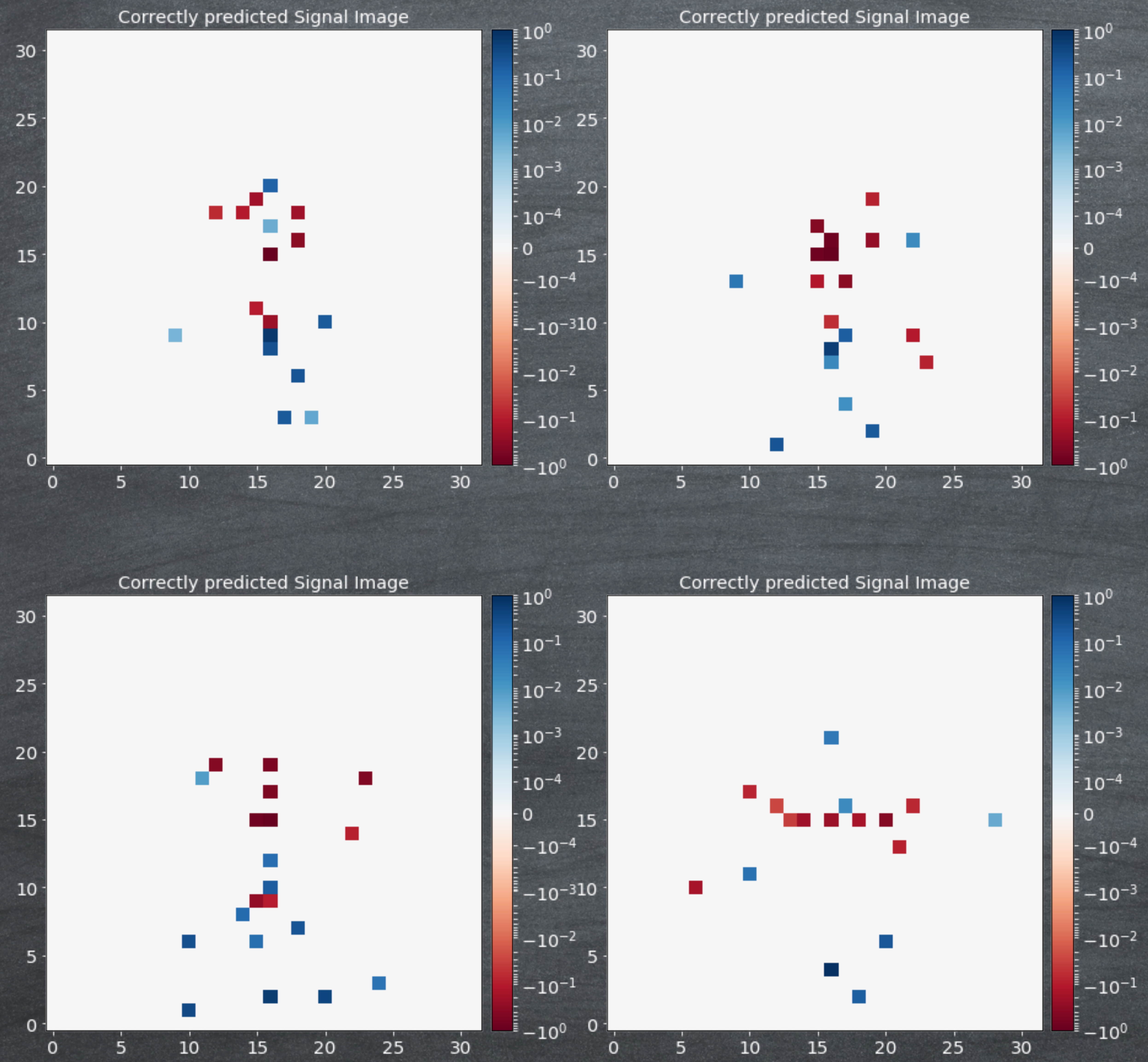


Toy Model - LRP



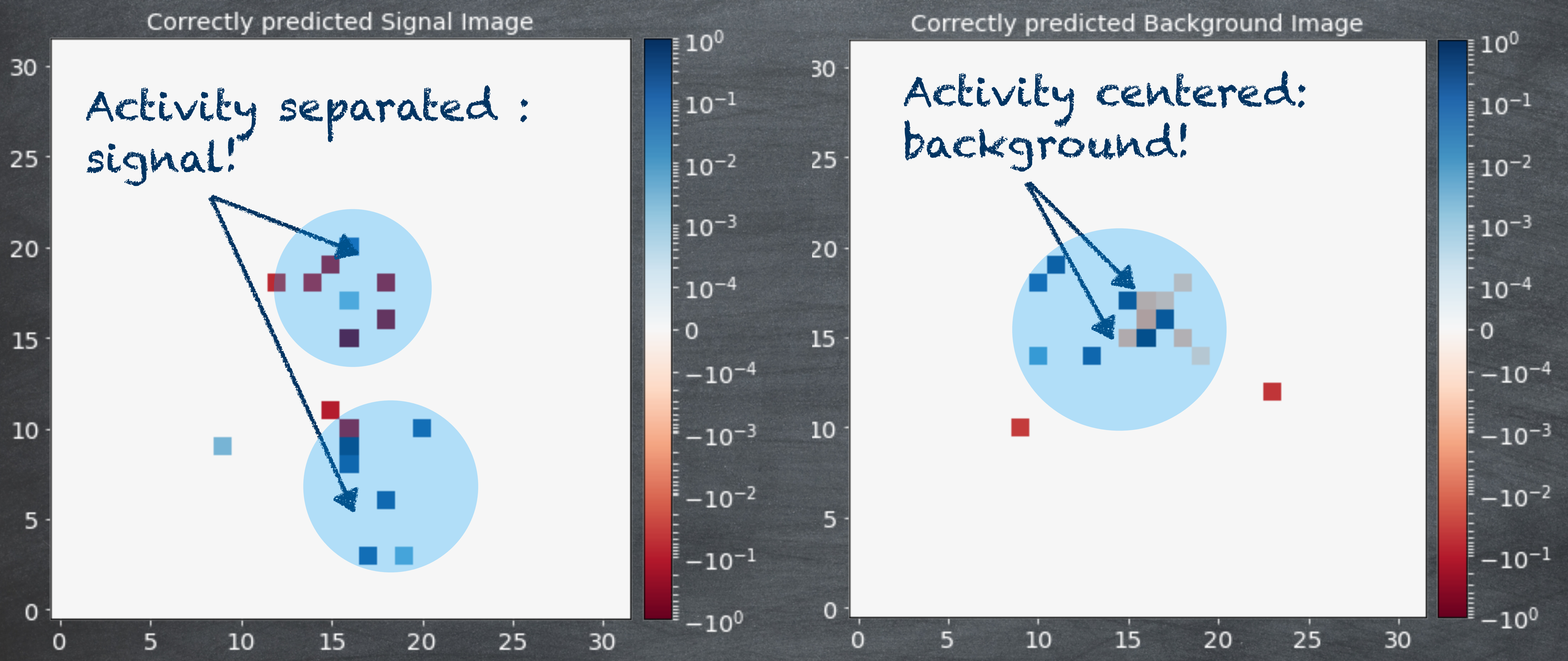

Correctly predicted Signal Images
Correctly predicted Background Images
0
5
10
15
20
25
30
10^0
10^-1
10^-2
10^-3
10^-4
0
-10^-4
-10^-3
-10^-2
-10^-1
-10^0


Summed Signal Image
separated subjets
Correctly predicted Signal Images
Looking for subjet
Summed Background Image
Correctly predicted Background Images
centralized subjets



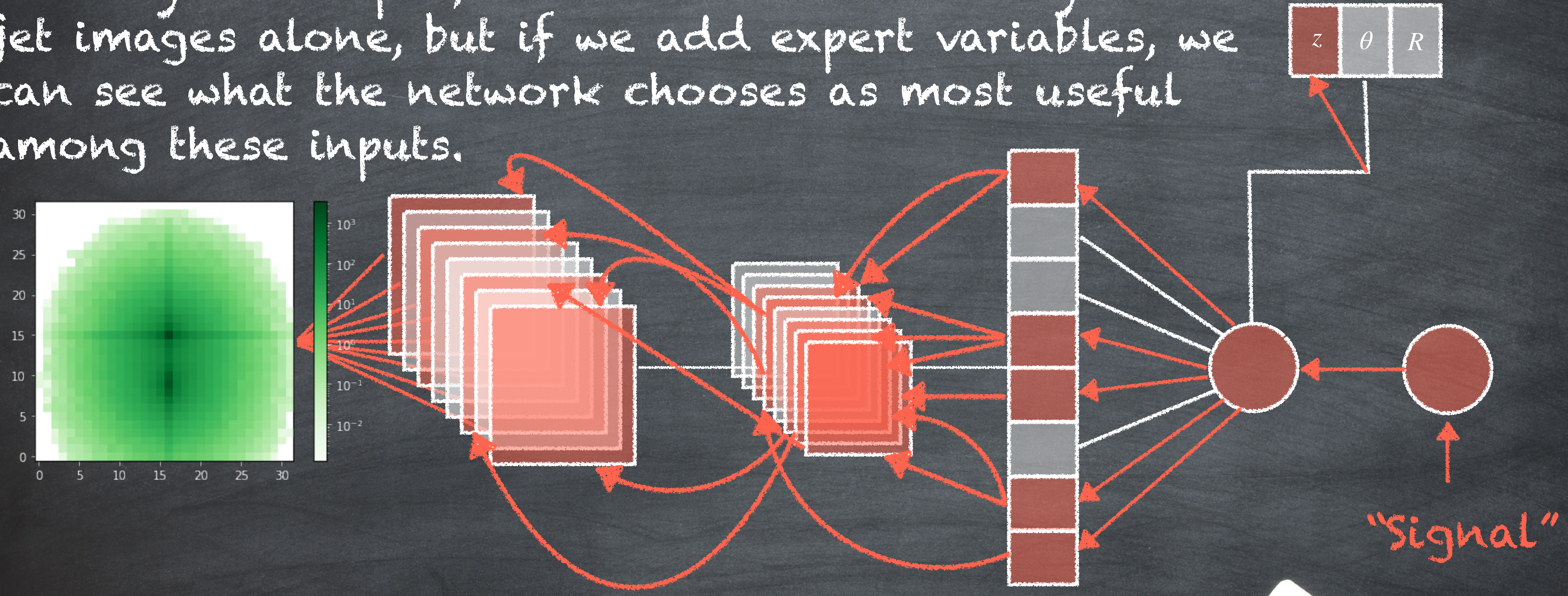
Correctly predicted Signal Image
Correctly predicted Signal Image
Correctly predicted Signal Image
Correctly predicted Signal Image

