ML4Jets 2021

Super-Resolution for QCD and Top Jets

Pierre Baldi[1], Lukas Blecher[2], Anja Butter[2], Julian Collado[1], Jessica Howard[3], Fabian Keilbach[2], Tilman Plehn[2], Gregor Kasieczka[4], and Daniel Whiteson[3]

July 7, 2021

Lukas Blecher

1 Department of Computer Science, University of California, Irvine, US
2 Institut für Theoretische Physik, Universität Heidelberg, Germany
3 Department of Physics and Astronomy, University of California, Irvine, US
4 Institut für Experimentalphysik, Universität Hamburg, Germany

Preprint: `https://arxiv.org/abs/2012.11944`

Motivation

Definition: Super Resolution

In single image super resolution (SISR) the goal is to predict a sensible high resolution (HR), super resolved (SR) version of a given low resolution (LR) image

- Use established SR method as starting point: ESRGAN [1]
- Generative Adversarial Network [2] setup

(a) High Resolution (b) Low Resolution (c) Bicubic Upsampling (d) Super Resolution

Figure: Super resolution on the STL-10 [3] testset using the ESRGAN

(a) High Resolution (b) Low Resolution (c) Bicubic Upsampling (d) Super Resolution

Question

Can an upsampled jet image include more information than the original, low-resolution image?

Dataset

- Use PYTHIA [4] to generate $t\bar{t}$ and QCD dijet events
- Center of mass energy of $\sqrt{s} = 14\,\text{TeV}$
- Run DELPHES [5] with standard ATLAS card $\rightarrow$ HR version (160×160)

Dataset

- Use PYTHIA [4] to generate $t\bar{t}$ and QCD dijet events
- Center of mass energy of $\sqrt{s} = 14$ TeV
- Run DELPHES [5] with standard ATLAS card → HR version (160×160)
- Perform downsampling step (sum pooling $f = 8$) → LR version (20×20)
- Run anti-kt jet algorithm using FASTJET [6] on HR & LR
- Filter: Jet $p_T \in [550, 650]$ GeV, $|\eta|_{\text{jet}} < 2$ and $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.1$

We end up with paired dataset of LR & HR images of the same event.

Dataset

Sparse images: 99.80% empty
Individual constituents can have transverse momenta of up to $p_T = 500\,\text{GeV}$, but over 85% have $p_T < 20\,\text{GeV}$
$\Rightarrow$ Sharp and wide distribution, hard to learn for network
Solution: Raise image to a power $p \in (0, 1)$ in a pixel-wise fashion

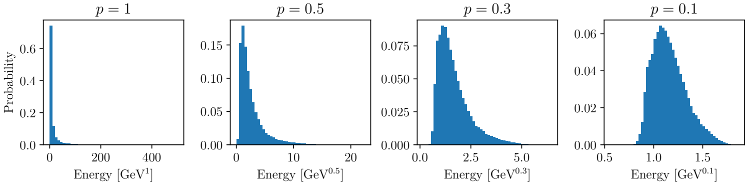


Figure: Energy distribution behaviour when raising it to different powers p

Training Process

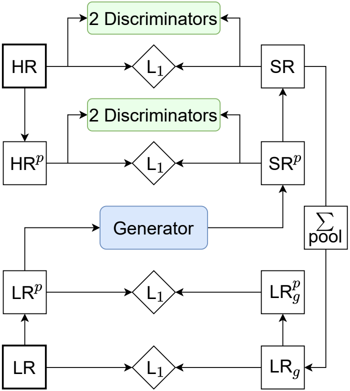


Figure: Training iteration

$$L_G = \sum_{s \in \{\text{std}, \text{pow}\}} \lambda_s \big(\lambda_{\text{HR}} \, L_{\text{HR}} + \lambda_{\text{LR}} \, L_{\text{LR}} + \lambda_{\text{adv}} \, L_{\text{adv}} + \lambda_{\text{patch}} \, L_{\text{patch}} \big)$$

