

Boosting DM searches at the LHC with deep learning

Thorben Finke

RWTH Aachen - Institute for Theoretical Particle Physics and Cosmology

finke@physik.rwth-aachen.de

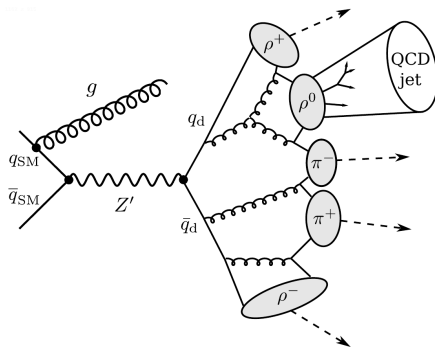
Based on [arXiv2204.11889](https://arxiv.org/abs/2204.11889) in collaboration with Michael Krämer, Maximilian Lipp,
Alexander Mück

July 21, 2022

Signal region

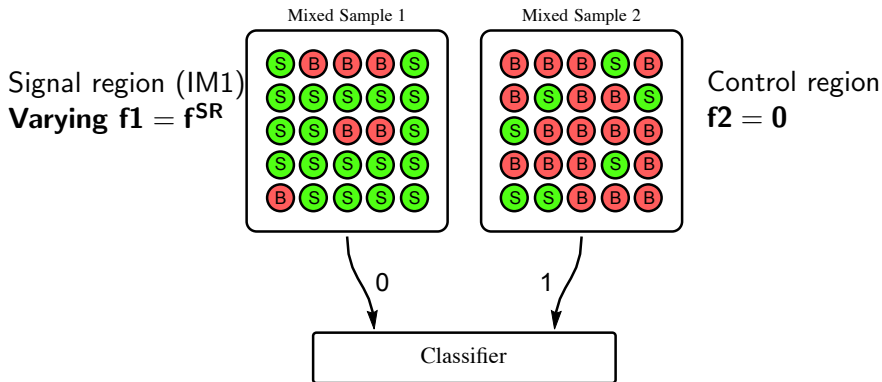
- ✓ $E_T^{\text{miss}} > 200 \text{ GeV}$ with various inclusive and exclusive bins
- ✓ leading anti- k_T $R = 0.4$ jet with $p_T^{\text{jet}} > 150 \text{ GeV}$, $|\eta^{\text{jet}}| < 2.4$
- ✓ at most 3 additional jets with $p_T^{\text{jet}} > 30 \text{ GeV}$, $|\eta^{\text{jet}}| < 2.8$
- ✓ $\Delta\phi(\mathbf{p}_T^{\text{jet}}, \mathbf{p}_T^{\text{miss}}) > 0.4$
- ✓ lepton veto

- $\mathcal{O}(10^6)$ background events in IM1 ($E_T^{\text{miss}} > 250 \text{ GeV}$)
- Main backgrounds:
 - **Z+jet production with invisibly decaying Z (61%)**
 - W+jet production with leptonically decaying W and non-identification of the charged lepton (31%)
 - Top quark production (3.5%)
 - Diboson production (2%)
- 95% CL exclusion of more than $\sim 40\text{k}$ signal events



- Dark sector coupling to the SM via vector mediator Z'
- On-shell production of Z' decaying to dark quarks
- Shower and hadronization in dark sector
- ρ_d^0 decay to standard model $\rightarrow r_{inv} = 0.75$

Mono-jet search is sensitive to this model, if **one of the dark jets stays invisible.**



- Two samples M_1 and M_2 with signal fractions f_1, f_2 with $f_1 > f_2$
- Optimal classifier on M_1 vs. M_2 is also optimal classifier for S vs. B
- Set classifier cut such that 1 per mille of control region events is kept