Toy Model - Single Events

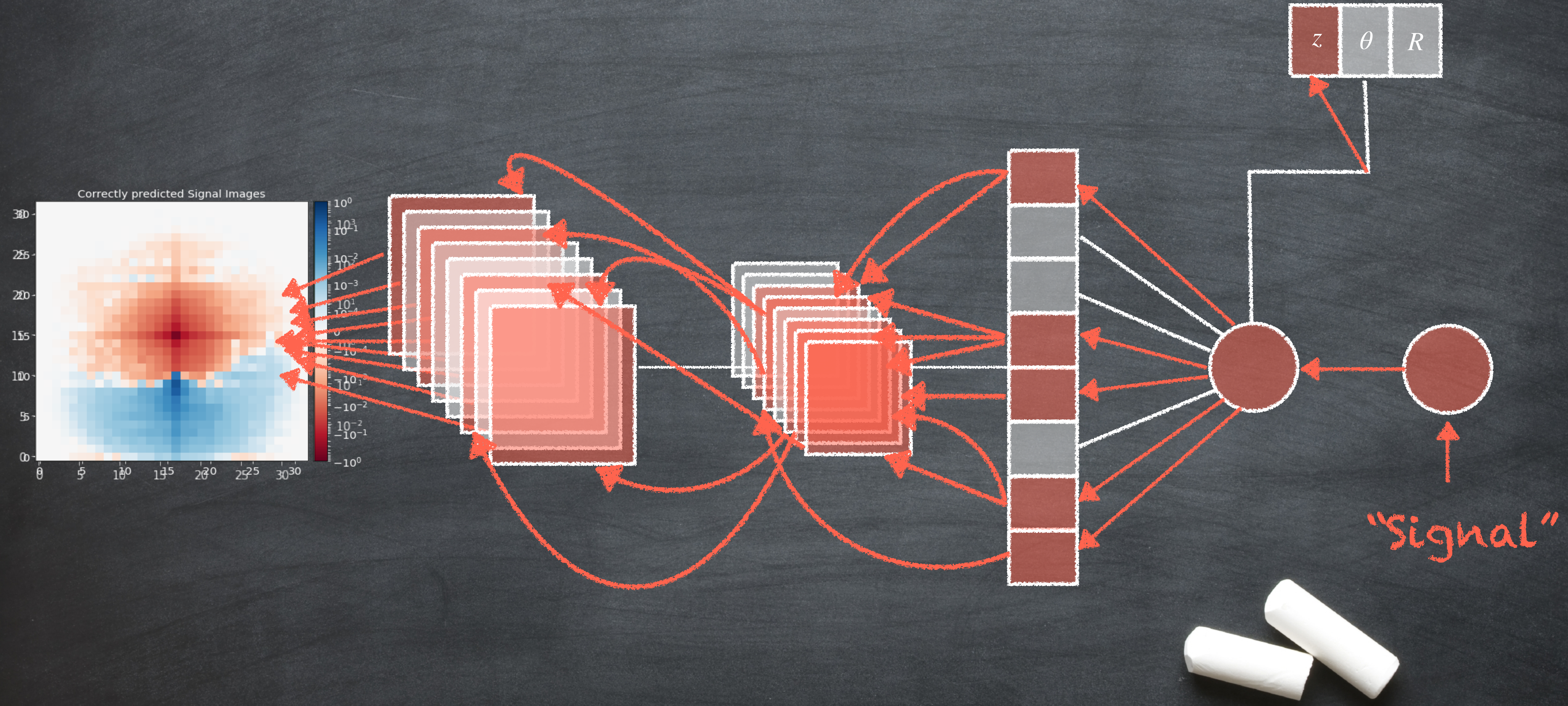


Toy Model w/ expert variables - LRP

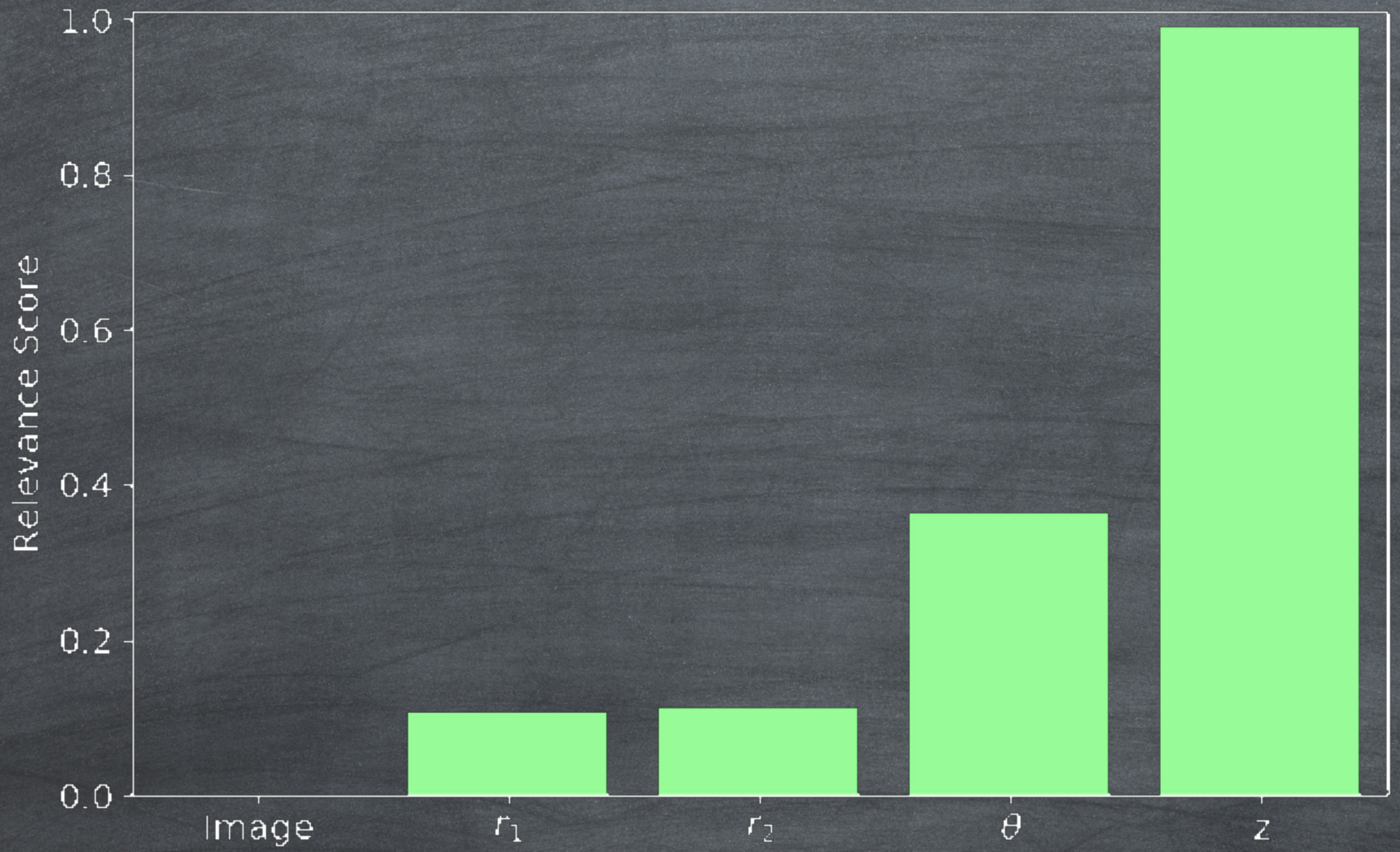
This toy model performs with 100% accuracy on the jet images alone, but if we add expert variables, we can see what the network chooses as most useful among these inputs.