Patch Loss

Patch loss helps to balance the spread of constituents

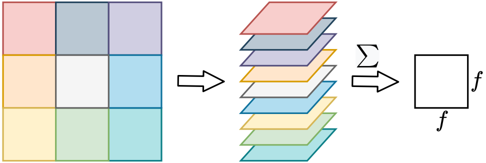


Figure: Patch rearrangement. f is the upscaling factor

Sum over created dimension.
Compare using **Mean Squared Error**

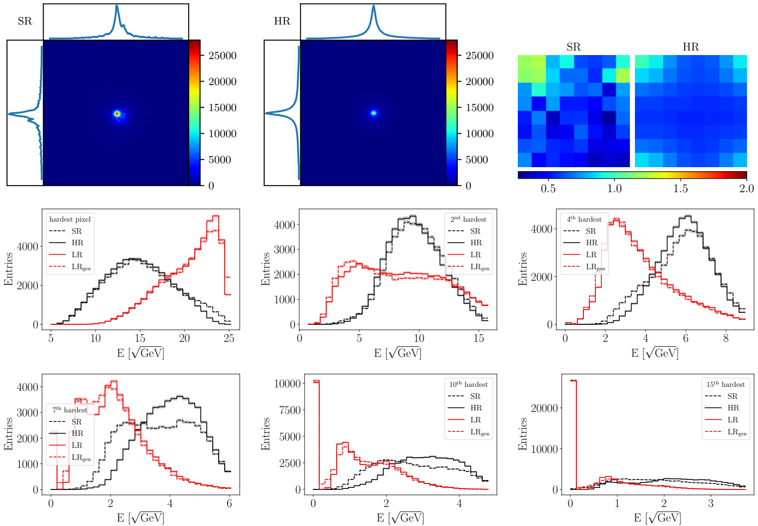
$$L_{\text{patch}} = L_2\left(\text{patch}(\text{SR}), \text{patch}(\text{HR})\right) \tag{1}$$

Jet observables

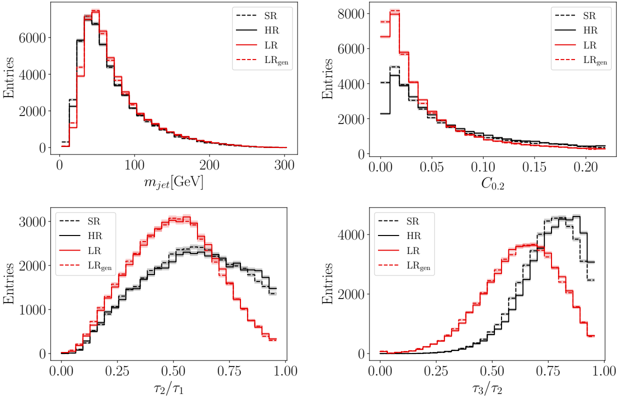
Compare p_T distribution for the n^th hardest jet and set of high-level jet observables [7–10]

$$
\begin{aligned}
m_\text{jet} &= \left(\sum_i p_i^\mu\right)^2 \\
C_{0.2} &= \frac{\sum_{i,j} p_{\mathrm{T},i} p_{\mathrm{T},j} (\Delta R_{i,j})^{0.2}}{(\sum_i p_{\mathrm{T},i})^2} \\
\tau_N &= \frac{\sum_k p_{\mathrm{T},k} \min(\Delta R_{1,k}, \ldots, \Delta R_{N,k})}{\sum_k p_{\mathrm{T},k} R_0}
\end{aligned}
$$