Proof of principle – using only Z + jet

For an increasing fraction of semi-visible jets in the signal region we see increasing sensitivity:

f^{SR}	$n_{\text{exp}}^{\text{SR}}$	n^{SR}	n_{A}^{SR}	n_{B}^{SR}	$(n^{\text{SR}} - n_{\text{exp}}^{\text{SR}}) / \sqrt{2 n_{\text{exp}}^{\text{SR}}}$
0%	1000	1048	0	1048	1.07
0.5%	1000	1175	184	991	3.91
0.6%	1000	1306	247	1059	6.84
1%	1000	1666	625	1041	14.89

→ no fake signal produced

Proof of principle – using only Z + jet

For an increasing fraction of semi-visible jets in the signal region we see increasing sensitivity:

f^{SR}	$n_{\text{exp}}^{\text{SR}}$	n^{SR}	n_{A}^{SR}	n_{B}^{SR}	$(n^{\text{SR}} - n_{\text{exp}}^{\text{SR}}) / \sqrt{2 n_{\text{exp}}^{\text{SR}}}$
0%	1000	1048	0	1048	1.07
0.5%	1000	1175	184	991	3.91
0.6%	1000	1306	247	1059	6.84
1%	1000	1666	625	1041	14.89

→ no fake signal produced

→ steep drop in performance once the classifier cannot train anymore

Proof of principle – using only Z + jet

For an increasing fraction of semi-visible jets in the signal region we see increasing sensitivity:

f^{SR}	$n_{\text{exp}}^{\text{SR}}$	n^{SR}	n_{A}^{SR}	n_{B}^{SR}	$(n^{\text{SR}} - n_{\text{exp}}^{\text{SR}}) / \sqrt{2 n_{\text{exp}}^{\text{SR}}}$
0%	1000	1048	0	1048	1.07
0.5%	1000	1175	184	991	3.91
0.6%	1000	1306	247	1059	6.84
1%	1000	1666	625	1041	14.89

→ no fake signal produced

→ steep drop in performance once the classifier cannot train anymore

→ High sensitivity far below the 95% CL of the original search

Additional backgrounds

We want to see how well we need to model the background – we construct a **signal region with 3.5% top events and 2% dibosons** in addition to 1% semi-visible jets.

$r_{t\bar{t}}^{\text{CR}}$	r_{VV}^{CR}	$n_{Z+\text{jet}}^{\text{CR}}$	$n_{t\bar{t}}^{\text{CR}}$	n_{VV}^{CR}	$n_{Z+\text{jet}}^{\text{SR}}$	$n_{t\bar{t}}^{\text{SR}}$	n_{VV}^{SR}	n_A^{SR}	n^{SR}
0%	0%	1000	0	0	1089	1245	1826	223	4383
2.80%	1.6%	876	108	16	838	134	37	456	1465
3.50%	2.0%	963	23	14	996	25	32	633	1686
5.00%	3.0%	971	12	17	1034	8	10	575	1627

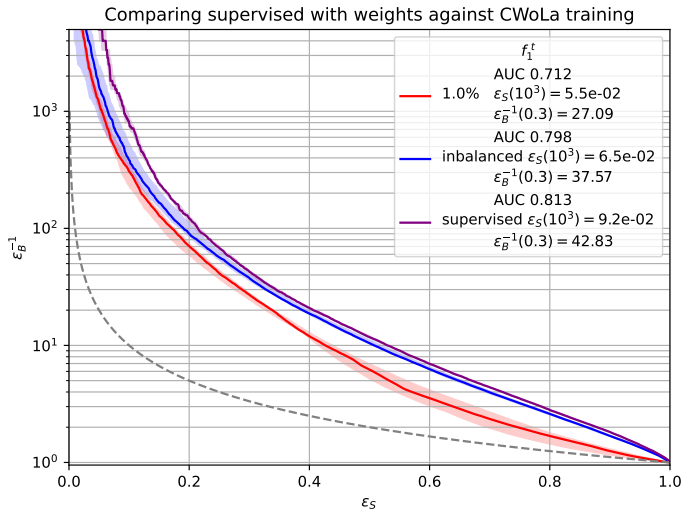
- Classification without labels (CWoLa) can be used to boost the discovery potential of existing searches
- The method will find any discrepancy between signal and control sample
 - Can be applied in an **iterative process**, explaining observed excesses until no excess (or no explanation) is left
- After setting a signal region, the method is completely **model agnostic**, which allows the search for very different signatures and model parameters