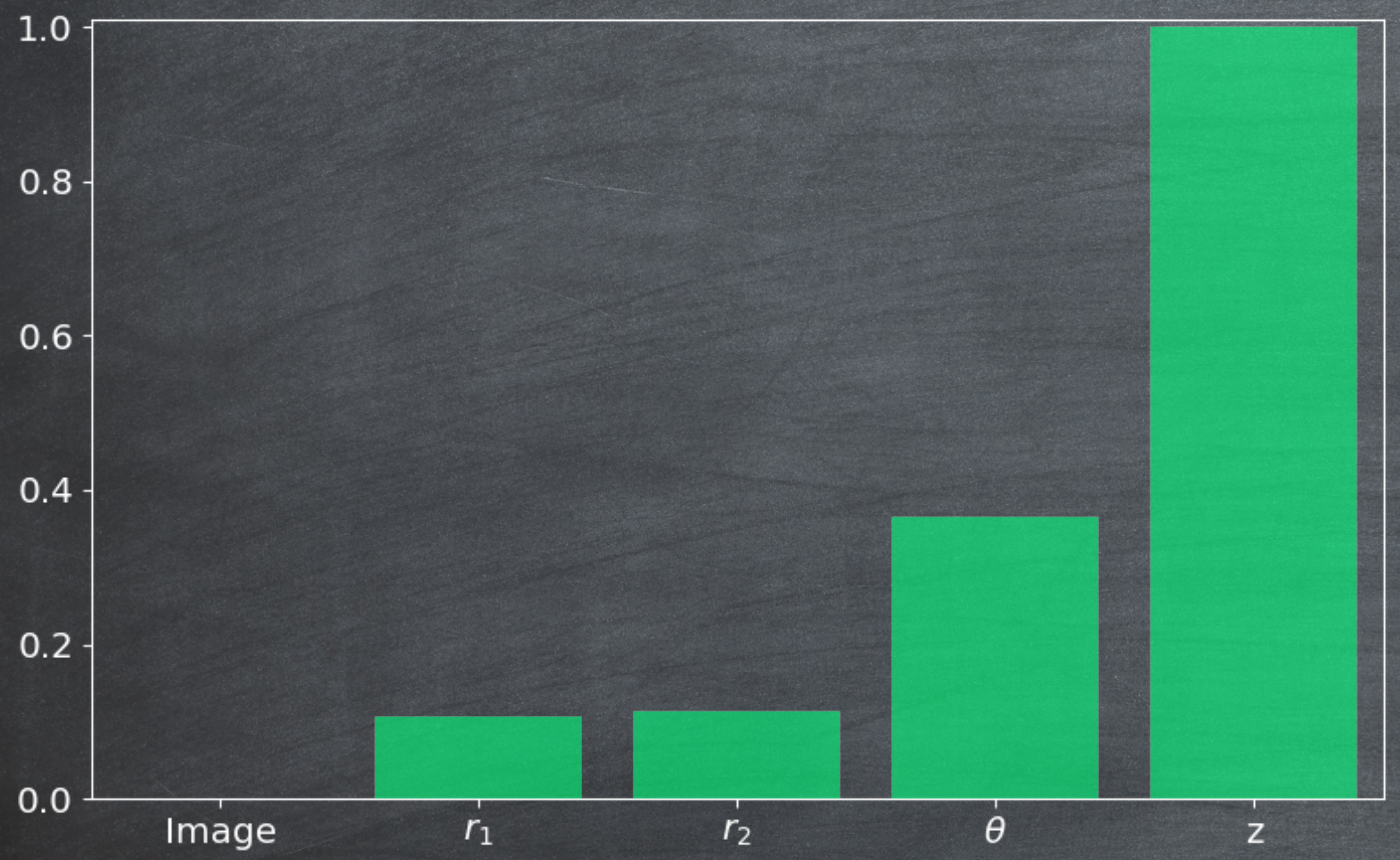
Toy Model w/ expert variables - LRP

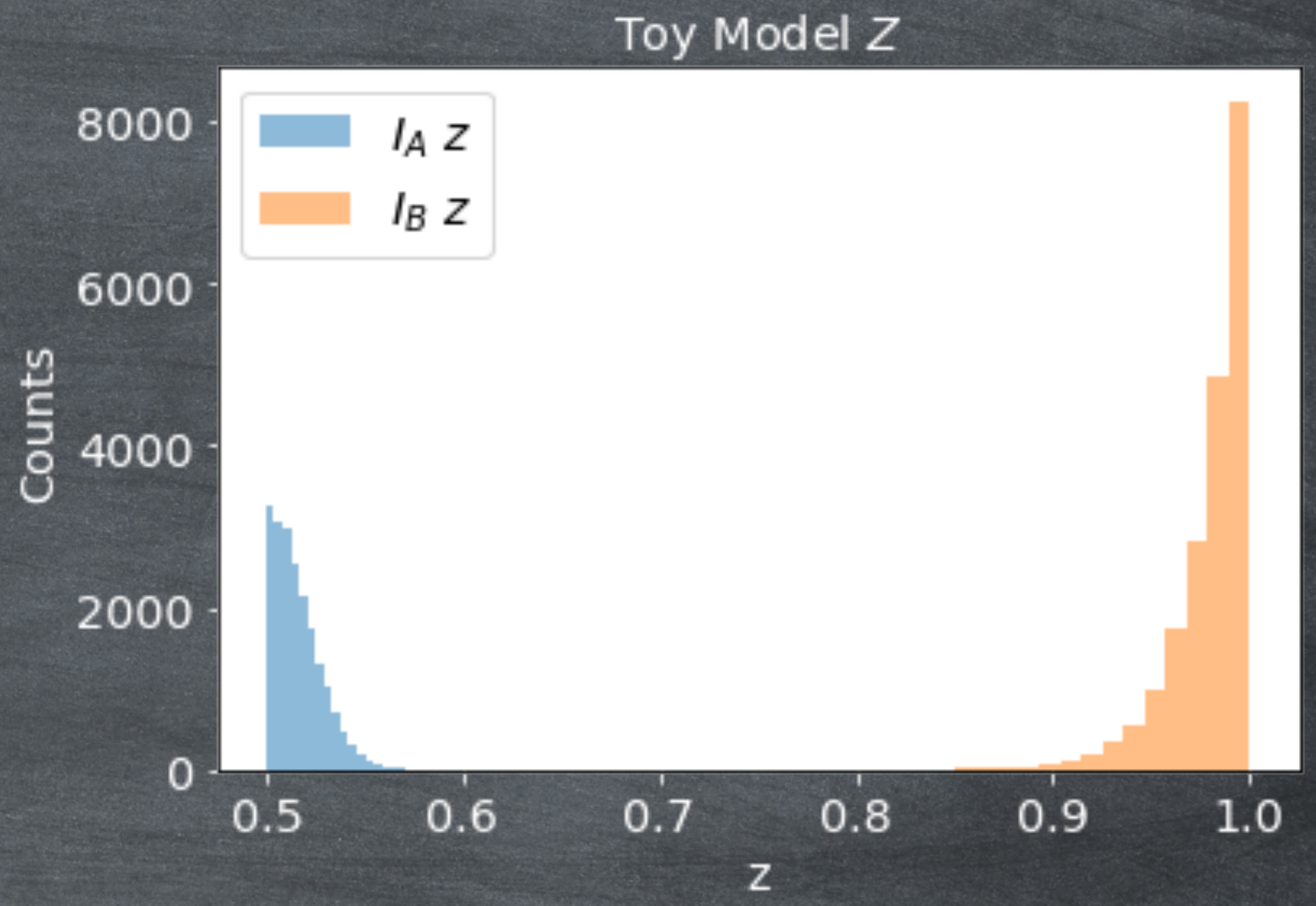


Toy Model – LRP for XAUG Variables



Toy Model - LRP for XAUG Variables

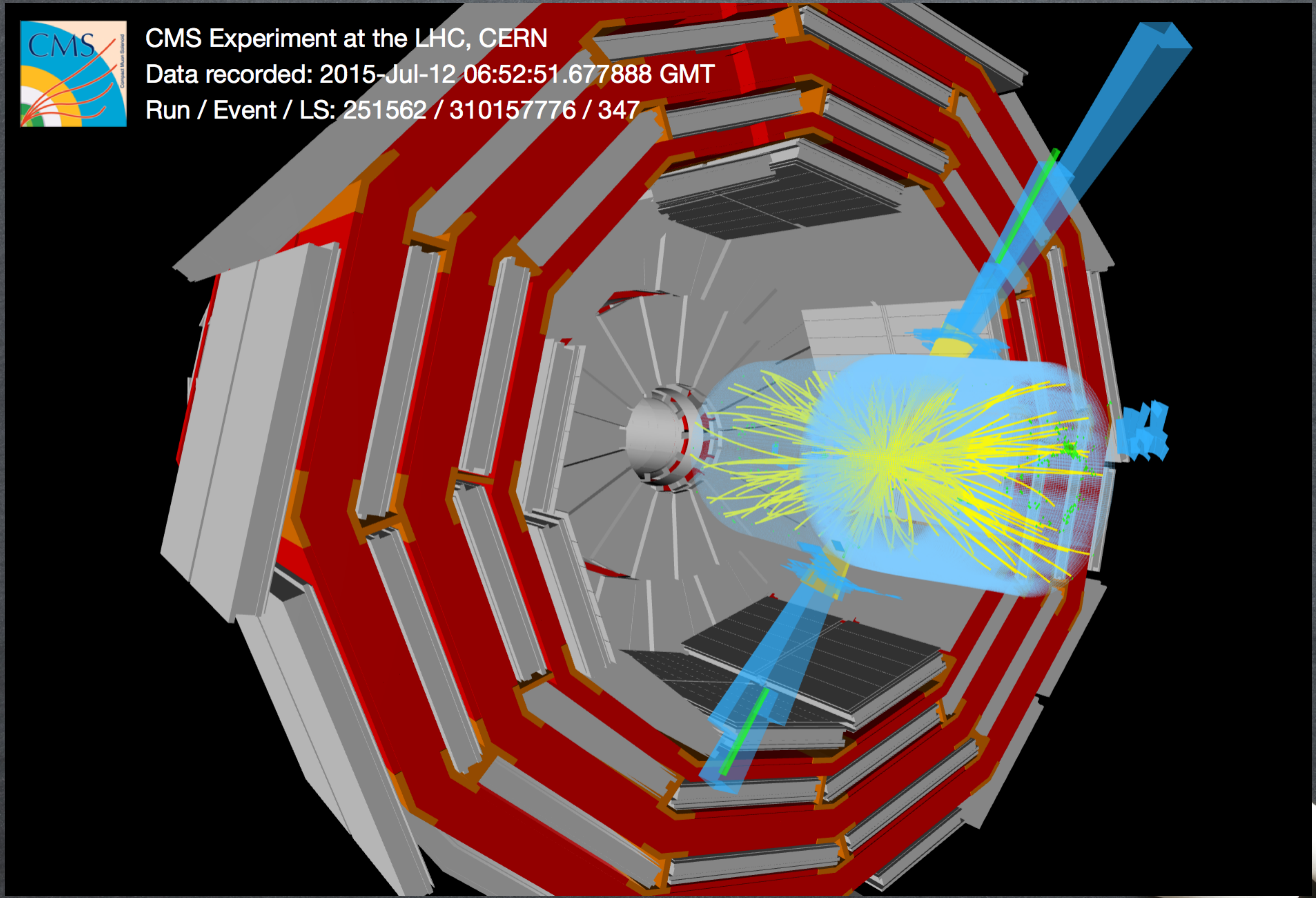




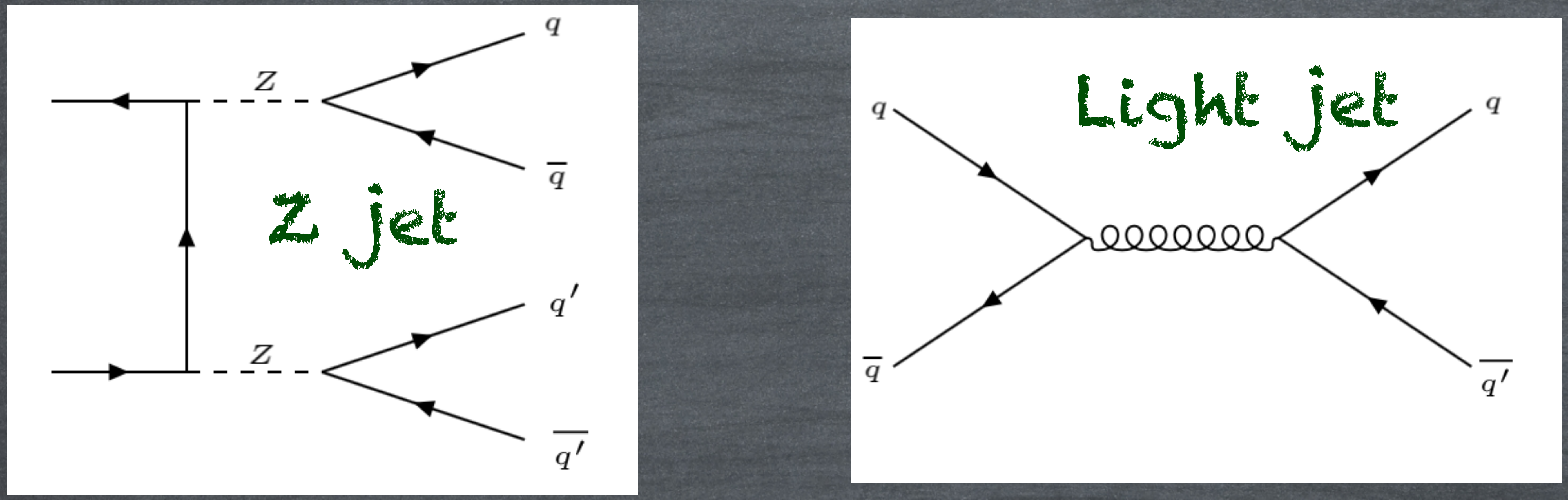
We see that the z input, that with the greatest separation between the toy "signal" and "background", has the greatest relevance to the XAUG toy model.