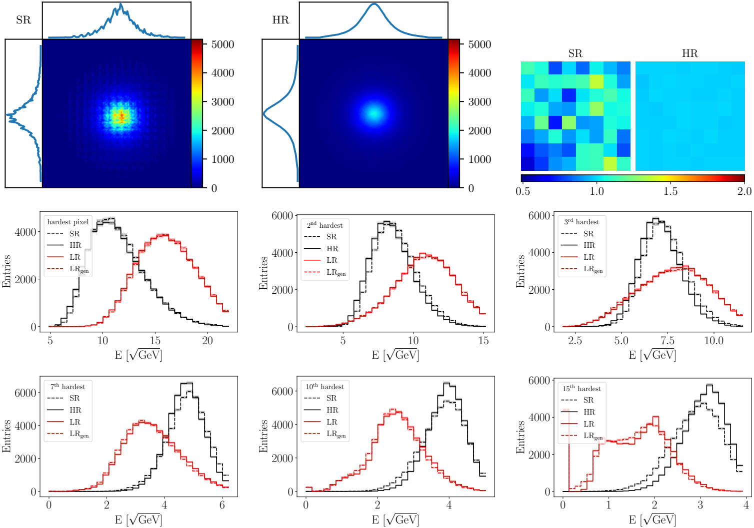
Performance for QCD Jets – Low level



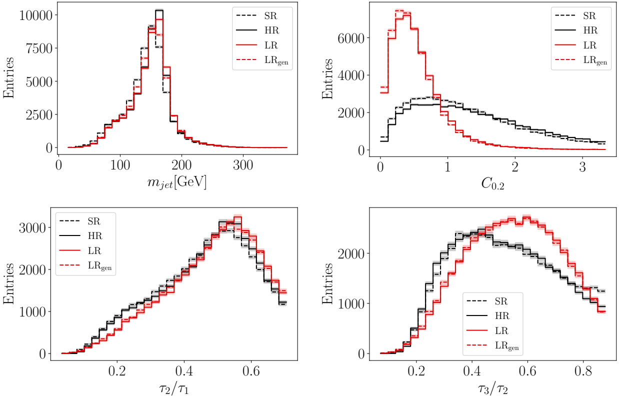
Performance for QCD Jets – High level



Performance for Top Jets – Low level



Performance for Top Jets – High level



Event-Event

Until now: Distributions over entire test set
Number of samples drown out individual effect

Question

Does up-sampling add information to an individual jet?

Goal is not to reconstruct the true HR jet. SR jet needs to be *consistent* with LR jet.
For given observable look at deviations from true value on event by event basis

$$\frac{\mathrm{HR}-\mathrm{LR}}{\mathrm{HR}} \quad \text{and} \quad \frac{\mathrm{HR}-\mathrm{SR}}{\mathrm{HR}}$$

Also look at relation of deviations

$$\frac{|\mathrm{HR}-\mathrm{SR}|}{|\mathrm{HR}-\mathrm{LR}|} = \begin{cases} <1, & \text{SR describes HR better than LR} \\ >1, & \text{LR describes HR better than SR} \end{cases}$$