BACKUP

Proof of principle, full table

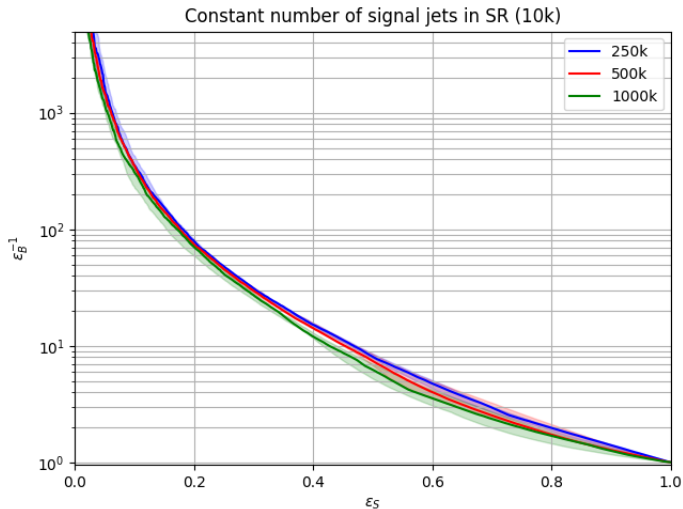
f^{SR}	$n_{\text{exp}}^{\text{SR}}$	n^{SR}	n_{A}^{SR}	n_{B}^{SR}	$(n^{\text{SR}} - n_{\text{exp}}^{\text{SR}}) / \sqrt{2 n_{\text{exp}}^{\text{SR}}}$
0%	1000	1048	0	1048	1.07
0.2%	1000	1065	47	1018	1.45
0.4%	1000	1107	100	1007	2.39
0.5%	1000	1175	184	991	3.91
0.6%	1000	1306	247	1059	6.84
0.7%	1000	1389	367	1022	8.70
0.8%	1000	1500	419	1081	11.18
1%	1000	1666	625	1041	14.89
2%	1000	2357	1392	965	30.34
4%	1000	4182	3269	913	71.15

Comparison supervised vs. CWoLa



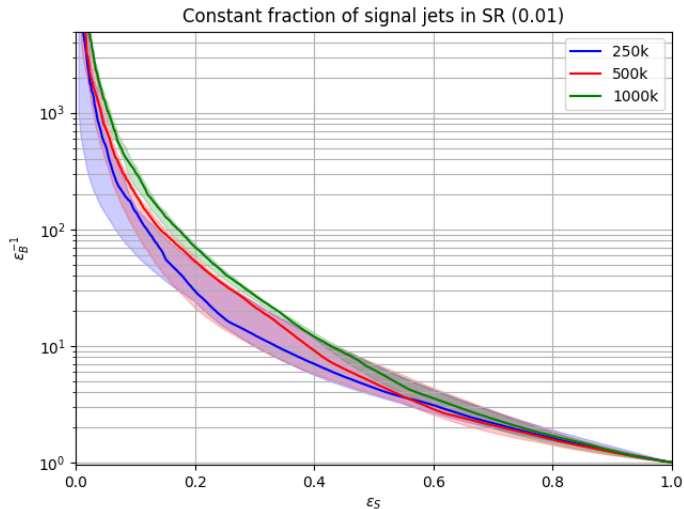
Changing the region size with fixed number of signal jets

Always using 10k signal jets in the signal region (only Z+jet background)



Changing the region size with fixed fraction of signal jets

Always using $f^{SR} = 0.01$ (only Z+jet background)



Classifier scores

Signal fraction $f^{SR} = 0.01$, t is the chosen threshold