Pythia Simulation

Pythia Simulation



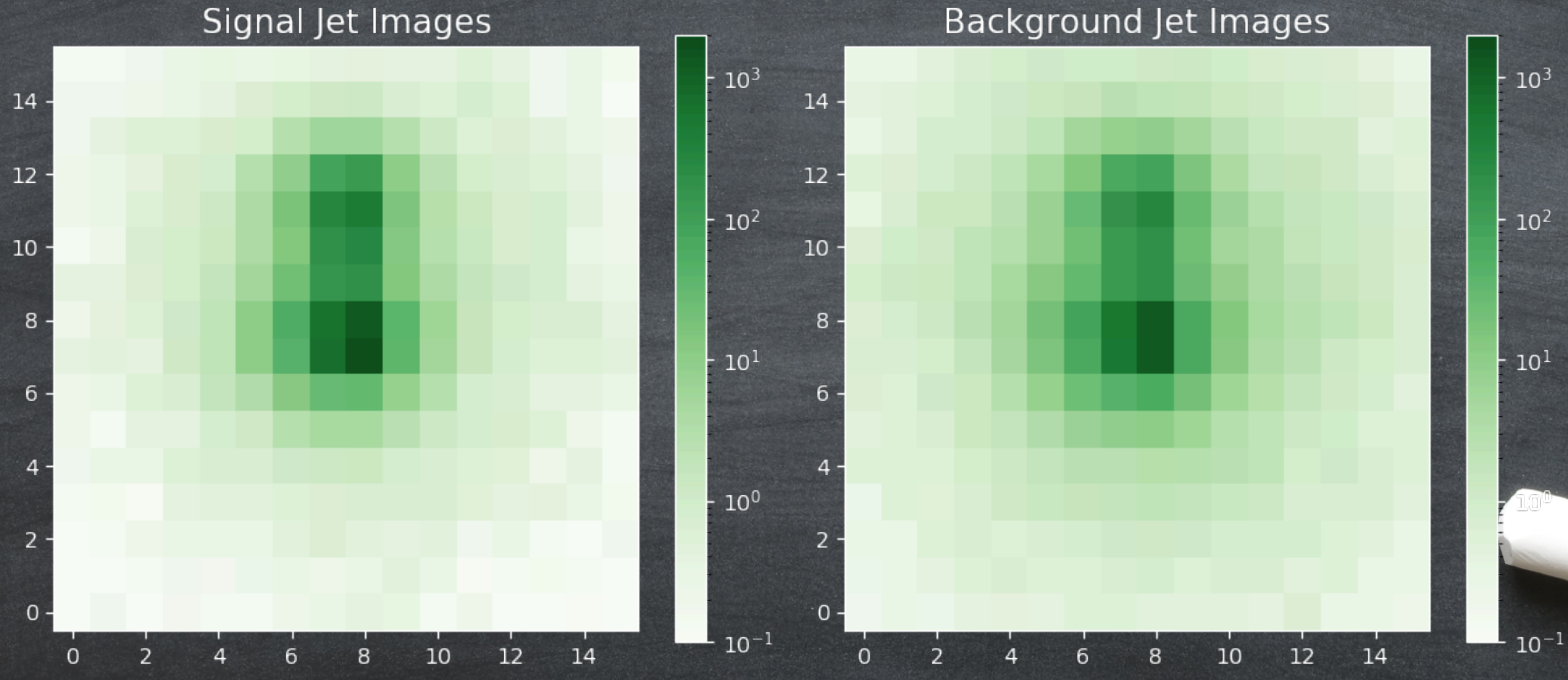
Pythia Simulation



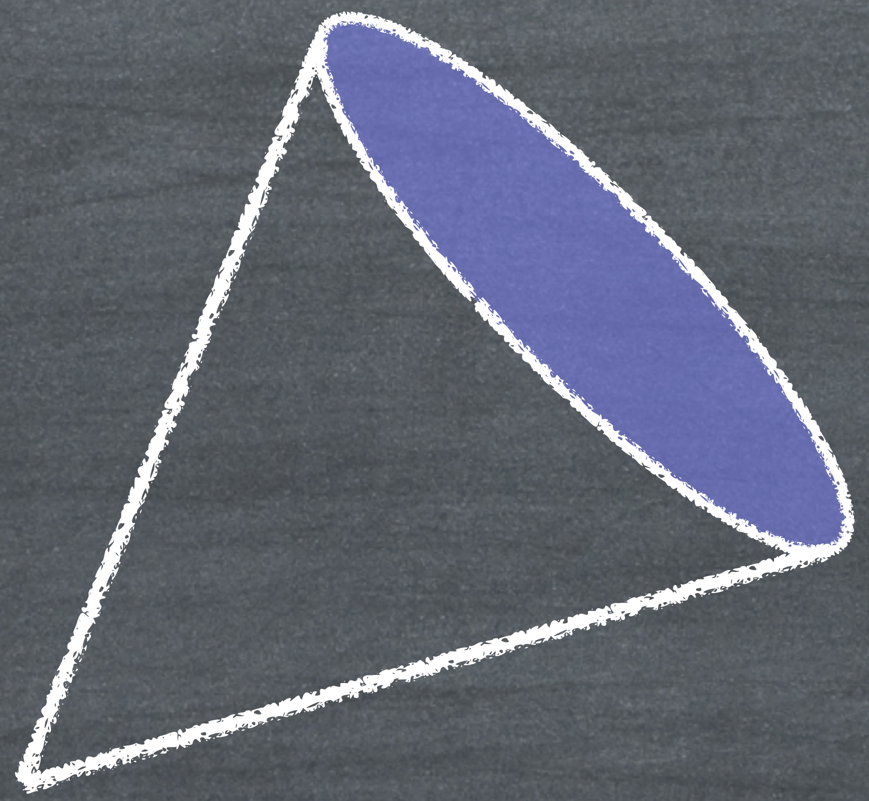
Simulated with pythia8,

AK8 jets from fastjet
Pt > 200 GeV

N-subjettiness from
fastjet-contrib:
WTA KT axis
Normalized



Pythia Simulation - Preprocessing



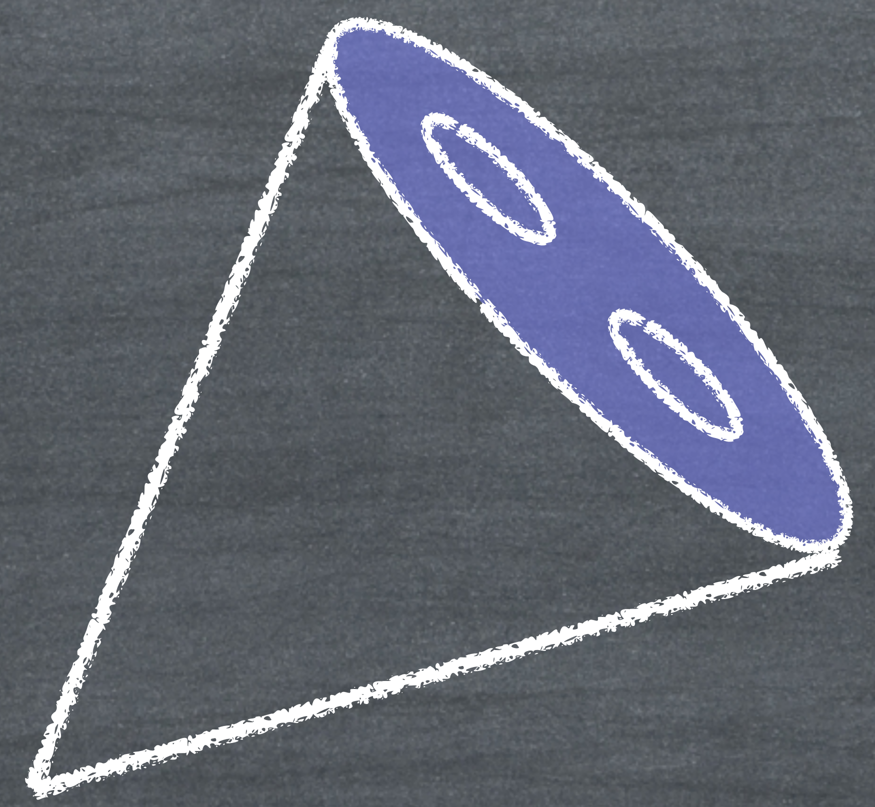
Compute AK8 jets

Simulated with pythia8
SM ZZ and QCD

AK8 jets from fastjet
Pt > 200 GeV

N-subjettiness from
fastjet-contrib:
WTA KT axis
Normalized

Pythia Simulation - Preprocessing



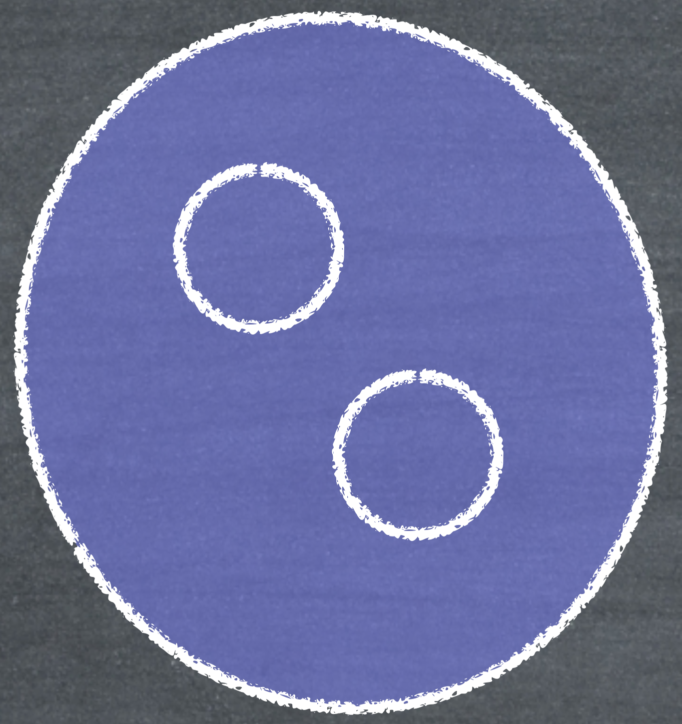
Calculate subjets

mMDT / soft drop

$\beta = 0$ with $z_{cut} = 0.1$

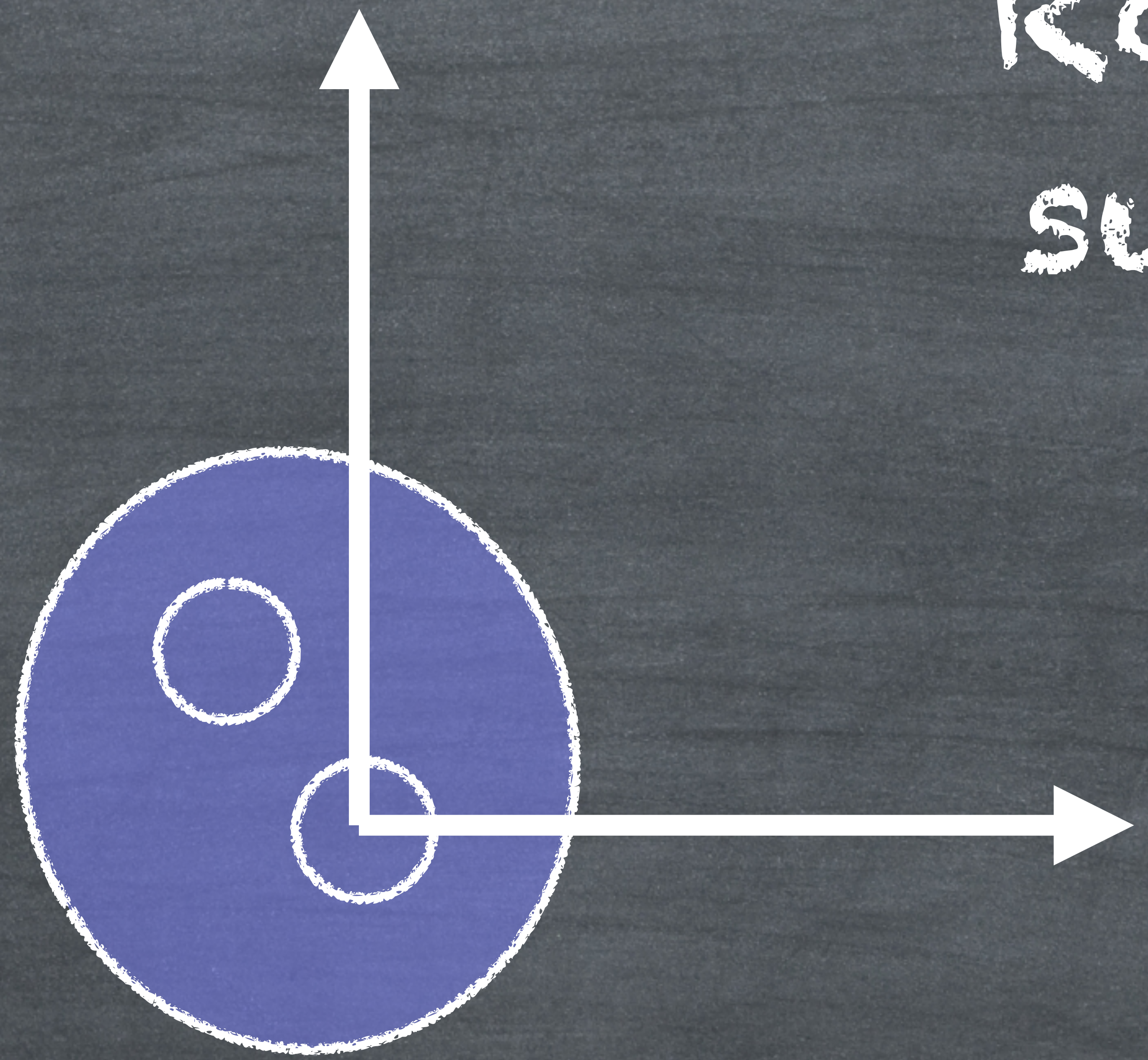
Pythia Simulation - Preprocessing

Project to 2d plane

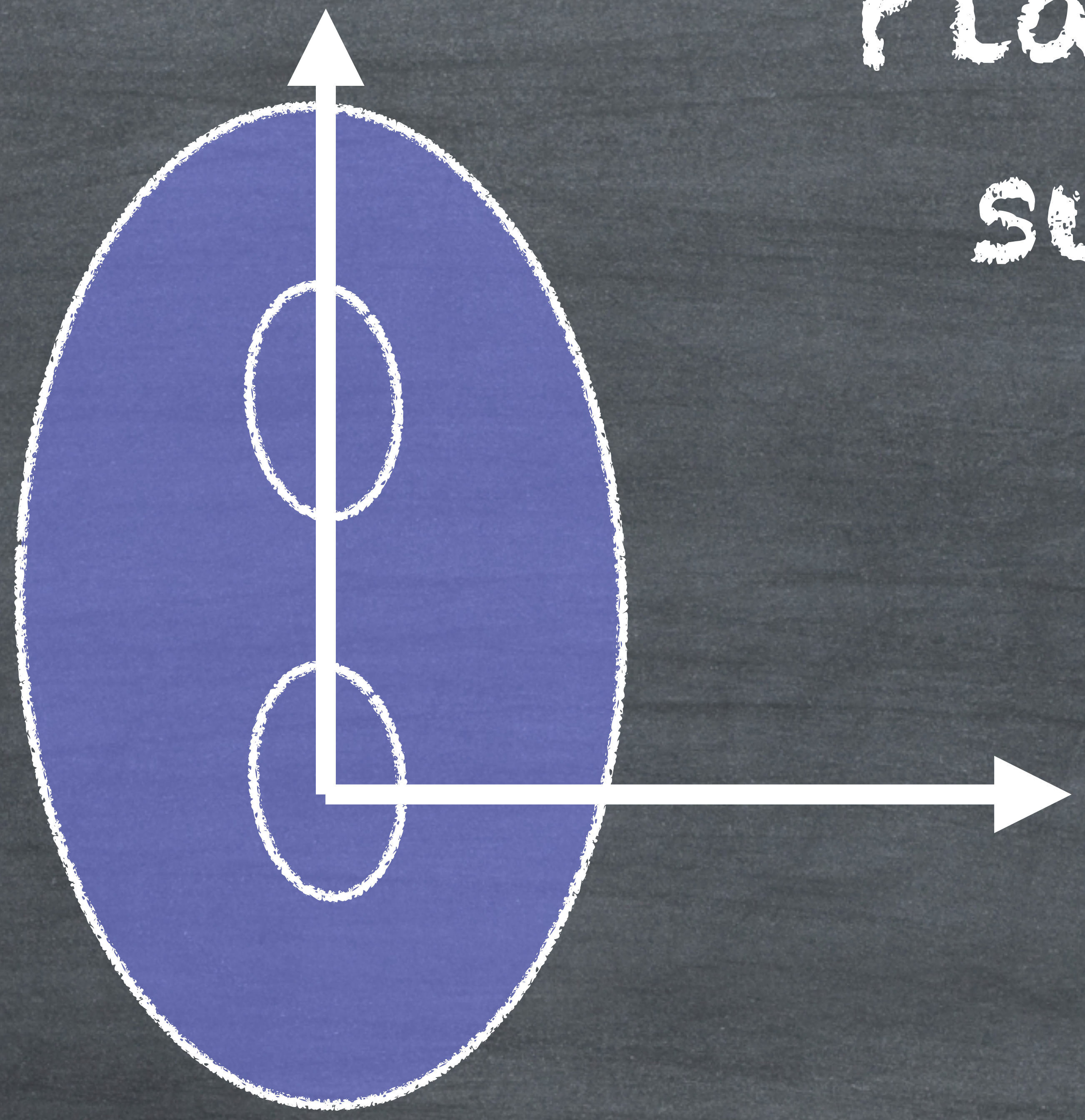


Pythia Simulation - Preprocessing

Rotate leading
subjet to (0,0)