Event-Event

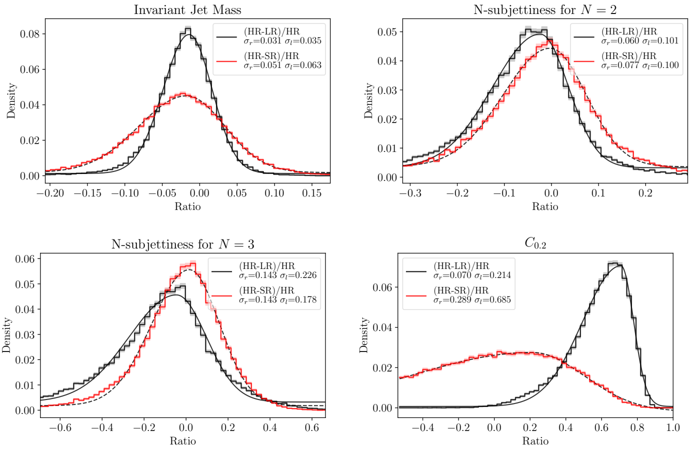


Figure: Top jets: Relative ratio for jet observables

Event-Event

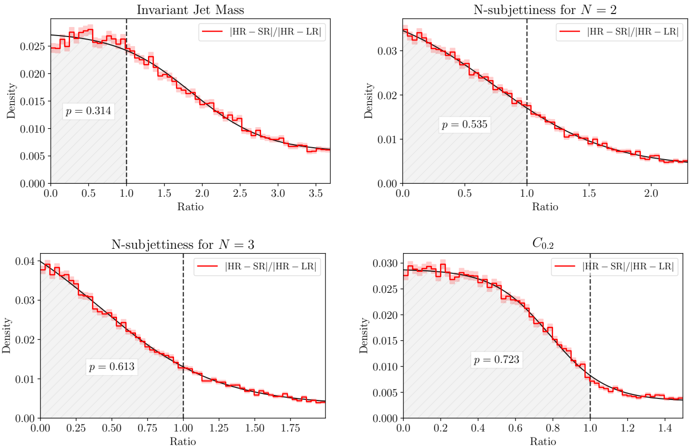


Figure: Top jets: Correlation of ratios for jet observables

Conclusion

- Introduce new application of deep learning to jet physics
- Able to generate sensible 8-fold super-resolved jet images
- Super-resolution networks can provide additional information
- Can be used to enhance jet measurements in regions with poor calorimeter resolution

References I

[1] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, C. C. Loy, Y. Qiao, and X. Tang, "ESRGAN: enhanced super-resolution generative adversarial networks", CoRR **abs/1809.00219** (2018).

[2] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Generative adversarial networks*, 2014.

[3] A. Y. N. Adam Coates Honglak Lee, "An analysis of single layer networks in unsupervised feature learning", (2011).

[4] 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**, 159–177 (2015).

[5] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi, "Delphes 3: a modular framework for fast simulation of a generic collider experiment", Journal of High Energy Physics **2014**, 10.1007/jhep02(2014)057 (2014).

[6] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual", Eur. Phys. J. C **72**, 1896 (2012).

[7] J. Gallicchio, J. Huth, M. Kagan, M. D. Schwartz, K. Black, and B. Tweedie, "Multivariate discrimination and the Higgs + W/Z search", JHEP **04**, 069 (2011).

[8] A. J. Larkoski, G. P. Salam, and J. Thaler, "Energy Correlation Functions for Jet Substructure", JHEP **06**, 108 (2013).

[9] J. Thaler and K. Van Tilburg, "Identifying Boosted Objects with N-subjettiness", JHEP **03**, 015 (2011).

[10] G. Kasieczka, N. Kiefer, T. Plehn, and J. M. Thompson, "Quark-Gluon Tagging: Machine Learning vs Detector", SciPost Phys. **6**, 069 (2019).

[11] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network", CoRR **abs/1609.04802** (2016).

Generator

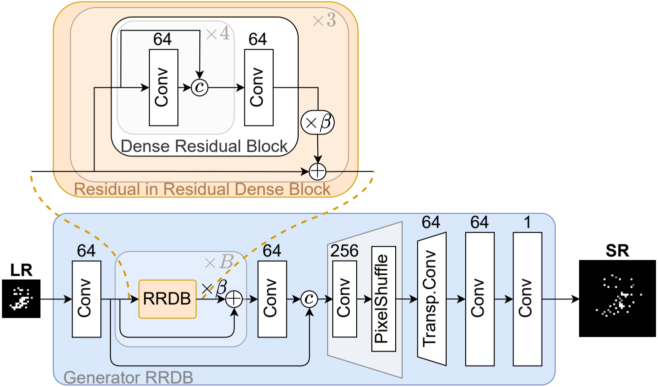


Figure: Generator structure [1, 11]

Discriminator

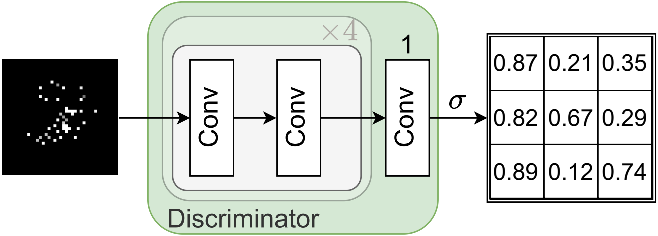


Figure: Markovian discriminator

Multi-Power Learning

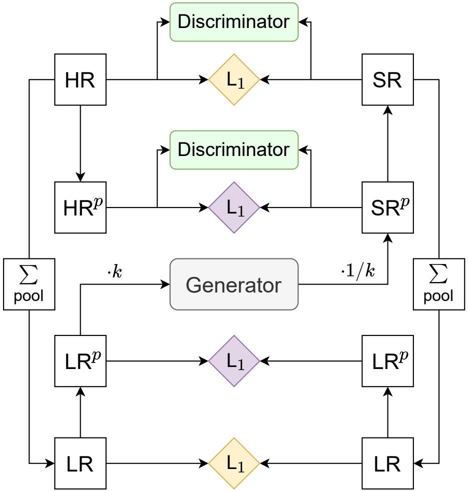


Figure: Training iteration – multi-power loss

Multi-Power Learning

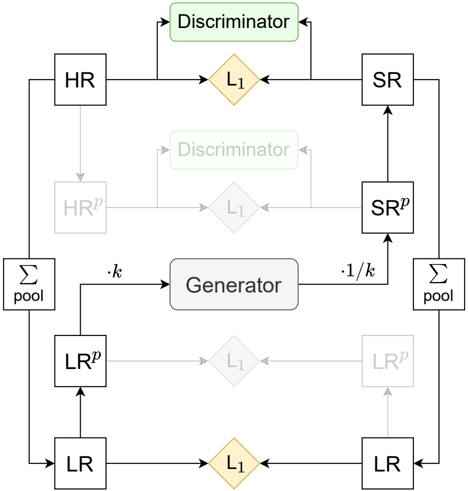


Figure: Training iteration – standard loss

Multi-Power Learning

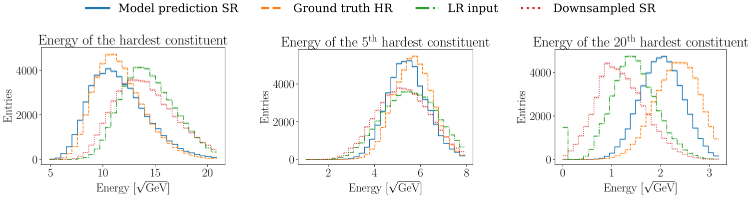


Figure: Energy distributions with only the standard loss

Multi-Power Learning

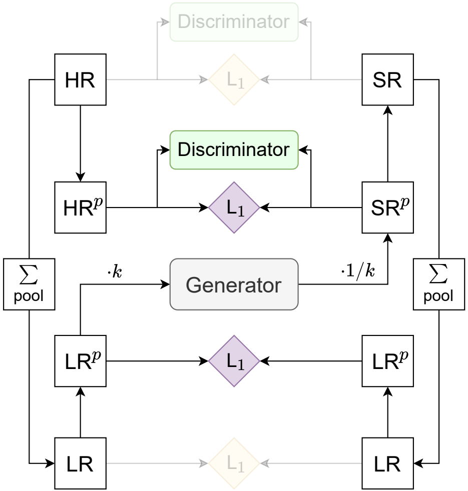


Figure: Training iteration – power loss

Multi-Power Learning

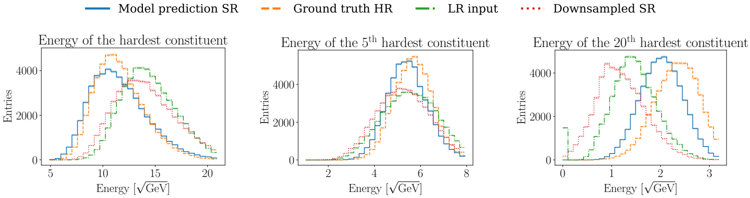


Figure: Energy distributions with only the standard loss

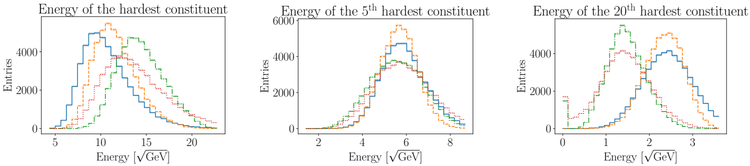


Figure: Energy distributions with only the power loss