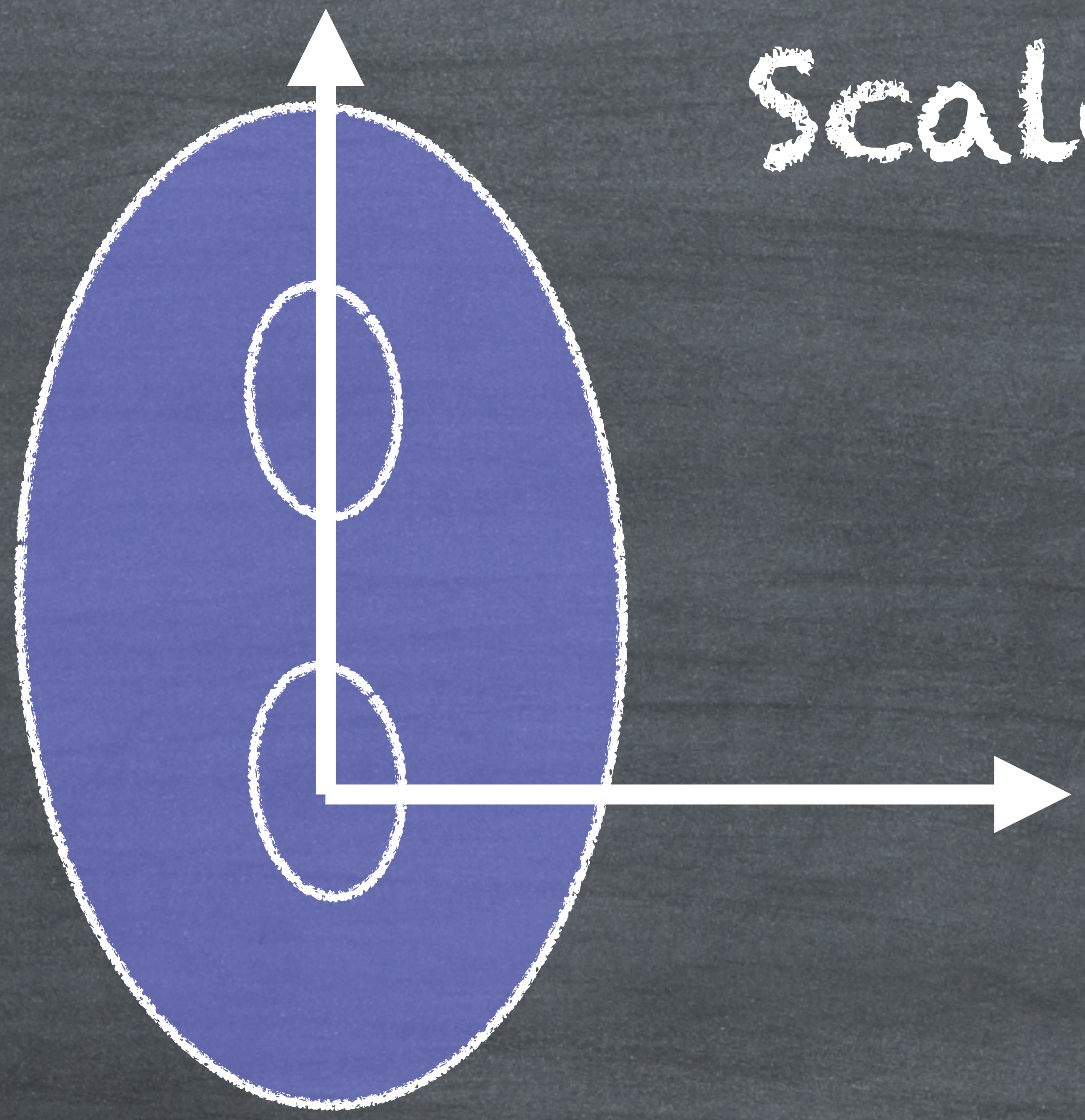


Pythia Simulation - Preprocessing



Place subleading subjet at (0,1)

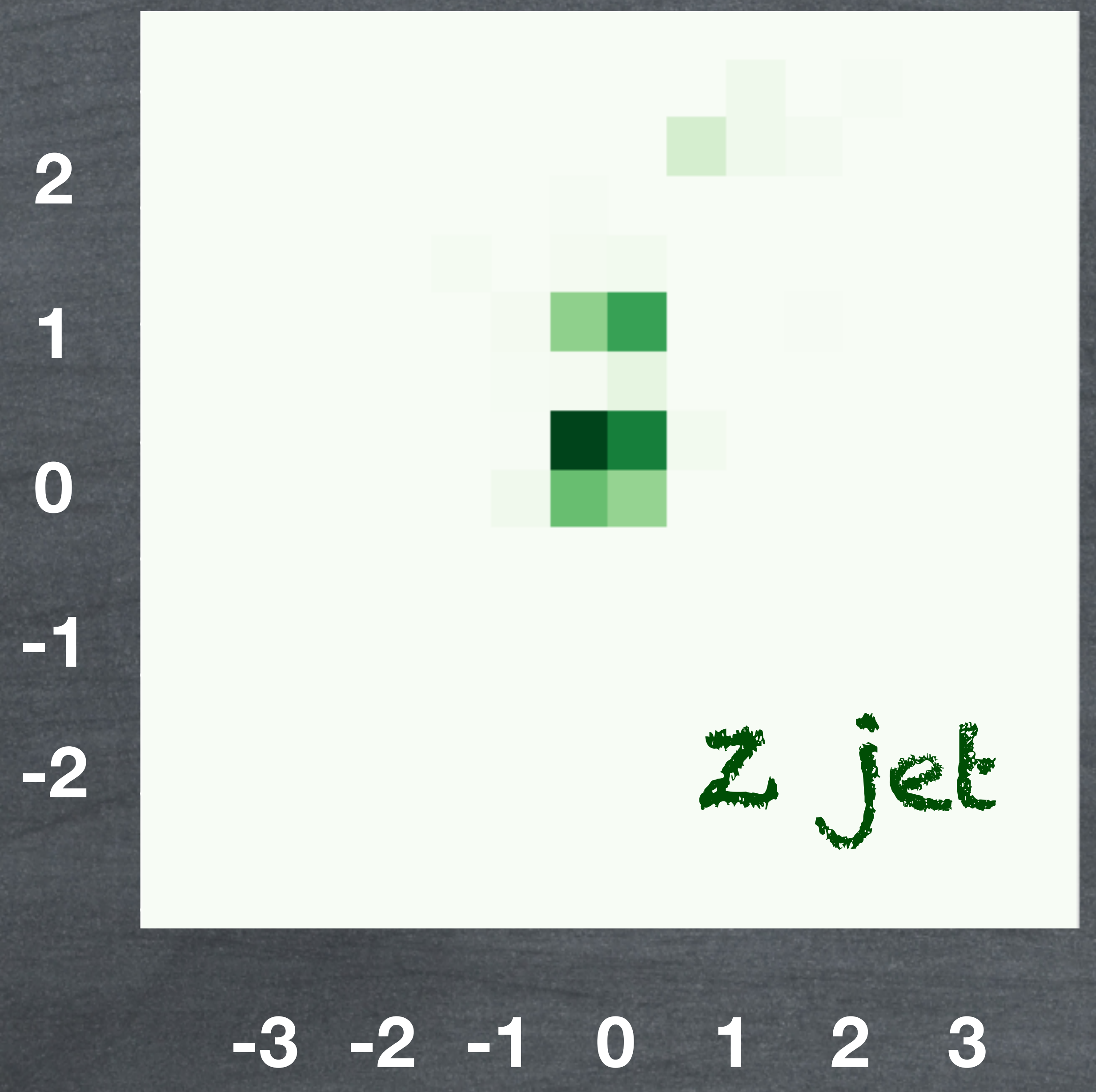
Pythia Simulation - Preprocessing



Scale intensities as

$$p_T/p_{T;jet}$$

Pythia Simulation - Preprocessing



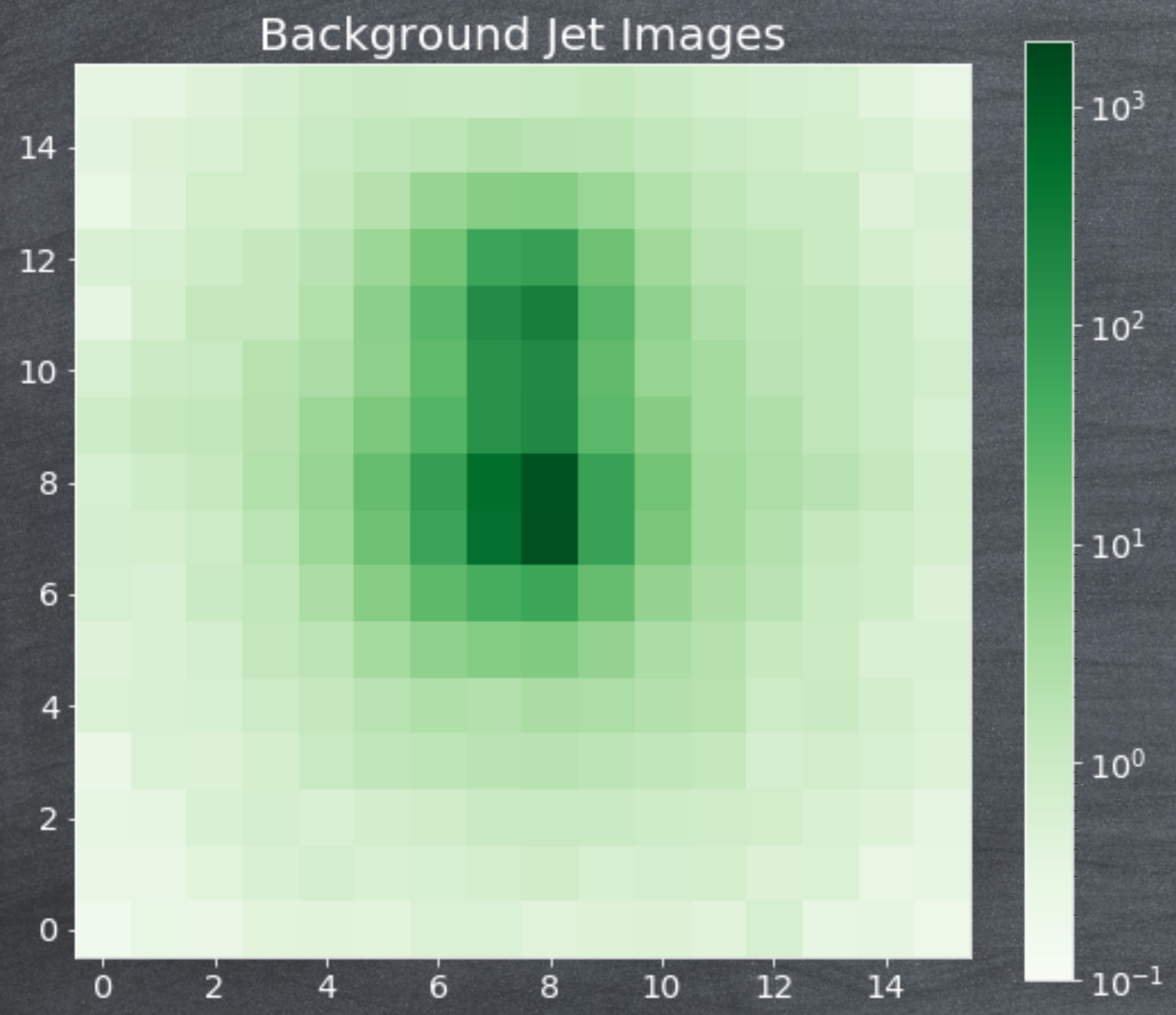
Tada!

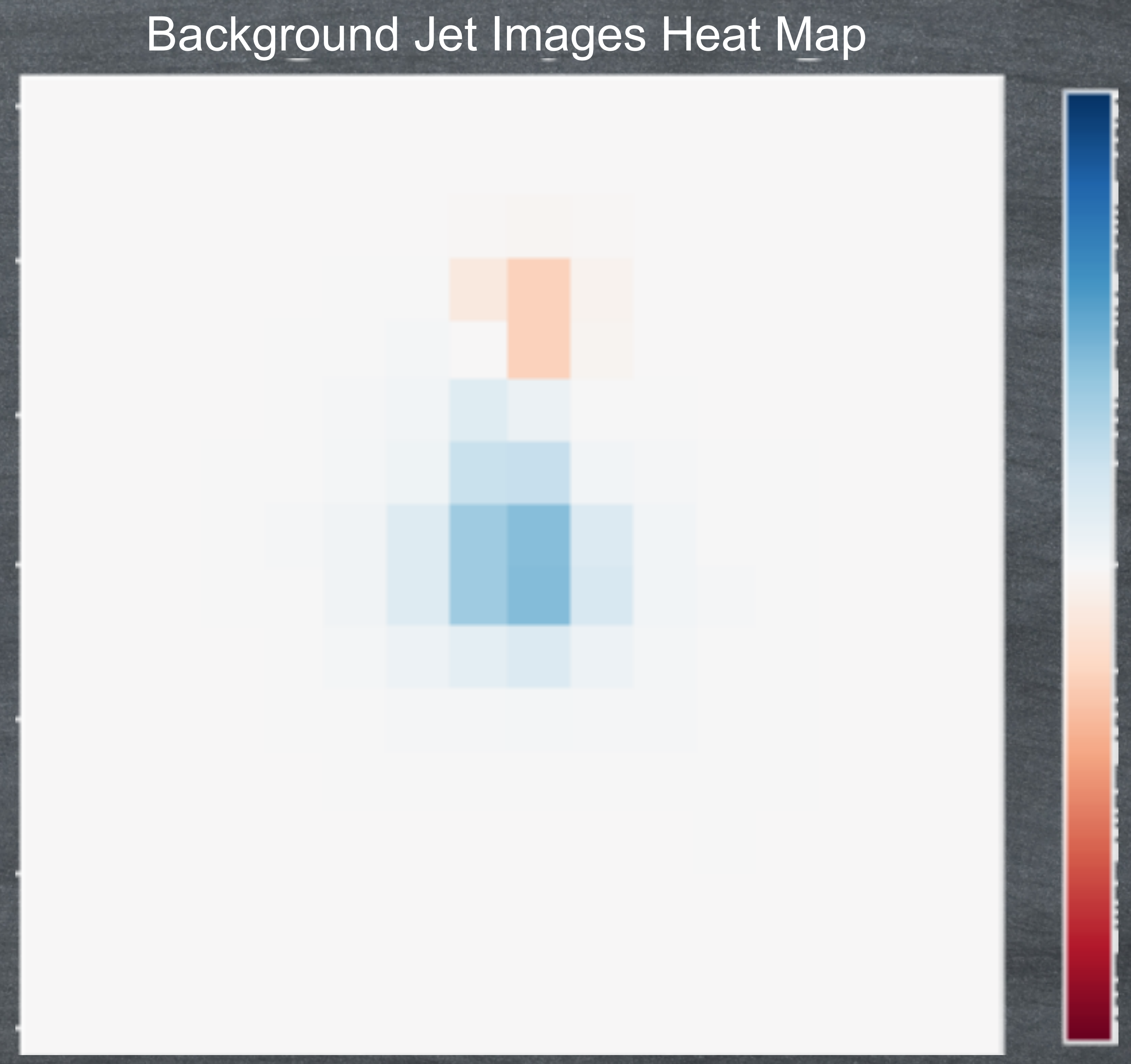
Pythia Simulation – Forward Propagation

Pythia Simulation - LRP

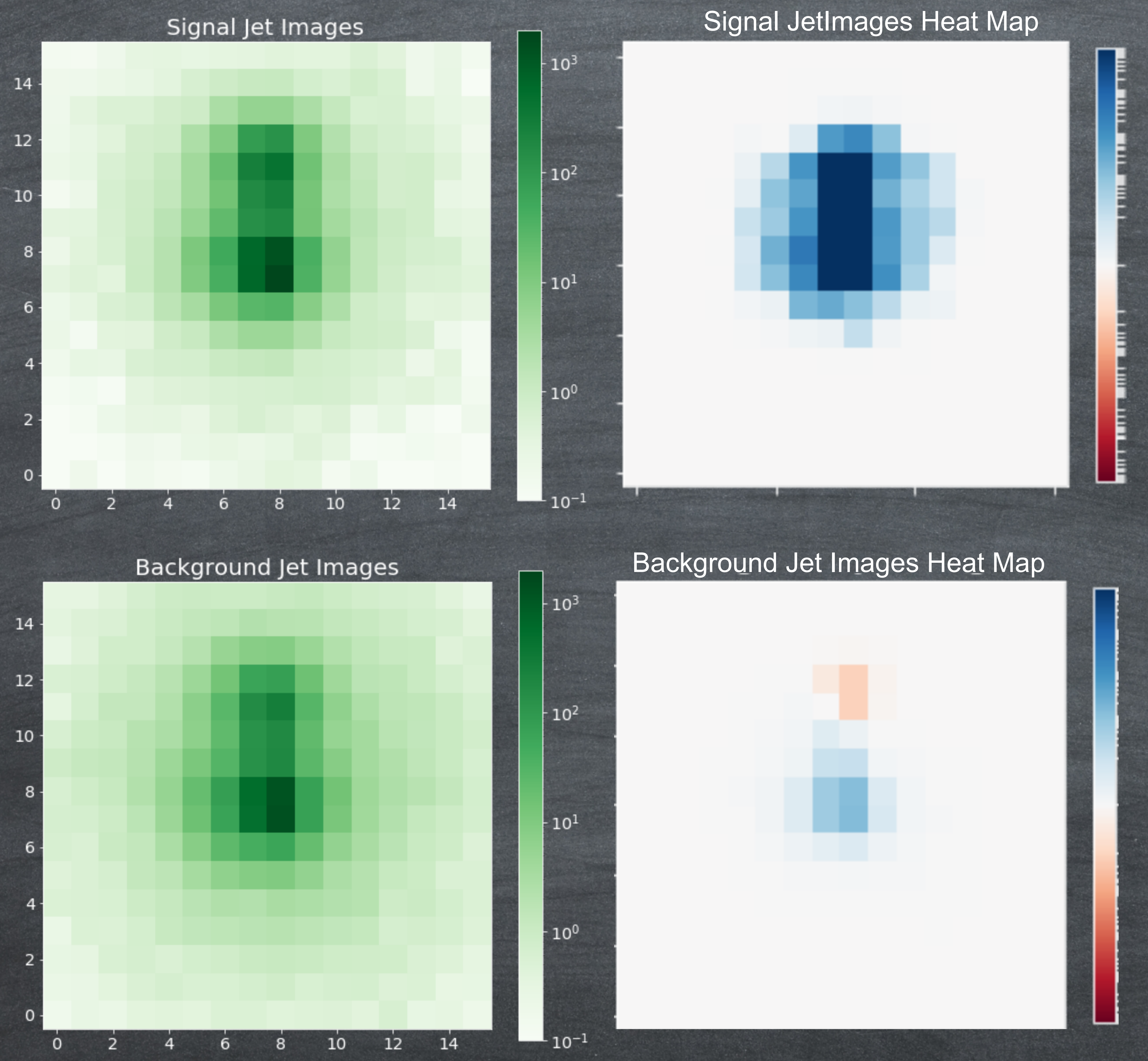
Pythia Simulation - LRP

Pythia Simulation – LRP

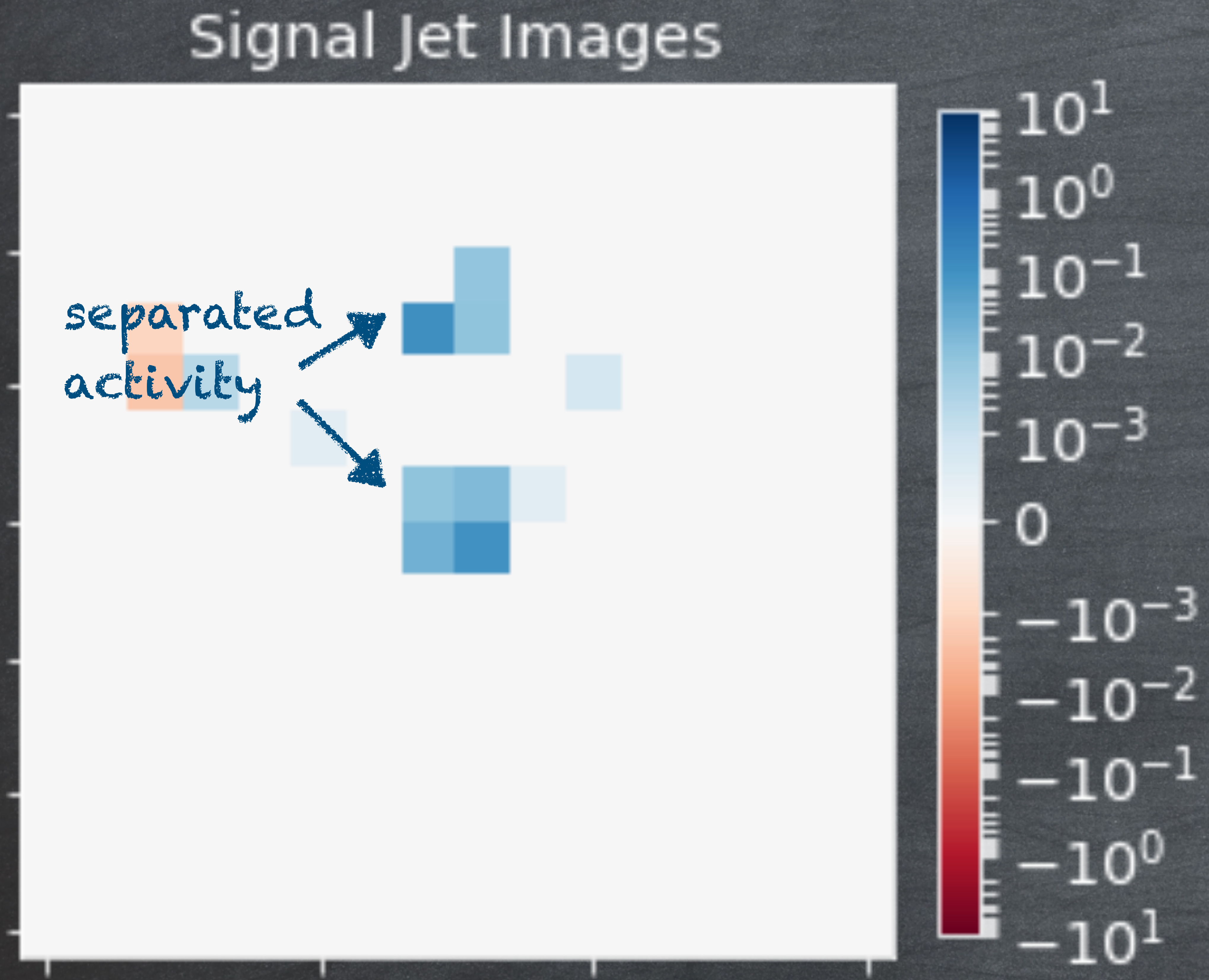


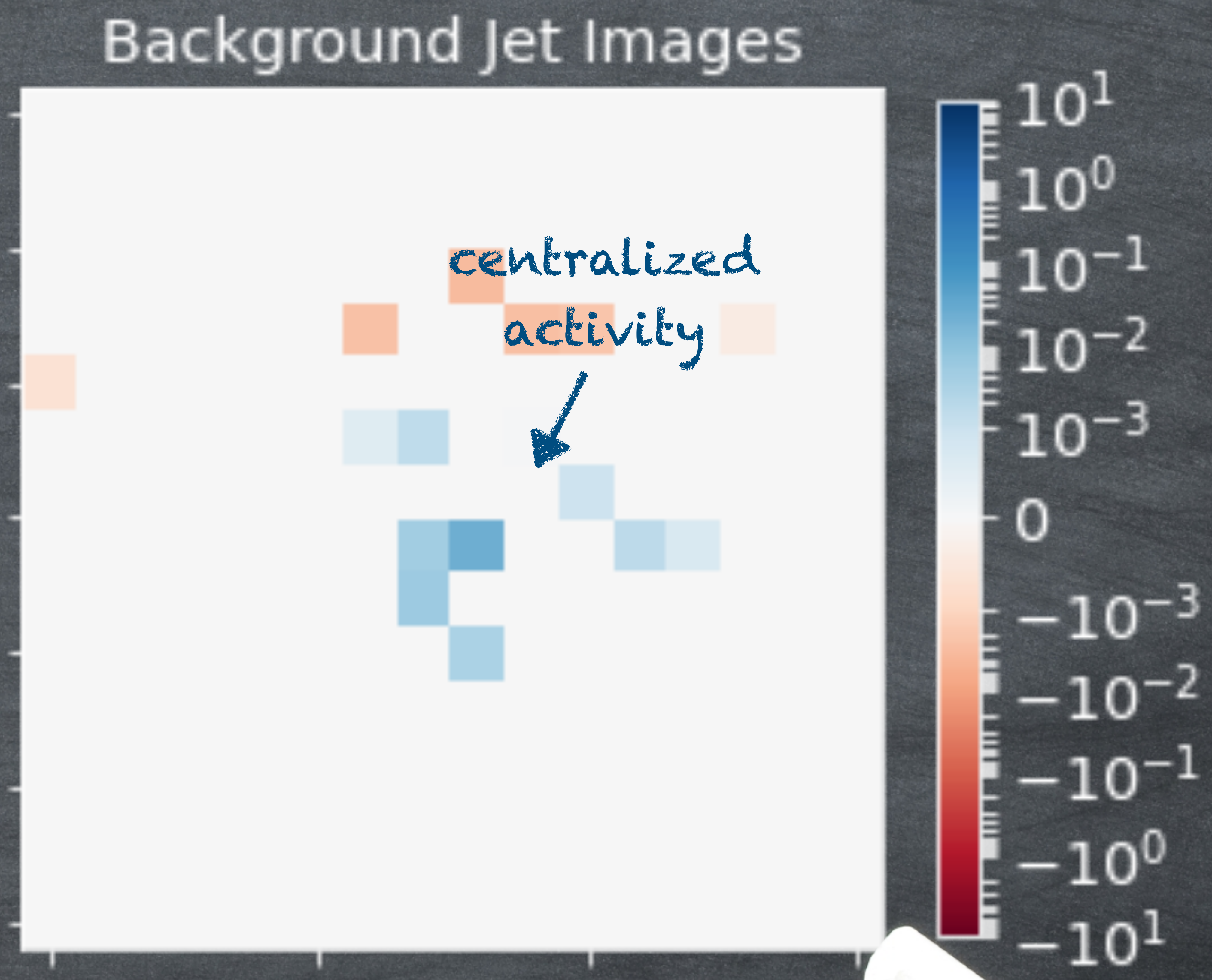


Pythia Simulation - LRP



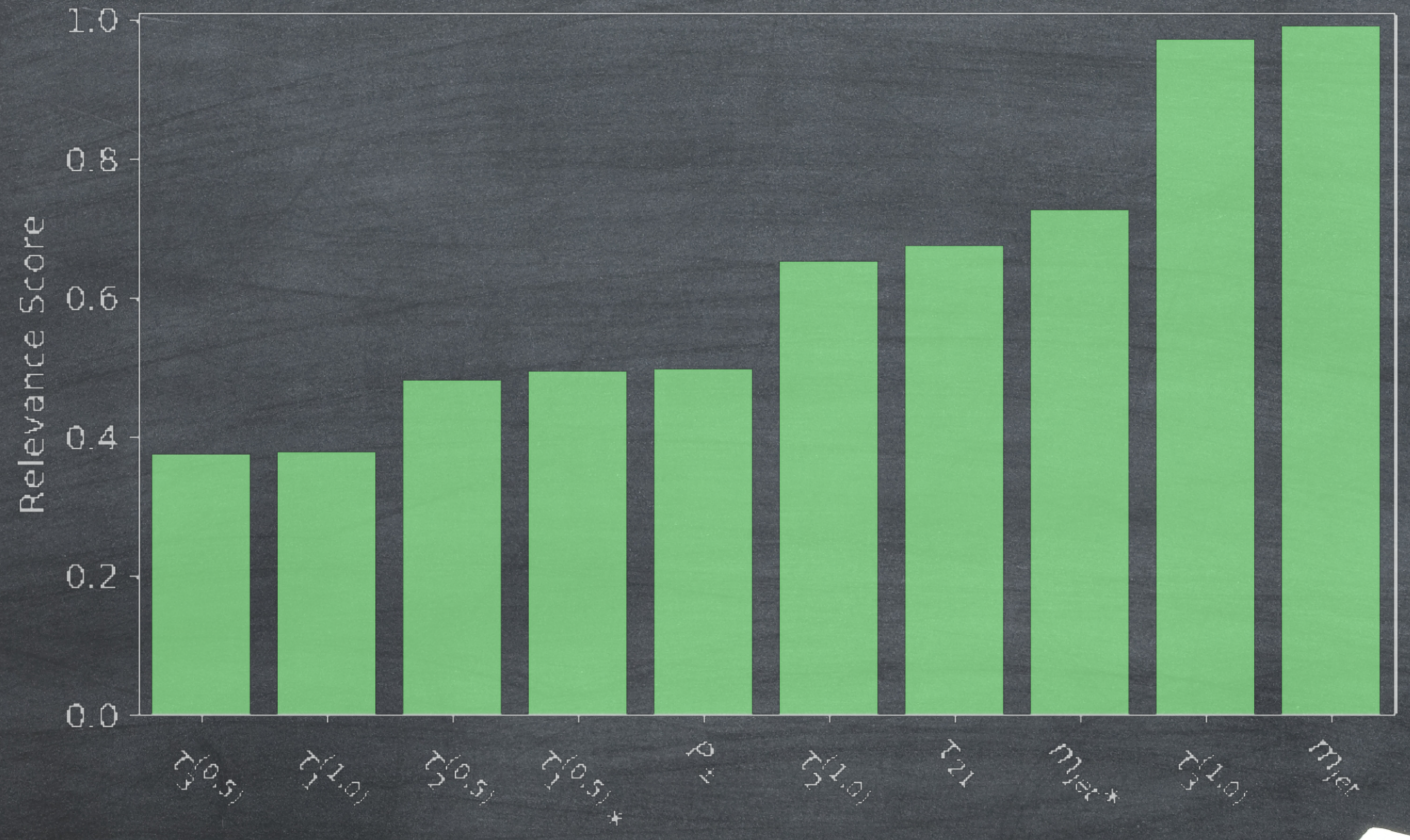
Pythia Simulation - Single Events





Pythia Simulation – LRP

XAUGS

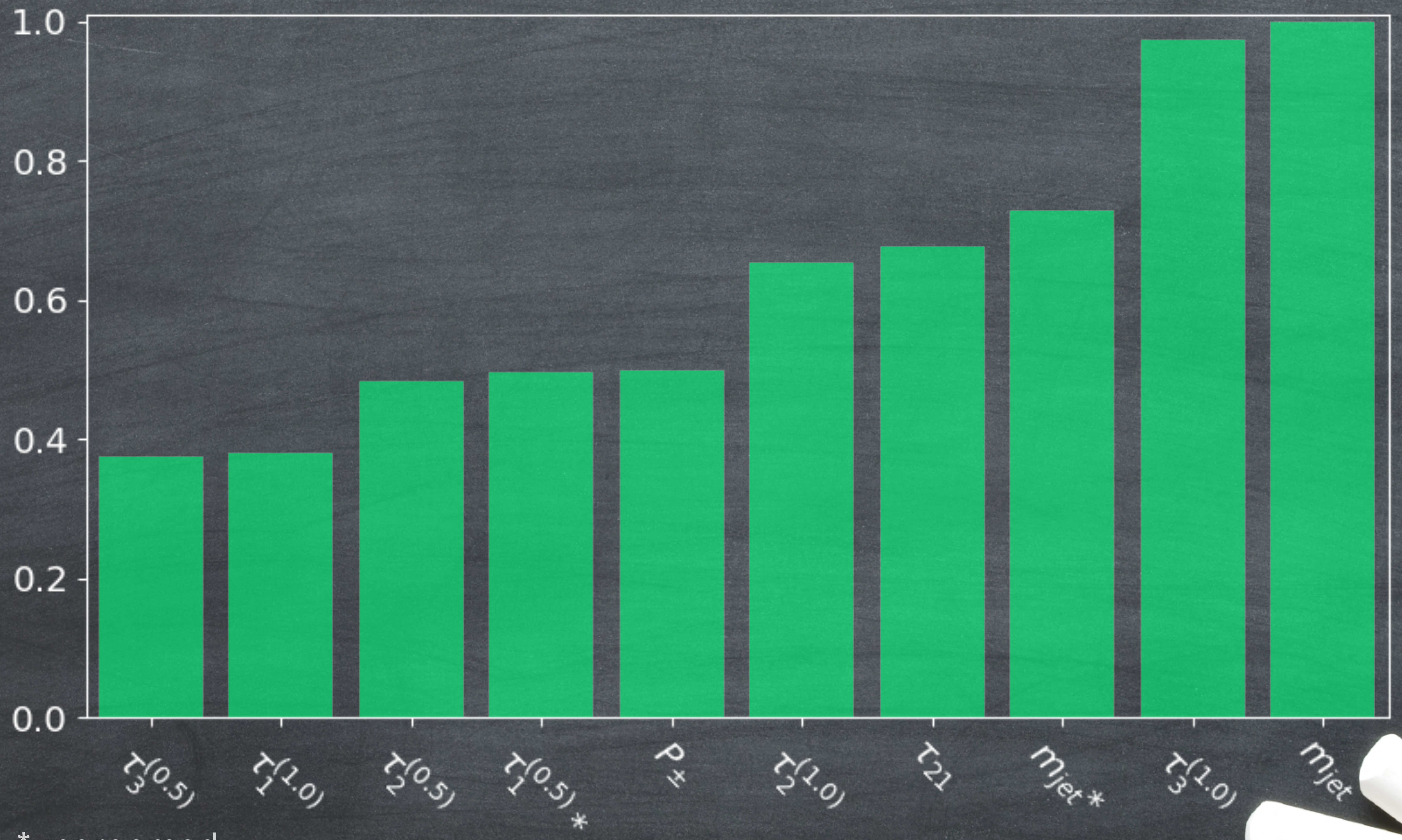


*ungroomed

p_T	η	ϕ	τ	P_e	P_μ	P_γ	$P_\pm$	P_0	θ_P

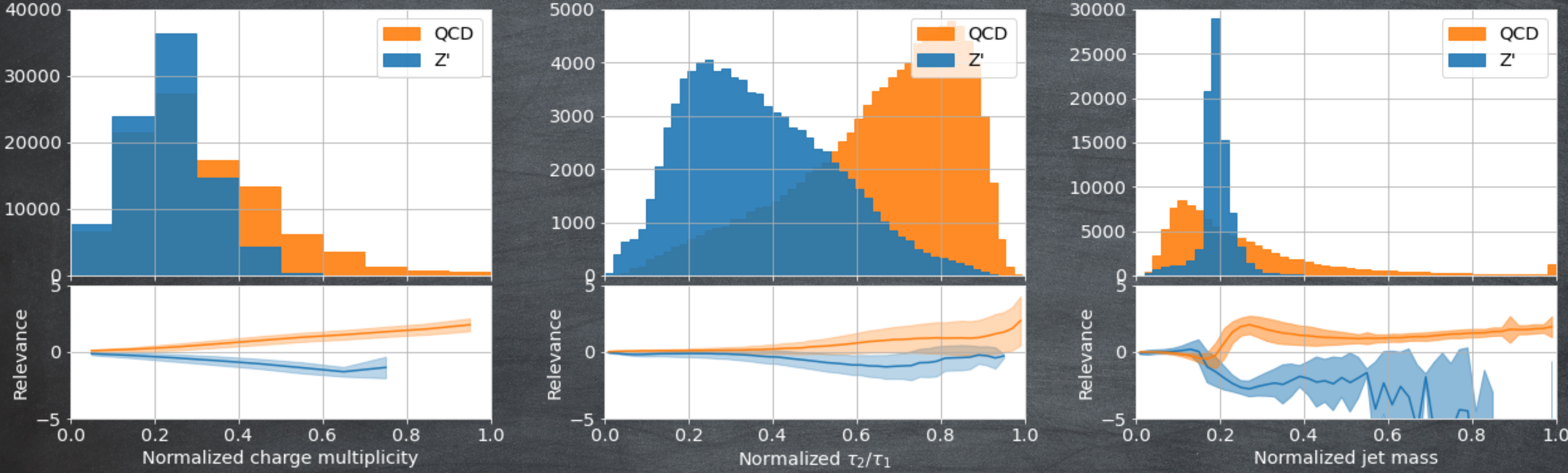
Pythia Simulation – LRP

XAUGS

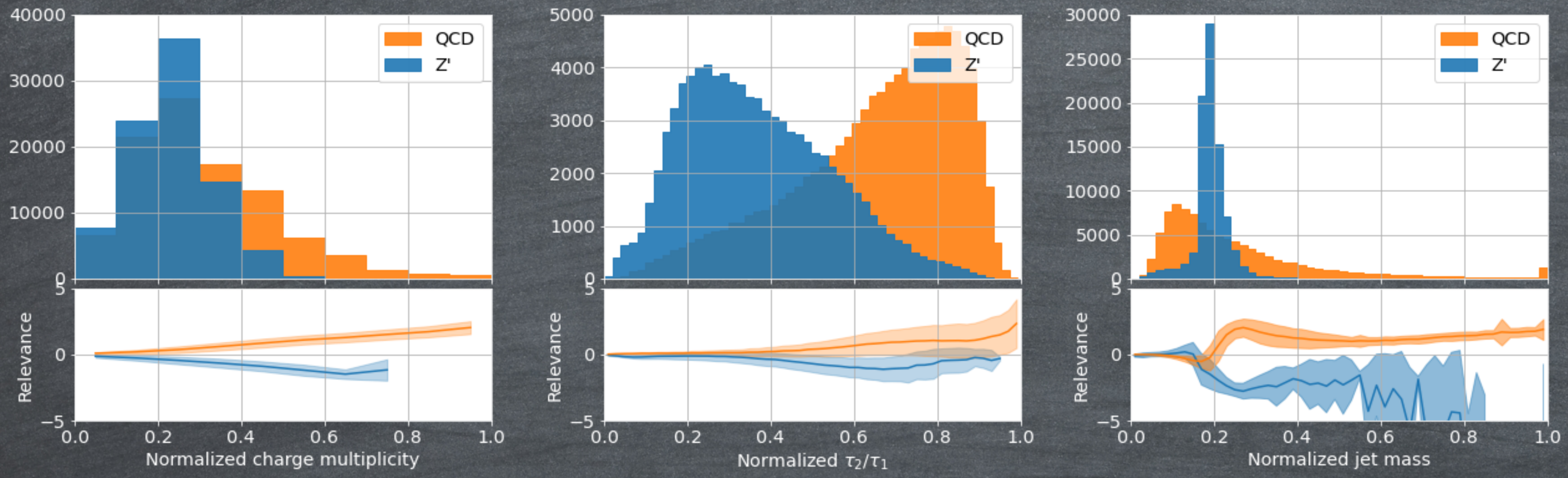


*ungroomed

Pythia Simulation – LRP Profiles of three most relevant inputs



LRP Profiles



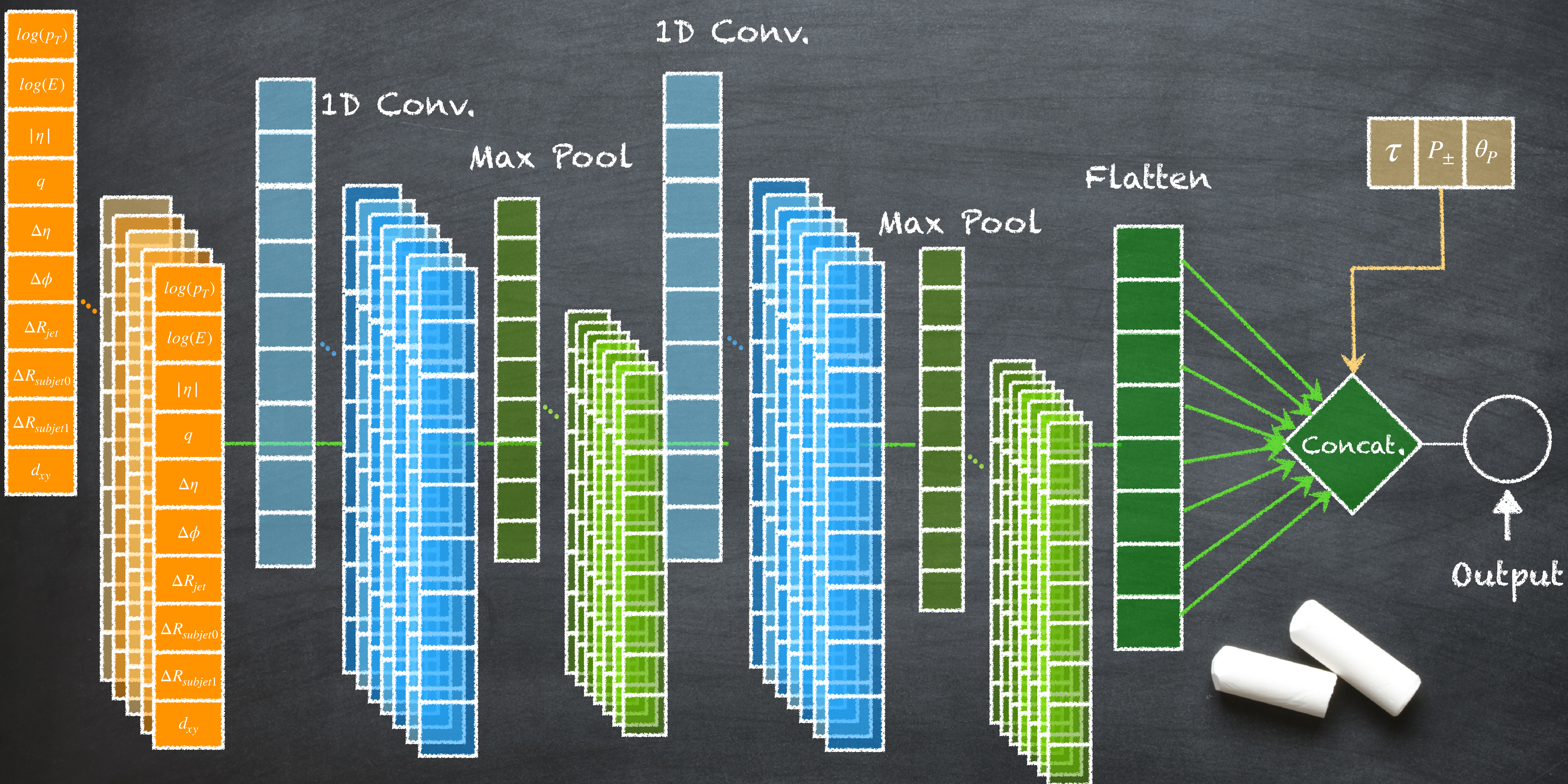
Gives insights into network behavior

XAUG Variables capture features in network

Can improve performance of networks, or reveal when they exhaust available information

Coming soon...

Particle-List CNN



LRP with XAUGs can...

Reduce complexity of taggers

Explore new expert variables

Apply to anomaly detection

References

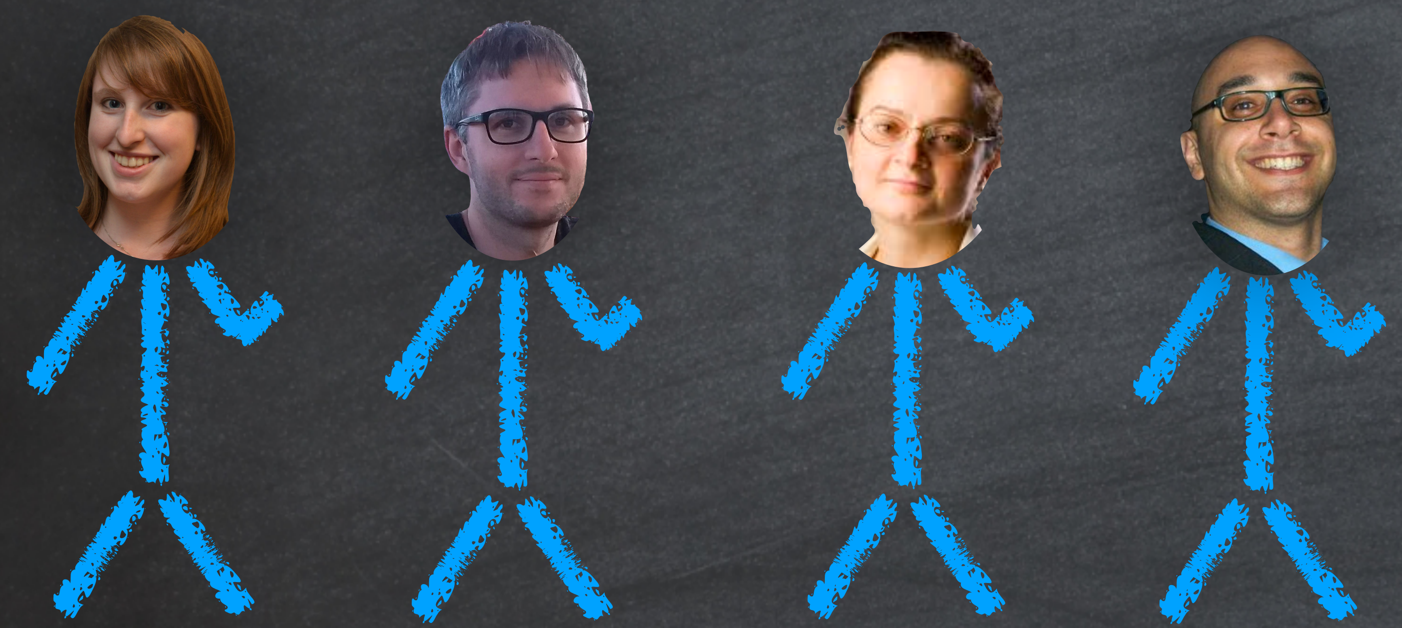
- A. J. Larkoski, I. Moult, and B. Nachman, Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning, Phys. Rept. 841 (2020) 1–63, [arXiv:1709.0446].
- R. Kogler et al., Jet Substructure at the Large Hadron Collider: Experimental Review, Rev. Mod. Phys. 91 (2019), no. 4 045003, [arXiv:1803.0699]
- W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, Definitions, methods, and applications in interpretable machine learning, Proceedings of the National Academy of Sciences 116 (Oct, 2019) 22071–22080.
- S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. MÃŒller, and W. Samek, On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation, PLOS ONE 10 (07, 2015) 1–46.
- K. Datta and A. Larkoski, How Much Information is in a Jet?, JHEP 06 (2017) 073, [arXiv:1704.0824].
- CMS Collaboration, C. Collaboration, Machine learning-based identification of highly Lorentz-boosted hadronically decaying particles at the CMS experiment, CMS-JME-18-002
- J. Cogan, M. Kagan, E. Strauss, and A. Schwarztman, Jet-Images: Computer Vision Inspired Techniques for Jet Tagging, JHEP 02(2015) 118, [arXiv:1407.5675]
- L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman, Jet-images — deep learning edition, JHEP 07 (2016) 069, [arXiv:1511.0519].
- J. Thaler and K. Van Tilburg, Identifying Boosted Objects with N-subjettiness, JHEP 03(2011) 015, [arXiv:1011.2268]
- J. Thaler and K. Van Tilburg, Maximizing Boosted Top Identification by Minimizing N-subjettiness, JHEP 02 (2012) 093, [arXiv:1108.2701]
- T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel C. O. Rasmussen, and P. Z. Skands, An Introduction to PYTHIA 8.2, Comput. Phys. Commun. 191 (2015) 159–177, [arXiv:1410.3012]
- M. Cacciari, G. P. Salam, and G. Soyez, FastJet User Manual, Eur. Phys. J. C72 (2012) 1896, [arXiv:1111.6097]
- M. Cacciari, G. P. Salam, and G. Soyez, The anti-kT jet clustering algorithm, JHEP04(2008) 063, [arXiv:0802.1189]
- M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, Towards an understanding of jet substructure, JHEP 09 (2013) 029, [arXiv:1307.0007]
- A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, Soft Drop,JHEP 05(2014) 146,[arXiv:1402.2657]

Our team

Our team

Lauren Hay | Benjamin Mannix | Margaret Morris | Garvita Agarwal

Graduate students



Christine McLean | Ulrich Schubert | Ia Iashvili | Sal Rappoccio

Postdocs | Professors

Boostemos!

1846